



ISSN: 1646-9895

Revista Ibérica de Sistemas e Tecnologias de Informação
Revista Ibérica de Sistemas y Tecnologías de Información

J u n h o 2 3 • J u n e 2 3



©RISTI 2023 <http://www.risti.xyz>

Nº 50

Edição / Edición

Nº 50, 06/2023

Tiragem / Tirage: 1000

Preço por número / Precio por número: 17,5€

Subscrição anual / Suscripción anual: 30€ (2 números)

ISSN: 1646-9895

Depósito legal:

Indexação / Indexación

Academic Journals Database, Dialnet, DOAJ, DOI, EBSCO, GALE, Google Scholar, IndexCopernicus, Information Systems Journal, Latindex, ProQuest, QUALIS, SciELO, SCImago, Scopus, SIS, Ulrich's, Web of Science

Publicação / Publicación

RISTI – Revista Ibérica de Sistemas e Tecnologias de Informação

Pr. 9 de Abril, nº 26, 4200-422 Porto, Portugal

E-mail: aistic@gmail.com

Web: <http://www.risti.xyz>

Director

Álvaro Rocha, ISEG, Universidade de Lisboa, PT

Coordenadores da Edição / Coordinadores de la Edición

Álvaro Rocha, ISEG, Universidade de Lisboa, PT

Maria José Sousa, ISCTE-IUL, PT

Conselho Editorial / Consejo Editorial

A. Patrícia Oliveira Universidade de Aveiro, PT

Abel Méndez, Instituto Tecnológico de Costa Rica, CR

Alejandro Peña, Universidad EAFIT, CO

Alma Gómez-Rodríguez, Universidade de Vigo, ES

Ana Melro, Universidade de Aveiro, PT

Ana Paula Afonso, P.PORTO/ISCAP.IPP/CEOS.PP, PT

Ana Rita Calvão, ESTGA, Universidad de Aveiro, PT

Ania Cravero, Universidad de La Frontera, CL

António Abreu Silva, ISCAP, Instituto Politécnico do Porto, PT

Antonio Garcia, Universidad de Santiago de Compostela, ES

António Godinho, ISEC, Instituto Politécnico de Coimbra, PT

Antonio Jiménez-Martín, Universidad Politécnica de Madrid, ES

Arturo J. Méndez, Universidad de Vigo, ES

August Climent, Ramon LLull University, ES

Beatriz Rodríguez, Universidad de la Republica, UY

Borja Bordel, Universidad Politécnica de Madrid, ES

Brenda L. Flores-Rios, Universidad Autónoma de Baja California, MX

Bruno Miguel Ferreira Gonçalves, Instituto Politécnico de Bragança, PT

Carlos Alexandre Silva, Instituto Federal de Minas Gerais, BR

Carlos Carreto, Instituto Politécnico da Guarda, PT

Carlos Morais, Instituto Politécnico de Bragança, PT

Carlos Rompante Cunha, UNIAG & CeDRI & Instituto Politécnico Bragança, PT

Carlos Vaz de Carvalho, Instituto Politécnico do Porto, PT

Célio Marques, Instituto Politécnico de Tomar, PT

Ciro Martins, Universidade do Algarve, PT

Cristina Caridade, Instituto Politécnico de Coimbra, PT

Daniel Polónia, Universidade de Aveiro, PT
Dante Carrizo, Universidad de Atacama, CL
Edwin Cedeño Herrera, Universidad de Panamá, PA
Fábio Marques, ESTGA, Universidade de Aveiro, PT
Felipe Vasquez, Universidad de La Frontera, CL
Fernando Moreira, REMIT, IJP, Universidade Portucalense & IEETA,
Universidade de Aveiro, PT
Fernando Ribeiro, Instituto Politécnico de Castelo Branco, PT
Fernando Bandeira, PT
Filipe Cardoso, Instituto Politécnico de Viseu, PT
Flor Gomes de María Sánchez Aguirre, Universidad César Vallejo, PE
Francisco Javier Lena-Acebo, Universidad de Cantabria, ES
Gabriel Guerrero-Contreras, Universidad de Cádiz, ES
Gerardo González Filgueira, Universidade da Coruña, ES
Gloria Valencia, Universidad de las Fuerzas Armadas, EC
Guilllhermo Davila, Universidad de Lima, PE
Hélder Gomes, Universidade de Aveiro, PT
Helder Zagalo, Universidade de Aveiro, PT
Hélia Guerra, Universidade dos Açores, PT
Henrique Gil, Instituto Politécnico de Castelo Branco, PT
Henrique S. Mamede, Universidade Aberta, Portugal
Inês Domingues, ISEC, Instituto Politécnico de Coimbra, PT
Isaias Bianchi, Universidade Federal de Santa Catarina, BR
Isidro Navarro, Universidad Politécnica de Cataluña, ES
Ismael Etxeberria-Agiriano, Universidad del País Vasco, ES
Ivan Garcia, Universidad Tecnológica de la Mixteca, MX
Jeimy Cano, Universidad de los Andes, CO
João Balsa, LASIGE, Universidade de Lisboa, Portugal
João Emílio Almeida, ISTE - Porto, PT
João Paulo Ferreira, Instituto Politécnico de Coimbra, PT
Joao Roberto de Toledo Quadros, CEFET/RJ, BR
Joao Tavares, Universidade do Porto, PT
Joaquim Reis, ISCTE - Instituto Universitário de Lisboa, PT
Jorge Eduardo Ibarra Esque, Universidad Autónoma de Baja California, MX

Jorge Hochstetter, Universidad de La Frontera, CL
José Álvarez-García, Universidad de Extremadura, ES
Jose Felipe Cocon Juarez, Universidad Autónoma del Carmen, MX
José Lousado, Instituto Politécnico de Viseu, PT
José Luis Pastrana Brincones, Universidad de Málaga, ES
Jose M. Molina, Universidad Carlos III de Madrid, ES
José Ribeiro, Politécnico de Leiria, PT
Jose Silvestre Silva, Academia Militar, PT
Josep Maria Marco-Simó, Universitat Oberta de Catalunya, ES
Juan Angel Contreras Vas, Universidad de Extremadura, ES
Juan M. Santos-Gago, Universidad de Vigo, ES
Juan Pablo DAmato, Universidad Nacional del Centro de la
Provincia de Buenos Aires, AR
Leila Weitzel, Universidade Federal Fluminense, BR
Leonardo Bermón Angarita, Universidad Nacional de Colombia, CO
Lilia Muñoz, Universidad Tecnológica de Panamá, PA
Lucila Romero, Universidad Nacional del Litoral, AR
Luis Álvarez Sabucedo, Universidad de Vigo, ES
Luís Bruno, Instituto Politécnico de Beja, PT
Luis Enrique Sánchez Crespo, Universidad de Castilla-la Mancha, ES
Luisa María Romero-Moreno, Universidad Sevilla, ES
Luz María Hernández Cruz, Universidad Autónoma de Campeche, MX
Luz Sussy Bayona Oré, Universidad Nacional Mayor de San Carlos, PE
Marcelo Marciszack, Universidad Tecnológica Nacional, AR
Marcelo Zambrano Vizuite, Universidad Tecnica Del Norte
Marco Painho, NOVA IMS, PT
Margarita Ramirez Ramirez, Universidad Autónoma de Baja California, MX
Maria Cristina Marcelino Bento, UNIFATEA, BR
María de la Cruz del Río-Rama, Universidad de Vigo, ES
Maria de los Milagros Gutierrez, Universidad Tecnológica Nacional, AR
Maria do Rosário Bernardo, Universidade Aberta, PT
Maria Hallo, Escola Politécnica Nacional, EC
Maria João Ferreira, Universidade Portucalense, PT

Maria João Gomes, Universidade do Minho, PT
Maria Sousa, ISCTE - Instituto Universitário de Lisboa, PT
Marisol B. Correia, ESGHT - Universidade do Algarve & CiTUR, PT
Maristela Holanda, Universidade de Brasília, BR
Martin Llamas Nistal, Universidad de Vigo, ES
Miguel Casquilho, Universidade de Lisboa, PT
Miguel Ramón González Castro, Centro Tecnológico Aimen, ES
Mirna Ariadna Muñoz Mata, CIMAT, MX
Nelson Rocha, Universidade de Aveiro, PT
Nuno Melão, Instituto Politécnico de Viseu, PT
Nuno Ribeiro, Universidade Fernando Pessoa, Portugal
Patricia Dias, Universidade do Estado de Minas Gerais, BR
Paula Prata, Universidade da Beira Interior, PT
Paulo Martins, Universidade de Trás-os-Montes e Alto Douro, PT
Paulo Pinto, Universidade Nova de Lisboa, PT
Paulo Rurato, Universidade Fernando Pessoa, Portugal
Pedro Araújo, Universidade da Beira Interior, PT
Pedro R. Palos- Sanchez, Universidad de Sevilla, ES
Pedro Sanz-Angulo, Universidad de Valladolid, ES
Pedro Sobral, Universidade Fernando Pessoa, Portugal
Pedro Sousa, Universidade do Minho, Portugal
René Faruk Garzozi-Pincay, Universidad Estatal Península de Santa Elena, EC
Rita Oliveira, DigiMedia, Universidade de Aveiro, PT
Rita Santos, ESTGA/DigiMedia, Universidade de Aveiro, Portugal
Ruben Pereira, ISCTE – Instituto Universitário de Lisboa, PT
Rui Pedro Marques, Universidade de Aveiro, PT
Rui Silva Moreira, Universidade Fernando Pessoa, PT
Samuel Sepúlveda, Universidad de La Frontera, CL
Santiago Raul Gonzales Sanches, Universidad Cesar Vallejo, PE
Sara Balderas-Díaz, Universidad de Cádiz, ES
Sergio Araya Guzmán, Universidad del Bío-Bío, CL
Sergio F. Lopes, Universidade do Minho, PT
Sergio Gálvez Rojas, Universidad de Málaga, ES

Solange N Alves-Souza, Universidade de São Paulo, BR
Telmo Silva, DigiMedia, Universidade de Aveiro, PT
Teresa Guarda, Universidad Estatal Península de Santa Elena, EC
Thiago Dias, Centro Federal de Educação Tecnológica de Minas Gerais, BR
Vitor Carvalho, Instituto Politécnico do Cávado e do Ave, PT
Vitor Santos, Universidade Nova de Lisboa - NOVA IMS, PT
Wagner Tanaka Botelho, Universidade Federal do ABC, BR
Yamid Hernández Julio, Universidad del Sinú, CO

Índice / Index

EDITORIAL

Quinze anos de RISTI – Revista Ibérica de Sistemas e Tecnologias de Informação: cinquenta números regulares e sessenta números especiais	1
--	---

Álvaro Rocha, Maria José Sousa

ARTIGOS / ARTICULOS / ARTICLES

Una revisión sistemática de Modelos de clasificación de dengue utilizando machine learning	5
--	---

Gisella Luisa Elena Maquen-Niño, Jessie Bravo, Roger Alarcón, Ivan Adrianzén-Olano, Hugo Vega-Huerta

Influência dos indicadores relacionados com governo eletrônico no Índice de Transparência Municipal (ITM) de Portugal.....	28
--	----

Kaio Vieira

Efeitos da realidade aumentada na aprendizagem das ciências naturais: uma revisão sistemática da literatura.....	44
--	----

António Faria, Guilhermina Lobato Miranda

Sistemas Automáticos de Calificación de Tareas de Programación: Revisión y Crítica.....	58
---	----

Carlos Cares, Ricardo Gacitúa

Modelo de tendencias para series complejas: caso de estudio tarifas hoteleras	73
---	----

Abraham Briones Juárez, Judith Alejandra Velázquez Castro

Divergencia en los niveles de madurez de e-gobierno en México	91
---	----

Oscar Yahev Carrera Mora

El efecto de un programa de psicoterapia de telesalud en la salud mental de jóvenes en alto riesgo psicosocial	106
--	-----

Manchego Carnero Brígida-Aurora, Sivincha Huamani Karen, Manchego Carnero Rocío-Edith

Sistemas conversacionais: aspetos culturais influenciam a avaliação de UX de um protótipo conceptual de assistente de voz	118
---	-----

Juliana Camargo, José Nunes, Maria João Antunes, Óscar Mealha, Carolina Abrantes e Luís Nóbrega

Quinze anos de RISTI – Revista Ibérica de Sistemas e Tecnologias de Informação: cinquenta números regulares e sessenta números especiais

Fifteen years of RISTI – Revista Ibérica de Sistemas e Tecnologias de Informação: fifty regular issues and sixty special issues

Álvaro Rocha¹, Maria José Sousa²

amr@iseg.ulisboa.pt; maria.jose.sousa@iscte-iul.pt

¹ ADVANCE, ISEG, Universidade de Lisboa, Rua Miguel Lupi, n.º 20, 1249-078 Lisboa, Portugal.

² Departamento de Ciência Política e Políticas Públicas, ISCTE-Instituto Universitário de Lisboa, Lisboa, Avenida das Forças Armadas. 1649-026, Lisboa, Portugal.

DOI: 10.17013/risti.50.1-4

1. Introdução

Ao publicar este número regular de junho de 2023, a RISTI – Revista Ibérica de Sistemas e Tecnologias de Informação – completa quinze anos de existência, com a publicação ininterrupta de números regulares e ainda de números não regulares através da publicação de números especiais. Os números regulares começaram por ser publicados semestralmente a partir de junho de 2008, tendo passado em 2014 para uma periodicidade trimestral. No computo geral, são cinquenta números regulares publicados, acrescidos de sessenta números especiais, ou seja, cento e dez edições já publicadas em quinze anos de vida da revista. Estes brilhantes números deixam-nos muito orgulhosos e esperançados para o futuro da RISTI, porque não é fácil manter uma dinâmica de longevidade e continuidade, face às adversidades que vão emergindo do meio ambiente, por vezes de fontes e origens inesperadas.

Este bonito e bem-sucedido percurso da RISTI somente foi possível devido ao esforço, resiliência e colaboração de muitos académicos ibero-americanos do domínio dos sistemas e tecnologias de informação, desde o fundador e diretor da revista, até aos

autores de artigos e de editoriais, passando por todo o conselho editorial, que avalia minuciosamente cada um dos artigos submetidos pelos autores, sem esquecer os leitores finais dos artigos publicados na revista.

Neste número histórico da RISTI foram submetidos pelos autores um total de oitenta e quatro artigos científicos para avaliação pelo conselho editorial, tendo sido selecionados para publicação os oito considerados como melhores, o que corresponde a uma apertadíssima e exigentíssima taxa de aceitação de 9,5%.

Que venham mais quinze anos de publicação ininterrupta da RISTI!

2. Estrutura

Este número da RISTI apresenta os seguintes artigos:

O primeiro artigo “A systematic review of dengue classification models using machine learning” revê 33 estudos científicos utilizando a metodologia PRISMA. Os métodos de aprendizagem de máquina mais usados são redes neurais e máquina vetorial de suporte. Os investigadores utilizam frequentemente variáveis climáticas e demográficas para melhorar a precisão. Predominam redes neurais, recorrentes, profundas e multicamadas, bem como métodos de séries temporais como LSTM e ARIMA para prever surtos de dengue. Os resultados obtidos são que os algoritmos de machine learning mais utilizados são as redes neurais (NN) e a máquina vetorial de suporte (SVM). A tendência é incluir variáveis climáticas e demográficas nos modelos de previsão.

O segundo artigo “Influência dos indicadores relacionados com a administração pública em linha no Índice de Transparência Municipal (ITM) de Portugal” investiga o impacto dos indicadores da administração pública em linha na transparência municipal em Portugal, examinando 15 indicadores. Através de análises estatísticas, determina-se a sua capacidade de explicar o Índice de Transparência Municipal (ITM) e identificam-se semelhanças entre municípios e indicadores. Os resultados revelam que alguns indicadores têm um poder explicativo significativo do ITM, destacando-se aqueles que são mais influentes. Isto demonstra a relação entre a administração pública em linha e a transparência, fornecendo informações sobre a melhoria da transparência através da administração pública em linha nos municípios portugueses.

O terceiro artigo “Efeitos da realidade aumentada na aprendizagem das ciências naturais: uma revisão sistemática da literatura” analisa o impacto da realidade aumentada na aprendizagem de ciências em alunos do 3.º ciclo e do ensino secundário. Utilizando a metodologia PRISMA, foram selecionados 10 artigos entre 2010 e 2022 de diversas bases de dados. A maioria dos estudos (8) utilizou métodos experimentais. Os resultados sugerem que a realidade aumentada teve efeitos positivos na aprendizagem, motivação, competências visual e espaciais e envolvimento dos alunos nas tarefas de aprendizagem.

O quarto artigo “Automatic Programming Task Grading Systems: Review and Critique” analisa sistemas e abordagens de correção automatizada de tarefas de programação, fornecendo uma visão geral de suas funções e desafios. Tendo sido desenvolvidos sistemas automáticos de feedback e classificação, sendo que os resultados apontam

para diferenças notáveis entre a academia e a indústria neste contexto, além da falta de estruturas de avaliação padronizadas para essas ferramentas.

O quinto artigo “Modelo de tendência para séries complexas: estudo de caso de tarifas de hotel” apresenta-se uma combinação de teoria da complexidade e suavização exponencial que é usada para entender os padrões nas tarifas de hotel. As séries de dados são divididas em partições e cinco métodos de plotagem são aplicados. O gráfico de comprimento R mostra anti-persistência com um ciclo de comprimento $n=8$, utilizado no modelo aditivo Holt-Winters com parâmetros ($a=0,36$, $b=0,78$, $y=0,67$), minimizando o erro. Esta combinação de métodos permite um ajustamento sazonal eficaz dos dados tarifários.

O sexto artigo “Divergência nos níveis de maturidade do governo eletrônico no México” tem dois objetivos principais: primeiro, demonstrar que os sites de governo eletrônico no mesmo nível de maturidade podem apresentar diferenças significativas; em segundo lugar, propor categorias derivadas de um nível de maturidade do governo eletrônico para avaliar com precisão os sites de governo eletrônico de nível municipal no México. Utilizando uma abordagem descritiva, foram analisados os 2470 municípios das 32 entidades mexicanas. Os resultados revelam divergências notáveis entre sites que estão no nível interativo, sugerindo dois tipos de sites interativos, alguns com funções assíncronas e outros síncronos, implicando um nível básico e avançado de interação. Este estudo contribui para a compreensão das etapas da administração pública em linha.

O sétimo artigo “El efecto de un programa de psicoterapia de telesalud en la salud mental de jóvenes en alto riesgo psicossocial” apresenta um estudo experimental que avaliou o impacto da psicoterapia de telessaúde na saúde mental de jovens em risco utilizando o Goldberg General Health Questionnaire. Os participantes foram divididos em um grupo intervenção que recebeu oficinas síncronas e um grupo controle sem intervenção. Antes da intervenção, 67% do grupo controle e 93% do grupo experimental apresentavam sintomas psicopatológicos. Após a intervenção, estes valores diminuíram para 73% e 27%, respetivamente. A diferença não foi significativa no grupo controle, mas foi significativa no grupo experimental, sugerindo que a psicoterapia de telessaúde liderada por enfermeiros melhora significativamente a saúde mental de jovens em risco.

Finalmente, o oitavo e último artigo, intitulado “Sistemas conversacionais: aspetos culturais influenciam a avaliação UX de um protótipo conceitual de assistente de voz” examina a experiência de 100 participantes, divididos igualmente entre brasileiros e portugueses, utilizando um sistema de conversação com assistente de voz automatizado. A avaliação da experiência do usuário (UX) incluiu estatística descritiva, análise da comunicação verbal dos utilizadores e o uso da escala SUS. A abordagem metodológica revelou que o protótipo conceitual tem potencial para simplificar a rotina dos profissionais de estética e beleza, embora algumas funcionalidades, como a nomenclatura dos botões e a acessibilidade, precisem de melhorias. Além disso, foi destacado que as opiniões variam entre brasileiros e portugueses, com a preferência destes últimos pela solução.

3. Agradecimentos

Esta introdução termina agradecendo a todos os autores e membros do conselho editorial que participaram no processo de revisão dos artigos que compõem esta edição, desejando que este número da RISTI seja mais um elemento de avanço do conhecimento e um importante contributo ibero-americano para o campo dos sistemas e tecnologias de informação, que estão a redesenhar e a expandir cada vez mais a economia digital.

Um agradecimento especial às associações científicas AISTI e ITMA, promotoras da RISTI, assim como às Bases de Dados de Revistas Académicas como Academic Journals Database, Dialnet, DOAJ, DOI, EBSCO, GALE, Google Scholar, IndexCopernicus, Information Systems Journal, Latindex, ProQuest, QUALIS, SciELO, SCImago, Scopus, SIS, Ulrich's e Web of Science, entidades que contribuem para que a RISTI seja uma revista científica de referência internacional.

Una revisión sistemática de Modelos de clasificación de dengue utilizando machine learning

Gisella Luisa Elena Maquen-Niño¹, Jessie Bravo², Roger Alarcón³,
Ivan Adrianzen-Olano⁴, Hugo Vega-Huerta⁵

gmaquenn@unprg.edu.pe; jbravo@unprg.edu.pe; ralarcong@unprg.edu.pe;
ivan.adrianzen@untrm.edu.pe; hvegah@unmsm.edu.pe

^{1,2,3} Grupo de Investigación en transformación digital, Universidad Nacional Pedro Ruiz Gallo, Calle Juan XXIII 391, C.P. 14001, Lambayeque, Perú

⁴ Universidad Nacional Toribio Rodríguez de Mendoza, Calle Higos Urco N° 342, C.P. 01000, Chachapoyas, Perú

⁵ Universidad Nacional Mayor de San Marcos UNMSM, Lima, Perú

DOI: 10.17013/risti.50.5-27

Resumen: El dengue es una enfermedad arboviral que anualmente reporta un gran número de infectados en la costa norte y la selva peruana. Según las estadísticas, está aumentando cada año. Este artículo tiene como objetivo desarrollar una revisión sistemática de la literatura científica sobre las variables de estudio y los métodos de aprendizaje automático utilizados actualmente para detectar la infección por dengue. La metodología utilizada fue PRISMA, mapeando inicialmente la literatura de 274 artículos científicos, quedando seleccionados 33 artículos para la revisión sistemática. Los resultados obtenidos son que los algoritmos de aprendizaje automático más utilizados son las redes neuronales (NN) y support vector machine (SVM). Asimismo, se ha encontrado que los científicos tienden a realizar investigaciones con variables climáticas o demográficas para obtener mejores resultados. Se concluye que los métodos de aprendizaje automático que más se han utilizado son las redes neuronales de diferentes tipos: convolucional, recurrente, profunda y multicapa, y para la predicción de brotes de dengue predominaron los métodos de series de tiempo con LSTM y ARIMA, también se estableció que la tendencia es hacia la inclusión de variables climáticas y demográficas en los modelos de predicción.

Palabras-clave: Dengue, detección, machine learning, métodos de clasificación, algoritmos de clasificación, random forest, support vector machine, artificial neural networks

A systematic review of dengue classification models using machine learning

Abstract: Dengue is an arboviral disease that annually reports a large number of infected on the north coast and the Peruvian jungle. According to statistics, it is increasing yearly. This article aims to develop a systematic review of the scientific

literature on the study variables and the machine learning methods currently used for detecting dengue infection. The methodology used was PRISMA, initially mapping the literature of 274 scientific articles, leaving 33 articles selected for the systematic review. The results obtained are that the most used machine learning algorithms are neural networks (NN) and support vector machine (SVM). Likewise, it has been found that scientists tend to carry out research with climatic or demographic variables to obtain better results. It is concluded that the machine learning methods that have been used the most are neural networks of different types: convolutional, recurrent, deep, and multilayer, and for the prediction of dengue outbreaks the time series methods with LSTM and ARIMA were the predominant ones, it was also established that the trend is towards the inclusion of climatic and demographic variables in the prediction models.

Keywords: Dengue, detection, machine learning, classification methods, classification algorithms, random forest, support vector machine, artificial neural networks

1. Introducción

La inteligencia artificial ha brindado la generación de modelos y algoritmos que permiten predecir el resultado de una variable de estudio, de acuerdo a patrones encontrados en un conjunto de datos, en diferentes sectores de la actividad humana (Bravo et al., 2023; Cubas & Maquen-Niño, 2022; Maquen-Niño et al., 2022).

Recientemente, machine learning ha permitido mejorar la predicción de que un individuo tenga una enfermedad o padecimiento, en menor tiempo y con niveles cada vez más aceptables de acierto (Maquen-Niño et al., 2023). Es en este nuevo contexto, que cada vez son más los experimentos que se hacen en cuanto a detección de enfermedades arbovirósicas como es el dengue, ya que, al ser una enfermedad propagada por el mosquito *Aedes aegypti* y, en menor medida, *Aedes Albopictus*, de clima tropical, afecta a diferentes países en el mundo, y es declarada una enfermedad endémica en ciertas regiones que cíclicamente aparecen con fuertes rebrotes especialmente en épocas de lluvia intensa. Ecuador al igual que Perú por tener regiones de clima tropical sufren de rebrotes de dengue en ciertas épocas del año, así como en épocas del fenómeno del niño, lo que hace indispensable que haya un control epidemiológico de las regiones afectadas por el dengue (Pérez et al., 2022).

En este aspecto, la Organización mundial de la salud (OMS) expresa que la cantidad de contagios de dengue, o también conocido como Dengue Fever (DF) o Dengue hemorrhagic fever (DHF), en el mundo ha aumentado, y actualmente se estima que hay entre 100 y 400 millones de infecciones cada año, aunque solo el 80% de ellas son generalmente leves y asintomáticas, hay un porcentaje que pone la vida en riesgo (OMS, 2022). La inteligencia artificial y el internet de las cosas se han convertido en grandes aliados en el sistema de salud, como lo demuestra (Kamesh & Sivakumar, 2020), a través de una aplicación móvil que utiliza el algoritmo SVM, también para (Ho et al., 2020), utilizaron los datos médicos de pacientes confirmados con dengue y aplicaron modelos de predicción de la red neuronal profunda (DNN) de última generación, además árbol de decisión de convocatoria (DT) y los modelos de regresión logística (LR).

En otro trabajo de investigación de (Khan Dourjoy et al., 2021) se han utilizado support vector machine (SVM) y el algoritmo de clasificador de bosques aleatorios para la predicción del dengue, en cambio en (Siddiq et al., 2021), se desarrollaron y se comparó modelos lineales (regresión lineal) y tres modelos no lineales (regresión de vectores de soporte, regresión de árboles de decisión y regresión del bosque aleatorio) para determinar el modelo con la mayor precisión en la predicción de la transmisión de DF.

Cabe indicar que hay trabajos de investigación que se centran en los datos sintomatológicos de los casos de dengue, pero hay varios trabajos actualmente que se están centrando en los datos climatológicos como tenemos los trabajos de (Bokonda et al., 2021; McGough et al., 2021; Nguyen et al., 2022; Rahman et al., 2021; Xu et al., 2020; Yavari Nejad & Varathan, 2021), quienes han utilizado dichos datos para definir modelos de predicción mucho más precisos utilizando diversos algoritmos de machine learning. Se han aplicado estudios en diferentes lugares, como en Bangladesh donde el dengue ha aumentado, el trabajo de investigación de (Sarma et al., 2020), trabaja con los datos del paciente, que contiene información sobre el informe de diagnóstico del paciente, el historial médico y los síntomas, aplicaron algoritmos de aprendizaje automático, como: árbol de decisión (DT) y Random Forest (RF) en el modelo de clasificación propuesto, a esto se suma el trabajo de (Sukama et al., 2020), en esta investigación, aplican la red neuronal recurrente (RNN), para predecir incidentes de DHF en DKI Yakarta mediante el uso de datos históricos de casos DHF de 2009 a 2017. Además la investigación de (Sukama et al., 2020), quienes aplicaron la red neuronal recurrente (RNN), para predecir incidentes de DHF en DKI Yakarta mediante el uso de datos históricos de casos DHF de 2009 a 2017, otro caso fue el realizado por (Huang et al., 2020), quienes trabajaron con los datos de casos de dengue de un Hospital en Taiwan, entre julio a noviembre del 2015.

A nivel Latinoamérica, tenemos algunos trabajos en Brasil, como el de McGough et al. (2021), quienes tomaron información de los casos de dengue ocurridos en Brasil entre los años 2001 a 2007 y los datos climatológicos en los años 2000 a 2016 para predecir con mayor precisión esta enfermedad, se suma el trabajo de (Bauxell et al., 2021), quienes presentan un algoritmo para la predicción del riesgo de dengue en Brasil, se basan en técnicas de aprendizaje automático y utilizan variables ambientales y socioeconómicas para predecir el riesgo de propagación del dengue en todo Brasil. Otros estudios en Brasil son los realizados por (Aleixo et al., 2022; Koplewitz et al., 2022; Souza et al., 2022).

En este sentido, los hallazgos anteriores llevaron a la elaboración de las preguntas de investigación: ¿Cuáles son las variables de estudio en la detección del contagio por dengue?, ¿Cuáles son los modelos y algoritmos de machine learning usados? y ¿Cuáles son los rangos de precisión encontrados?

2. Métodos y materiales

La revisión sistemática fue desarrollada siguiendo la metodología PRISMA. Se utilizó las bases de datos indexadas: Scopus, Web of Science, IEEE Xplore y Google académico porque albergan artículos científicos que siguen criterios rigurosos aceptados por la comunidad científica, así como publicaciones de diferentes países que están a la

vanguardia en temas de machine learning orientados a la detección del dengue. Para realizar esta investigación, se contó como materia prima artículos científicos en inglés que fueron encontrados aplicando ecuaciones de búsqueda específica y sus sinónimos, con el uso de operadores booleanos AND y OR. Para el concepto: “Classification” se tomaron en cuenta los sinónimos: Classification models, prediction y para el concepto “Dengue” se tomaron en cuenta los sinónimos: Dengue fever, DENV. Por tanto, la ecuación de búsqueda quedo planteada en idioma inglés conforme se visualiza en la Tabla 1.

Idioma	Ecuación de búsqueda
English	(classification OR prediction OR “Classification models”) AND “machine learning” AND (dengue OR “dengue fever” OR DENV)

Tabla 1 – Ecuación de búsqueda

Se aplicó como filtro de búsqueda que los años de publicación de los artículos sean entre el año 2014 hasta el año 2022. De 274 publicaciones encontradas en ese rango de años, 184 pasaron la etapa de cribado, 96 cumplieron los criterios de elegibilidad y en la etapa final 33 cumplieron con los criterios de inclusión, como se observa en la Tabla 2. Los criterios de cribado corresponden a dos aspectos: la pertinencia y la accesibilidad. Por pertinencia se entiende que, en base a la lectura del resumen y conclusiones, los investigadores determinaron la pertinencia o no del artículo dentro de la presente investigación. Por accesibilidad se entiende que los archivos son de acceso abierto y descargables.

Los principales criterios de elegibilidad son los siguientes:

1. Los objetivos de la investigación están claramente especificados y se relacionan con la clasificación del dengue usando Machine Learning.
2. La investigación contiene aportes teóricos sobre las tendencias de la clasificación del dengue usando Machine Learning
3. La investigación define y explican claramente las variables de estudio para la detección del dengue.
4. Los modelos y algoritmos de clasificación de dengue obtenidos en la investigación son pertinentes para conseguir sus objetivos
5. El score obtenido en cada modelo de machine learning es mayor a 0.7

Database	Step 1: Identificación	Step 1: Cribado	Step 2: Eligibilidad	Step 3: Inclusión
Scopus	151	129	75	27
Web of science	76	35	10	1
IEEE Xplore	43	16	10	4
Google Scholar	4	4	1	1
Total	274	184	96	33

Tabla 2 – Bases de datos y fases desarrolladas para obtener información

La aplicación de criterios identificación, cribado, elegibilidad e inclusión puede verse en el diagrama de flujo PRISMA, detallado en la Figura 1.

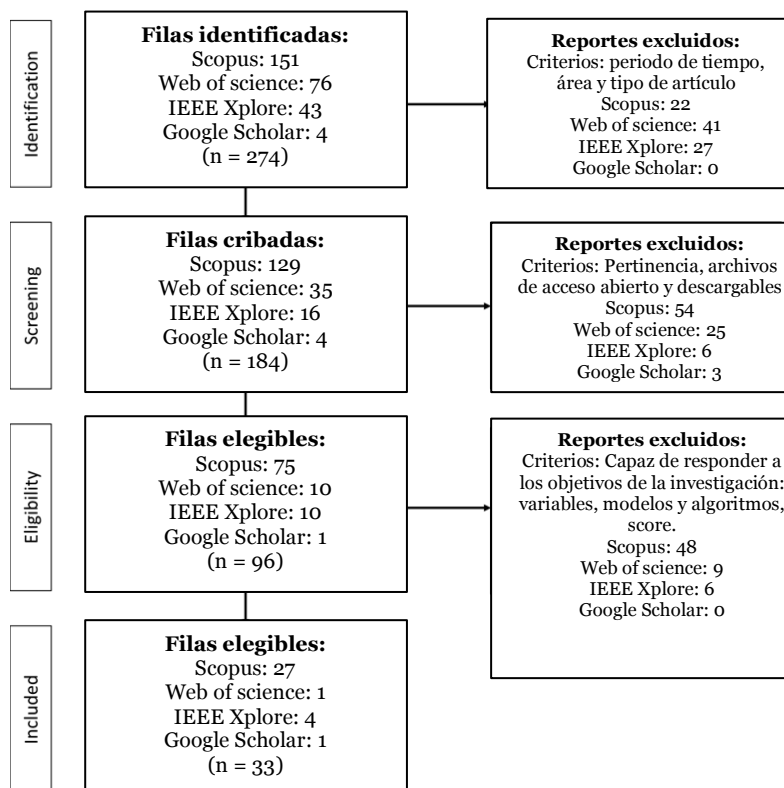


Figura 1 – Flujo de trabajo de PRISMA para la selección de los estudios relevantes

3. Resultados

Para realizar el análisis bibliométrico se utilizó la herramienta VOS-Viewer de los 33 artículos elegidos tomando como análisis las co-ocurrencias de todas las palabras claves.

Los resultados obtenidos del análisis bibliométrico realizado de la búsqueda inicial nos muestran tres cluster de las co-ocurrencias de todos los keywords de los artículos encontrados, como se observa en la Figura 2.

Del análisis de los cuatro clústeres encontrados, el clúster de color verde consta de 10 ítems y se centra en las técnicas para la predicción del dengue: machine learning, Decision Tree, Random Forest, time series analysis, prediction, forecasting. Otros conceptos relacionados que tienen gran nivel de ocurrencia son: dengue, aedes aegypti, virus infection, tropical countries.

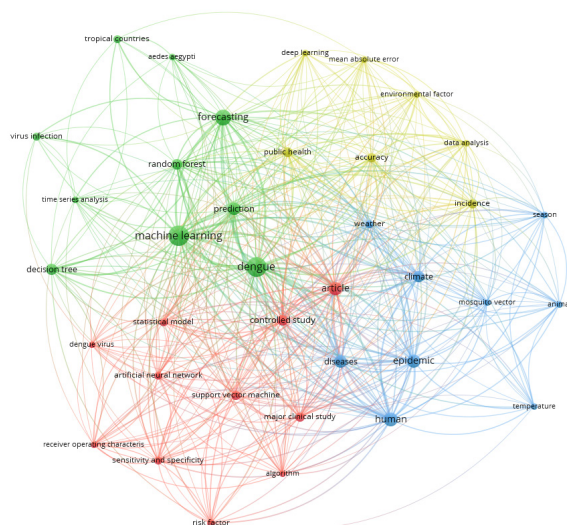


Figura 2 – Análisis de co-ocurrencias de palabras clave

El segundo clúster, es el clúster amarillo que consta de 7 ítems, centrando su atención en las medidas de evaluación del modelo como son: accuracy, mean absolute error, incidence, data analysis. Otros conceptos relacionados son: environmental factor, public health y deep learning.

El tercer clúster, consta de 11 ítems, es el clúster rojo, que hace referencia principalmente a modelos: statistical model, support vector machine, artificial neural networks, algorithm. Otros conceptos relacionados son: dengue virus, controled study, risk factor, receiver operating characterisc, major clinical study, article, sensitivity and especifity.

El cuarto clúster, es el clúster azul que consta de 9 ítems, que hace referencia a los atributos de las variables utilizadas para la predicción del dengue: weather, climate, season, temperature, epidemic, mosquito vector, diseases. También encontramos otros conceptos que tienen gran ocurrencia en los artículos estudiados como son: animal y human.

Un total de 33 artículos científicos fueron seleccionados para responder a las preguntas de investigación formuladas. Por cada artículo, se detalló la base de datos en la cual fue encontrado, el autor y el año de publicación, el país del cual la evidencia científica fue recolectada, sus palabras clave, así como las categorías o dimensiones investigadas. Los resultados fueron sistematizados en la Tabla 3, siendo las categorías denotadas por los siguientes números: Epidemiológica (Categoría 1), Clima (Categoría 2), Síntomas (Categoría 3), Demográficas (Categoría 4), Análisis de laboratorio Clínico (Categoría 5), socioeconómica (Categoría 6) y geográfica (Categoría 7).

DB	No.	Autor(es)	País de los datos	Título de artículo	Cate-goría
S C O P U S	1	(Appice et al., 2020)	Mexico	A Multi-Stage Machine Learning Approach to Predict Dengue Incidence: A Case Study in Mexico	1, 2
	2	(Srivastava et al., 2020)	India	An online learning approach for dengue fever classification	3
	3	(Huang et al., 2020)	Taiwan	Assessing the risk of dengue severity using demographic information and laboratory test results with machine learning	4, 5
	4	(Sukama et al., 2020)	DKI Jakarta	Comparing activation functions in predicting dengue hemorrhagic fever cases in DKI Jakarta using recurrent neural networks	3
	5	(Ho et al., 2020)	Taiwan	Comparing machine learning with case-control models to identify confirmed dengue cases	4, 5, 3
	6	(Sarma et al., 2020)	Bangladesh	Dengue Prediction using Machine Learning Algorithms	4, 5, 3
	7	(Xu et al., 2020)	China	Forecast of dengue cases in 20 Chinese cities based on the deep learning method	1, 2
	8	(Balamurugan et al., 2020)	India	Improved prediction of dengue outbreak using combinatorial feature selector and classifier based on the entropy-weighted score based optimal ranking	5
	9	(Raizada et al., 2020)	India	Vector-borne disease outbreak prediction by machine learning	1, 2
	10	(Benedum et al., 2020)	Iquitos, Puerto Rico y Singapore	Weekly dengue forecasts in Iquitos, Peru; San Juan Puerto Rico; and Singapore	1, 2
	11	(Khan Dourjoy et al., 2021)	Bangladesh – South Asia	A Comparative Study on Prediction of Dengue Fever Using Machine Learning Algorithm	3
	12	(McGough et al., 2021)	Brazil	A dynamic, ensemble learning approach to forecast dengue fever epidemic years in Brazil using weather and population susceptibility cycles	1, 2
	13	(Kaur et al., 2021)	India	A Machine Learning Approach for Predicting Dengue Outbreak	1, 2
	14	(Ferdousi et al., 2021)	Brazil	A Windowed Correlation-Based Feature Selection Method to Improve Time Series Prediction of Dengue Fever Cases	1, 2

DB	No.	Autor(es)	País de los datos	Título de artículo	Cate-goría
S C O P U S	15	(Yavari Nejad & Varathan, 2021)	Malaysia	Identification of significant climatic risk factors and machine learning models in dengue outbreak prediction	1, 2
	16	(Bauxell et al., 2021)	Brazil	Machine learning techniques using environmental data from remote sensing applied to modeling dengue risk in Brazil	1, 6, 2
	17	(Rahman et al., 2021)	Thailand	Mapping the spatial distribution of the dengue vector Aedes aegypti and predicting its abundance in northeastern Thailand using machine-learning approach	6, 2, 3
	18	(Choubey et al., 2021)	San Jose e Iquitos	Soft computing techniques for dengue prediction	2
	19	(Tanawi et al., 2021)	DKI Jakarta	Support Vector Regression for Predicting the Number of Dengue Incidents in DKI Jakarta	2
	20	(Nguyen et al., 2022)	Vietnam	Deep learning models for forecasting dengue fever based on 2 data in Vietnam	1
	21	(Abdualgalil et al., 2022)	Yemen	Early Diagnosis for Dengue Disease Prediction Using Efficient Machine Learning Techniques Based on Clinical Data	4, 3, 5
	22	(Ochida et al., 2022)	New Caledonia	Modeling present and future climate risks of dengue outbreak, a case study in New Caledonia	2
	23	(Koplewitz et al., 2022)	Brazil	Predicting dengue incidence leveraging internet-based data sources. A case study in 20 cities in Brazil	1, 7, 2
	24	(Souza et al., 2022)	Brazil	Predicting dengue outbreaks in Brazil with manifold learning on 2 data	2
	25	(Aleixo et al., 2022)	Brazil	Predicting Dengue Outbreaks with Explainable Machine Learning	1, 4, 2
	26	(Kumar et al., 2022)	Bangladesh	Prediction of dengue incidents using hospitalized patients, meteorological and socioeconomic data in Bangladesh: A machine learning approach	1, 2
	27	(Nalini et al., 2022)	San Juan e Iquitos	Prediction of dengue infection using Machine Learning	2

DB	No.	Autor(es)	País de los datos	Título de artículo	Cate-goría
WoS	28	(Salim et al., 2021)	Malaysia	Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques	1, 2
IEEEExplore	29	(Baker et al., 2021)	San Juan e Iquitos	Forecasting Dengue Fever Using Machine Learning Regression Techniques	3, 2
	30	(Singh, 2021)	India	Particle Swarm Optimization assisted Support Vector Machine based Diagnostic System for Dengue prediction at the early stage	3
	31	(Siddiq et al., 2021)	Saudi Arabia	Predicting Dengue Fever Transmission Using Machine Learning Methods	1, 2
	32	(Rana et al., 2022)	Kaggle (internet repository)	Dengue Fever Prediction using Machine Learning Analytics	3
Google	33	(Chattopadhyay & Chattopadhyay, 2022)	India	VIRDOC: A VIRTUAL DOCTOR to predict dengue fatality.	3

Tabla 3 – Base de datos, lista de autores, país, título y categorías exploradas por artículo.

Del total de 33 artículos seleccionados, algunos fueron de Brasil (6), India (6), Bangladesh (3), Malaysia (2), Taiwan (2), DKI Jakarta (2), San Juan e Iquitos (2), San José e Iquitos (1), Iquitos, Puerto Rico y Singapore (1), Arabia Saudita (1), Vietnam (1), Thailandia (1), Yemen (1), New Caledonia (1), México (1), China (1) y Kaggle (1). Por otro lado, en las categorías exploradas por los autores, tenemos clima (21), epidemiológica (15), síntomas (11), análisis de laboratorio clínico (5), demográficas (5), socioeconómica (2) y geográfica (1) correspondientes a las variables de estudio, conforme se visualiza en la Figura 3.

Un análisis más profundo de cada Artículo científico permitió identificar aspectos relacionados a las variables/características seleccionadas, fuente de datos, algoritmo / método de machine learning y, método de validación en cada artículo (ver Tabla 4).

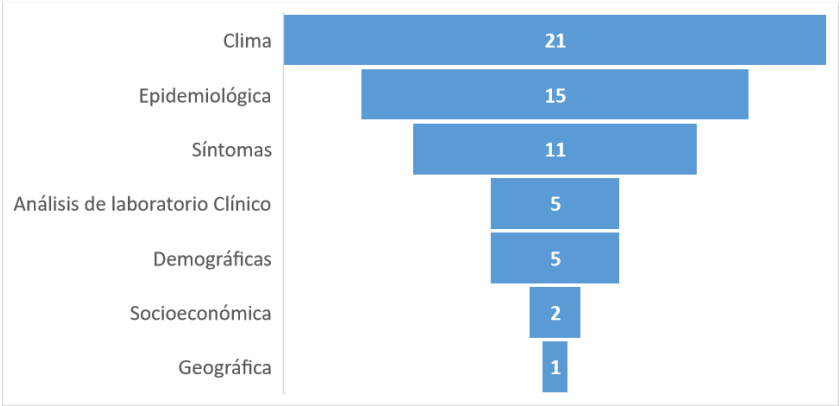


Figura 3 – Número de artículos por categoría

4. Discusión de resultados

La revisión sistemática permitió seleccionar un total de 33 artículos, muchos de los cuales son de Brasil, India, Bangladesh e Iquitos. La categoría más estudiada por los autores es la de Clima (21), seguido por la categoría epidemiológica (15) y la categoría síntomas (11), demostrando que existe una gran tendencia a los estudios con variables de factores climáticos que ayudan a mejorar la predicción de brotes de dengue y la detección de la enfermedad de acuerdo con la ubicación geográfica del paciente.

Esto debido a que el vector mosquito que produce la enfermedad, es de clima tropical y subtropical, lo cual está asociado con variables climatológicas como la temperatura ambiental, la humedad y sobre todo el nivel de precipitación. De esa forma, si es que hay temporadas de lluvias, el agua muchas veces se queda empozada y es criadero de zancudos propagadores del dengue.

Otra de las categorías que también es predominante en los estudios y ocupa el segundo lugar, es la de factores epidemiológicos, que está ligado altamente al control sanitario de cada país en las provincias donde el dengue es una enfermedad endémica, que tiene que ver con el registro de contagiados, hospitalizados y muertos que reportan cada cierto tiempo los hospitales y centros de salud, y que para ayudar a los modelos de machine learning deberían tener la misma distribución temporal, sea diaria, semanal o mensual, pero que no haya periodos en los que no hubo reporte de datos, ya que esto afecta la continuidad de los datos y para reemplazar esos faltantes se tienen que realizar estimaciones a través de métodos predictivos.

La categoría Síntomas que es la que ocupa el tercer lugar, se caracteriza por tener investigaciones que han utilizado datos descargados de repositorios oficiales de un determinado país y en su minoría base de datos construidas por los investigadores de datos obtenidos de clínicas y hospitales. Entre los síntomas que tienen mayor incidencia

Nº	Variables/ características	Categoría	Fuente de datos	Algoritmo/ método	Método de validación
1	Dos variables: Número de casos de dengue y dengue hemorrágico. Temperatura media del aire (en °C) a nivel mensual por cada uno de los 32 estados	1, 2	Serie de Tiempo de ambas variables recolectados durante el período 1985–2010 en 32 estados federales de México. Training (1985–2009) y test (2010)	Time Series Forecasting: AutoRegressive Integrated Moving Average (auto.arima) y Vector Autoregression (VAR). Regression Methods (M5'), Support Vector Regression (SVR), k-Nearest Neighbor (kNN) y la estrategia de aprendizaje propuesta de múltiples etapas llamada AutoTiC-NN. AutoTiC-NN demostró el menor RMSE en la mayor cantidad de estados evaluados.	Root Mean Square Error (RMSE).
2	14 variables correspondientes a Síntomas de pacientes e investigaciones de laboratorio de los pacientes	3	182 muestras de pacientes, 148 positivos y 34 negativos. (training 80% y test 20%)	Support Vector Machine (SVM), Random Forest (RF), así como múltiples clasificadores de Weka: Logistic, SGD, SGD text, Simple Logistic, SMO, Voted perceptron, Adaboost, Random subspace, Bayes Net, Naive Bayes (NB) y LWL. Clasificadores en Línea: aROW, PA2, ALMA, a JELLIP, CW, POE, aROMA, ROMA. (RF obtuvo 82.41%)	F - MEASURE
3	Variables demográficas y análisis de laboratorio clínicos	4, 5	798 casos (training set 90%, test set 10%)	Logistic Regression, SVM, RF, gradient boosting machine (GBM), Artificial Neural Network (ANN) (ANN: ROC: 0.8324 ± 0.0268, Accuracy: (0.7523 ± 0.0273))	10 - fold cross validation approach, ROC and Accuracy
4	Incidencias de Dengue hemorrágico	3	Data de Dengue hemorrágico (DHF) de 418 semanas de 5 municipalidades de Jakarta entre el 2009 y 2017 (Training 70% y test 30%)	Recurrent Neural Network (RNN) y Backpropagation Through Time (BPTT). Los resultados de la predicción de DHF en las cinco municipalidades de Jakarta obtuvieron mejores resultados en el RMSE y RME con la función de activación Sigmoid.	RMSE and MSE (The Mean Squared Error).
5	18 variables, siendo: demográficas, clínicas (signos vitales y comorbilidades), de análisis de laboratorio.	4, 5, 3	4894 casos de atención por emergencia en un Hospital	Logistic Regression, Deep Neural Network, Decision Tree (DT) (DNN fue el mejor con 86.95%±0.45%)	10-fold cross validation process

N°	Variables/ características	Categoría	Fuente de datos	Algoritmo/ método	Método de validación
6	23 atributos: demográficos, síntomas y análisis de laboratorio	4, 5, 3	101 000 casos de Bangladesh de dos ciudades principales densamente pobladas (Dhaka y Chittagong)	Sistema clasificador binario (0,84)	ROC curve
7	Número de casos de dengue humano por mes (variable epidemiológica) y 9 variables meteorológicas mensuales.	1, 2	Datos meteorológicos y Dato epidemiológico mensual del año 2005 al 2018 de 20 ciudades de China (Training data 2005- al 2016, Testing Data 2017-2018)	Redes neuronales recurrentes de memoria a corto plazo (LSTM) y Redes neuronales recurrentes de memoria a corto plazo con aprendizaje de transferencia (LSTM - TL) y su comparación con otros modelos: BPNN; Backpagation Neural Network; GAM; Generalized Additive Model; SVR; Support Vector Regression; GBM; Gradient Boosting Machine (LSTM 0.9087 /1.1600)	RMSE (The Root Mean Squared Error), RRSE (The Root Relative Squared Error)
8	20 variables correspondientes a los atributos de un análisis de laboratorio de sangre	5	480 muestras de pacientes con dengue	EWSORA Algoritmo de clasificación óptima basado en puntaje ponderado de entropía. Genetic Algorithm (GA), Particle Swarm Algorithm (PSO), Correlation based Feature Selector (CFS) Modelos de ML: J48, NB, SVM y Multilayer perceptron (MLP), regresión y series de tiempo (EWSORA con MLP con 98,72)	cross validation
9	Casos positivos del 2013 al 2017 para Chikungunya, malaria y dengue. También Variables meteorológicas: temperatura, humedad y precipitación.	1, 2	5415958 datos de casos positivos y datos meteorológicos de 36 estados del 2013 al 2017 fusionados. Tres posibles salidas: riesgo alto, moderado o bajo. (Training: 60%y Test 40%).	Convolutional Neural Network multimodal disease outbreak prediction CNN- MDOP y Artificial Neural Network multimodal disease outbreak prediction ANN - MDOP (CNN accuracy: 0.88)	Prediction accuracy
10	Recuentos de casos y brotes semanales de dengue en Iquitos, Perú; San Juan, Puerto Rico; y Singapur de 1990 a 2016. Variables meteorológicas de esos tres lugares (16 variables)	1, 2	Se crearon 500 conjuntos de entrenamiento que se usó para entrenar cada modelo y se promediaron las predicciones	Algoritmo ML predictivo: RF, RF con Algoritmo de señalización univariante (RF-UFA) Regresión de Poisson y regresión logística. Modelo de promedio móvil integrado autorregresivo (ARIMA) y modelo estacional ARIMA (SARIMA) (Modelos de ML - RF tenían un 21 % y un 33 % menos de error)	Cross Validation with fold =10 MAE

Nº	Variables/ características	Categoría	Fuente de datos	Algoritmo/ método	Método de validación
11	15 variables de sintomatología de Dengue, buscándose predicción para el próximo año	3	Encuestas online a más de 600 pacientes de Bangladesh y el método de topografía. Se han recopilado casi 1047 datos. Training 80 % y test 20%	SVM y el algoritmo clasificador de Random Forest (RFC), (SVM 0.695)	cross validation
12	Variables de pacientes con dengue y variables climáticas como: temperatura y precipitación diaria	1, 2	Datos sobre casos anuales de dengue (Ministerio de Salud de Brasil) para 2001–2017 y sobre temperatura y precipitación diaria (GMAO-NASA) para 2000–2016, para 20 municipios endémicos de dengue en Brasil	Contempla 11 modelos de aprendizaje ensamblado que usan SVM. En función de si es un año epidémico o no, el ‘strength’ de su modelo es del 63% o el 71% respectivamente.	Accuracy, Time- series
13	Variables meteorológicas y el conteo de casos de Dengue como variable epidemiológica.	1, 2	Datos obtenidos del Departamento Meteorológico de la India, Pune entre 2006-2015.	Hidden Markov Model (HMM) para datos de series de tiempo. La precisión obtenida para los cuatro distritos es 97% para Bathinda, 92% para Patiala, 97% para Ludhiana y 86% para Amritsar.	Accuracy
14	Casos de dengue y variables climáticas	1, 2	78 municipios en Brasil con casos 71348, 58424, 45319, 36743, 27103 Training 50%, validation 30% y test 20%	Recurrent neural networks (RNN), LSTM, GRU y el modelo Lineal. (GRU: total de casos MAE 0.13 y para prueba de varias ubicaciones MAE rango de 0.2435 a 0.2569)	Mean absolute error (MAE)
15	Variables sobre casos de Dengue (incidencias y casos confirmados) y variables climáticas como son Temperatura, humedad relativa y cantidad de lluvia.	1, 2	Datos obtenidos de dos fuentes oficiales. El incidente de dengue y los casos confirmados semanales del portal del Ministerio de Salud de Malasia (MOH) y, los datos climáticos se obtienen del Departamento Meteorológico de Malasia (MMD) en el periodo 2010 al 2013	Bayes Network models (BN), SVM, Radial Basis Function (RBF Tree), decision table (DT) y NB. (Bayes Networks con 92-35% de accuracy)	cross validation (ten times) with TempeRain factor (TRF)

N°	Variables/ características	Categoría	Fuente de datos	Algoritmo/ método	Método de validación
16	Variables clínicas de pacientes con dengue en Brasil. Variables ambientales y socioeconómicas	1, 6, 2	Datos de 2010 a 2013 de la distribución mensual de episodios de dengue del Sistema de Información de Enfermedades Notificables (SINAN), desarrollado por el Ministerio de Salud de Brasil. Datos ambientales de las misiones satelitales de observación de la Tierra (Training 80%, test 20%)	DT, RF (R2: 0.7635 y accuracy: 89,49% usando Index B) y neural networks (R2: 0.5041 and accuracy: 92,40% usando index B y un accuracy: 79.21% usando el index A). RF es el de mejor resultado.	R-squared (R2) and accuracy
17	Variables socioeconómicas, variables climáticas y síntomas de dengue	6, 2, 3	Tailandia de enero a diciembre del 2019	Modelos de aprendizaje supervisados, regresión logística (LR), SVM, vecino K-Nearest (KNN), red neuronal artificial (ANN) y RF. RF, que usó factores socioeconómicos, KAP y de paisaje como entradas, tuvo la mayor precisión del modelo de predicción, seguida por ANN, kNN, LR y SVM (socioeconómico: CA = 0,86 y F1 = 0,85; KAP: CA = 0,92 y F1 = 0,90 y paisaje: CA = 0,89 y F1 = 0,87)	AUC (área bajo la curva ROC), CA (precisión de clasificación) y puntaje F1
18	Variables climáticas cualitativas y cuantitativas	2	Datos de DengAI Data set y los datos climáticos de dos ciudades San José e Iquitos	XGBoost Regressor Model, Negative Binomial Regressor Model, DT Regressor Model, ARIMA Model. En base a los valores MAE y las compensaciones, el regresor binomial negativo (NBR) es el modelo con mejor rendimiento.	MAE, RMSE
19	Las variables meteorológicas, incluidos las variables sobre la lluvia, la humedad y la temperatura	2	Datos de Cinco regiones en DKI Yakarta desde el 6 de enero de 2009 hasta el 25 de septiembre de 2017, proporcionados por la Agencia de Meteorología, Climatología y Agencia Geofísica (BMKG)	Support Vector Regression (Cross-correlation, Augmented Dickey-Fuller test para calcular el lapso entre dos series de tiempo estacionarias). SVM con núcleo lineal dio errores relativamente pequeños en la predicción de incidentes de dengue (MAE < 8 y RMSE <10).	MAE, RMSE

Nº	Variables/ características	Categoría	Fuente de datos	Algoritmo/ método	Método de validación
20	12 factores meteorológicos en el mismo período, incluidas medidas de temperatura, lluvia, humedad, evaporación y horas de sol, casos y muertes por tasa mensual del dengue	1	Casos de dengue de 20 provincias/ciudades pertenecientes a tres regiones principales de Vietnam entre 1997 y 2016 del Instituto Nacional de Higiene y Epidemiología de Vietnam. Datos meteorológicos del Instituto de Meteorología, Hidrología y Cambio Climático de Vietnam. Training (1997 a 2013) y test (2014 a 2016) para provincia	Long short-term memory neural network based on the attention mechanism (LSTM- ATT), Long short-term memory neural network LSTM, Convolutional Neural Network (CNN), Transformer (TF) (LSTM-ATT, RMSE: 1,60 y MAE: 1,90)	MAE, RMSE
21	21 atributos: edad, sexo, fiebre, síntomas y variables clínicas.	4, 3, 5	Se obtuvo del Centro de Monitoreo Epidemiológico (EMC) - Oficina de Salud Pública y Población (PHPO) - Taiz City, Yemen del 2017 al 2019.	Técnicas eficientes de aprendizaje automático (EMLT): K-Nearest Neighbor (KNN), clasificador de impulso de gradiente (GBC), clasificador de árbol adicional (ETC), impulso de gradiente extremo (XGB) y máquina de impulso de gradiente de luz (LightGBM). El mejor modelo fue ETC con una accuracy del 99,03%, fi-score del 99,04%, precisión del 98,92% y recall del 99,17 %.	Accuracy, fi-score, recall, precision, AUC, and runtime
22	La lluvia diaria (RR) y la temperatura máxima (TX)	2	Los datos demográficos de Nueva Caledonia de los censos de 1969, 1976, 1983, 1989, 1996, 2004, 2009, 2014 and 2019. Los datos climáticos de una estación meteorológica de Météo- France de 1970 a 2020 en Faubourg Blanchot, Nouméa.	SVM. Se presenta un método para evaluar la probabilidad semanal de un brote de dengue para un lugar específico y proyectar este riesgo en un escenario de cambio climático.	Leave Time Out - Cross Validation (LTO - CV)

N°	Variables/ características	Categoría	Fuente de datos	Algoritmo/ método	Método de validación
23	Variables epidemiológicas, geográficos y climáticos	1, 7, 2	Incidenias semanales de años anteriores, variables climáticas semanales en Brasil y datos de búsqueda en Internet en tiempo real. Periodo de enero del 2010 a Julio del 2016.	RF y L1 – based (LASSO Regression). Se determinó que RF es más consistente y robusto que LASSO.	RMSE, relative RMSE, R2
24	Variables climáticas	2	Seis ciudades de Brasil	SVM. Random Guess (R-Guess) accuracy 0.8	Test Mean Accuracy
25	Variables epidemiológicas, meteorológicas, sociodemográficas	1, 4, 2	Datos del Sistema Nacional de Información sobre Enfermedades Notificables (SINAN), datos meteorológicos y datos sociodemográficos del Instituto Brasileño de Geografía y Estadísticas	Gradient boosting decision tree algorithm (CatBoost), accuracy 90%	MAE, RMSE, and R2
26	Casos de dengue del 2012 al 2019 y Variables climatológicos de 11 ciudades de Bangladesh	1, 2	11 distritos diferentes de Bangladesh	Modelo de aprendizaje automático, regresión lineal múltiple (MLR) y regresión de vectores de soporte (SVR). Support Vector Regression (SVR) 75% de precisión.	Precision, mean absolute error (MAE) of models
27	Temperatura máxima y mínima, temperatura promedio, precipitación total de la estación meteorológica climática GHCN de la NOAA.	2	Datos de las ciudades de San Juan e Iquitos	Regresión lineal, Series de tiempo, Heat Mapping, Moving mean. Obtuvo mejores resultados es Linear Regression con un MAE de 20.77	Mean absolute error (MAE)
28	Variables Epidemiológicas y Variables climáticas (temperatura, velocidad del viento, humedad y lluvia).	1, 2	Los datos contienen 5 años (2013 a 2017) de casos semanales para cinco distritos en Selangor, Malasia: Gombak, Hulu Selangor, Hulu Langat, Klang y Petaling.	Classification and Regression Tree (CART), Artificial Neural Network (MLP), SVM (LINEAR, POLYNOMIAL, RBF), and Bayes Network (TAN). SVM (núcleo lineal) exhibió el mejor rendimiento de predicción con una exactitud de 70 %.	Accuracy, sensitivity , and precision

Nº	Variables/ características	Categoría	Fuente de datos	Algoritmo/ método	Método de validación
29	Variables sintomatológicas y climatológicas (Location, Temperature, Precipitation, Humidity, and Vegetation)	3, 2	DrivenData Website de dos ciudades San Juan e Iquitos	Machine Learning Regression, regression neural network, RF, Negative binomial regression, Poisson Regression, Decision Tree Regression (DTR), NB, SVM, k-nearest neighbor (k-NN) Regression, AdaBoost, Bagging with decision tree. Obtuvo mejores resultados Recurrent Neural Network (RNN) con memoria a largo plazo (LSTM) con una precisión del 94%	Mean absolute error (MAE)
30	Variables sintomatológicas	3	Conjuntos de datos de clasificación de dengue estándar	J48, RF, DT y Algoritmo propuesto (PSO - SVM). Obtuvo mejores resultados el algoritmo propuesto con 87% de accuracy	MAE
31	Variables epidemiológicas y variables climatológicas: Temperatura y humedad	1, 2	Casos de DF reportados para Jeddah City, Arabia Saudita. La temperatura y humedad del Research Data Archive at the National Center for Atmospheric Research	Regresión lineal y tres modelos no lineales (regresión de vectores de soporte, regresión de árboles de decisión y regresión del bosque aleatorio). Obtuvo mejor resultados Support Vector Classification con 76 % de accuracy	Precision
32	Variables sintomatológicas	3	Conjunto de datos con 124 muestras y 8 características descargado de Kaggle	Análisis de aprendizaje automático (DFES- MLA) utiliza clasificadores de árbol de decisión (DT) y bosque aleatorio (RF). Obtuvo mejor resultado Random Forest con un accuracy de 0.97.	Accuracy.
33	Variables sintomatológicas	3	Conjuntos de datos de clasificación de dengue estándar	Regresión lineal múltiple (MLR), ANOVA. Obtuvo mejores resultados Multiple Linear Regression con 75% de accuracy.	accuracy and performance

Tabla 4 – Variables/características seleccionadas, categoría, fuente de datos, algoritmo / método de machine learning y método de validación en cada artículo.

en las investigaciones podemos mencionar: Fiebre, dolor de cabeza, dolor corporal, artralgias, mialgias, náuseas, mareo, hemorragia, entre otros.

A nivel mundial existe un consenso para clasificar los tipos de contagios por dengue, siendo tres: Dengue sin signos de alerta, Dengue con signos alerta y Dengue hemorrágico o grave, y siendo la última la más peligrosa para la salud, y aunque es el que se da en muy baja proporción, puede haber riesgo de muerte de no tratarse a tiempo, de las 33 investigaciones seleccionadas, 2 investigaciones han abordado detalladamente la detección del dengue hemorrágico.

Con respecto a las técnicas y algoritmos de machine learning se puede denotar que hay una fuerte tendencia al uso de redes neuronales (17 estudios), ya que, debido a su variedad de topologías, son muy adaptables en diferentes tipos de investigación tanto para clasificación, como para regresión. Dentro de los modelos de redes neuronales utilizados podemos citar: ANN (7 estudios), LSTM (3 estudios), MLP (2 estudios), CNN (2 estudios), RNN (1 estudio), BPNN (1 estudio) y Perceptrón (1 estudio). Otras de las técnicas de machine learning que tuvieron gran incidencia fueron: SVM (12 estudios) seguida por RF (8 estudios), DT (6 estudios), Naive Bayes (4 estudios) para la detección del contagio de la enfermedad. Otras investigaciones se enfocaron hacia la regresión de datos, siendo las de mayor incidencia Logistic regression (5 estudios), Support vector regression (5 estudios), Linear regression (4 estudios) y Machine learning regression (4 estudios). También se utilizó métodos de series de tiempo con el uso de LSTM (3 estudios) y ARIMA y sus variantes (4 estudios).

Considerando la precisión de los modelos para la detección del dengue, se encontró que 9 artículos tuvieron como accuracy un valor mayor o igual que 70% y menor de 80%, 6 artículos tuvieron un valor mayor o igual a 80% y menor de 90%, 11 artículos tuvieron un valor mayor o igual a 90% y menor o igual a 100%, presentándose también que 7 artículos no especificaban el valor de precisión.

5. Conclusiones

Se llega a la conclusión que los métodos de machine learning que son más utilizados en la detección del contagio de dengue son las redes neuronales siendo las más ampliamente utilizadas las redes neuronales artificiales convolucionales y perceptrón multicapa. Para lo que corresponde a predicción de brotes de dengue se utilizaron preferentemente los modelos de serie de tiempo con LSTM y los modelos ARIMA con sus variantes, estos últimos más ligados a modelos estadísticos.

Existe una fuerte tendencia hacia la inclusión de variables climáticas en los modelos de machine learning que ayuden a mejorar la predicción, debido a que el mosquito que lo transmite tiene un hábitat con variables climáticas establecidas. Así mismo, la segunda categoría con mayor incidencia fue la epidemiológica debido a que los países están dando mayor importancia al control y monitoreo de enfermedades arbovirósicas como lo es el dengue, a través de la publicación de tasas relacionadas con las infecciones en los portales web oficiales, siendo esta data de fácil acceso a los investigadores.

Por último, se debe señalar que una gran cantidad de artículos (11) mostraron un rango de precisión mayor o igual al 90%, lo que se puede considerar como un rango

de precisión aceptable para modelos de machine learning que busquen detectar la enfermedad del dengue.

Referencias

- Abdualgalil, B., Abraham, S., & Ismael, W. M. (2022). Early Diagnosis for Dengue Disease Prediction Using Efficient Machine Learning Techniques Based on Clinical Data. *Journal of Robotics and Control*, 3(3), 257–268. <https://doi.org/10.18196/jrc.v3i3.14387>
- Aleixo, R., Kon, F., Rocha, R., Camargo, M. S., & Camargo, R. Y. (2022). Predicting Dengue Outbreaks with Explainable Machine Learning. In 2022 22nd International Symposium on Cluster, Cloud and Internet Computing (CCGrid) 2022, (940–947). <https://doi.org/10.1109/CCGrid54584.2022.00114>
- Appice, A., Gel, Y. R., Iliev, I., Lyubchich, V., & Malerba, D. (2020). A Multi-Stage Machine Learning Approach to Predict Dengue Incidence: A Case Study in Mexico. *IEEE Access*, 8, 52713–52725. <https://doi.org/10.1109/ACCESS.2020.2980634>
- Baker, Q. B., Faraj, D., & Alguzo, A. (2021). Forecasting Dengue Fever Using Machine Learning Regression Techniques. In 2021 12th International Conference on Information and Communication Systems, ICICS 2021, May, 157–163. <https://doi.org/10.1109/ICICS52457.2021.9464619>
- Balamurugan, S. A., Mallick, M. S. M., & Chinthana, G. (2020). Improved prediction of dengue outbreak using combinatorial feature selector and classifier based on entropy weighted score based optimal ranking. *Informatics in Medicine Unlocked*, 20(March), 100400. <https://doi.org/10.1016/j.imu.2020.100400>
- Bauxell, J., Vall-Ilossera, M., & Gurgel, H. (2021). Machine learning techniques using enviromental data from remote sensing applied to modeling dengue risk in Brazil. In International Geoscience and Remote Sensing Symposium, (4608–4611). <https://doi.org/10.1109/IGARSS47720.2021.9554704>
- Benedum, C. M., Shea, K. M., Jenkins, H. E., Kim, L. Y., & Markuzon, N. (2020). Weekly dengue forecasts in Iquitos, Peru; San Juan, Puerto Rico; and Singapore. *PLoS Neglected Tropical Diseases*, 14(10), 1–26. <https://doi.org/10.1371/journal.pntd.0008710>
- Bokonda, P. L., Ouazzani-Touhami, K., & Souissi, N. (2021). Which Machine Learning method for outbreaks predictions? In 2021 IEEE 11th Annual Computing and Communication Workshop and Conference, CCWC 2021, ML, (825–828). <https://doi.org/10.1109/CCWC51732.2021.9376061>
- Bravo, J., Alarcón, R., Valdivia, C., & Serquén, O. (2023). Application of Machine Learning Techniques to Predict Visitors to the Tourist Attractions of the Moche Route in Peru. *Sustainability*, 15(11), 8967. <https://doi.org/10.3390/SU15118967>
- Chattopadhyay, A. K., & Chattopadhyay, S. (2022). VIRDOCD: A VIRtual DOCTOR to predict dengue fatality. *Expert Systems*, 39(1), 1–17. <https://doi.org/10.1111/exsy.12796>

- Choubey, D. K., Mishra, A., Pradhan, S. K., & Anand, N. (2021). Soft computing techniques for dengue prediction. In *Proceedings - 2021 IEEE 10th International Conference on Communication Systems and Network Technologies, CSNT 2021*, (648–653). <https://doi.org/10.1109/CSNT51715.2021.9509706>
- Cubas, J. V., & Maquen-Niño, G. L. E. (2022). Machine learning model in the detection of phishing websites. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (E52), 161–173. <https://www.scopus.com/record/display.uri?eid=2-s2.0-85143599912&origin=resultslist&sort=plf-f>
- Ferdousi, T., Cohnstaedt, L. W., & Scoglio, C. M. (2021). A Windowed Correlation-Based Feature Selection Method to Improve Time Series Prediction of Dengue Fever Cases. *IEEE Access*, 9, 141210–141222. <https://doi.org/10.1109/ACCESS.2021.3120309>
- Ho, T. S., Weng, T. C., Wang, J., Der Han, H. C., Cheng, H. C., Yang, C. C., Yu, C. H., Liu, Y. J., Hu, C. H., Huang, C. Y., Chen, M. H., King, C. C., Oyang, Y. J., & Liu, C. C. (2020). Comparing machine learning with case-control models to identify confirmed dengue cases. *PLoS Neglected Tropical Diseases*, 14(11), 1–21. <https://doi.org/10.1371/journal.pntd.0008843>
- Huang, S. W., Tsai, H. P., Hung, S. J., Ko, W. C., & Wang, J. R. (2020). Assessing the risk of dengue severity using demographic information and laboratory test results with machine learning. *PLoS Neglected Tropical Diseases*, 14(12) 1–19. <https://doi.org/10.1371/journal.pntd.0008960>
- Kamesh, R., & Sivakumar, N. (2020). Design and implementation of realtime detection of dengue using machine learning. *Journal of Advanced Research in Dynamical and Control Systems*, 12(5 Special Issue), 1013–1025. <https://doi.org/10.5373/JARDCS/V12SP5/20201851>
- Kaur, L., Bhatia, R., & Nagpal, S. (2021). A Machine Learning Approach for Predicting Dengue Outbreak. In *Proceedings - 2021 3rd International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2021*, (60–65). <https://doi.org/10.1109/ICAC3N53548.2021.9725539>
- Khan Dourjoy, S., Rabbani Rafi, A., Nasrin Tumpa, Z., & Saifuzzaman, M. (2021). A Comparative Study on Prediction of Dengue Fever Using Machine Learning Algorithm. In *Lecture Notes in Networks and Systems* (Vol. 127). https://doi.org/10.1007/978-981-15-4218-3_15
- Koplewitz, G., Lu, F., Clemente, L., Buckee, C., & Santillana, M. (2022). Predicting dengue incidence leveraging internet-based data sources. A case study in 20 cities in Brazil. *PLoS Neglected Tropical Diseases*, 16(1). <https://doi.org/10.1371/JOURNAL.PNTD.0010071>
- Kumar, S., Rahman, M., Howlader, A., Siddiqi, U., Mohi, K., Borhan, R., & Rahman, E. (2022). Prediction of dengue incidents using hospitalized patients, metrological and socio-economic data in Bangladesh_ A machine learning. *Plos One*. <https://doi.org/https://doi.org/10.1371/journal.pone.0270933>

- Maquen-Niño, G. L. E., Sandoval-Juarez, A. A., Veliz La Rosa, R. A., Carrión-Barco, G., Adrianzén-Olano, I., Vega-Huerta, H., & De-La-Cruz-VdV, P. (2023). Brain Tumor Classification Deep Learning Model Using Neural Networks. *International Journal of Online and Biomedical Engineering*, 19(9), 81–92. <https://doi.org/10.3991/IJOE.V19I09.38819>
- Maquen-Niño, G. L. E., Villegas-Cubas, J. E., Celis-Bravo, P. J., Carrión-Barco, G., & Adrianzen-Olano, I. (2022). Integrative methodology model for the creation of pathfinding systems. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, (E51), 46–57. https://www.researchgate.net/publication/370442097_Integrative_methodology_model_for_the_creation_of_pathfinding_systems
- McGough, S., Clemente, L., Kutz, N., & Santillana, M. (2021). A dinamyc, ensemble leanring approach to forecast dengue fever epidemic years in Brazil using weather and population susceptibility cycles. *RSIF*. 18(179). <https://doi.org/10.1098/rsif.2020.1006>
- Nalini, C., Shanthakumari, R., Prasanna, R. V., Nikilesh, A., & Pranav, S. M. N. (2022). Prediction of dengue infection using Machine Learning. In 2022 International Conference on Computer Communication and Informatics, ICCCI 2022; <https://doi.org/10.1109/ICCCI54379.2022.9740924>
- Nguyen, V.-H., Tuyet-Hanh, T. T., Mulhall, J., Minh, H. V., Duong, T. Q., Chien, N. V., Nhung, N. T. T., Lan, V. H., Minh, H. B., Cuong, D., Bich, N. N., Quyen, N. H., Linh, T. N. Q., Tho, N. T., Nghia, N. D., Anh, L. V. Q., Phan, D. T. M., Hung, N. Q. V., & Son, M. T. (2022). Deep learning models for forecasting dengue fever based on climate data in Vietnam. *PLOS Neglected Tropical Diseases*, 16(6), e0010509. <https://doi.org/10.1371/journal.pntd.0010509>
- Ochida, N., Mangeas, M., Dupont-Rouzeyrol, M., Dutheil, C., Forfait, C., Peltier, A., Descloux, E., & Menkes, C. (2022). Modeling present and future climate risk of dengue outbreak, a case study in New Caledonia. *Environmental Health: A Global Access Science Source*, 21(1), 1–10. <https://doi.org/10.1186/s12940-022-00829-z>
- OMS. (2022). Dengue y dengue grave. <https://www.who.int/es/news-room/fact-sheets/detail/dengue-and-severe-dengue>
- Pérez, L. G. A., García, M. I. B., Babici, V. R., & Monge, E. J. A. (2022). Dengue fever after the covid-19 pandemic: a statistical software analysis of its current situation in Ecuador. *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, (E53), 362–374.
- Rahman, M. S., Pientong, C., Zafar, S., Ekalaksananan, T., Paul, R. E., Haque, U., Rocklöv, J., & Overgaard, H. J. (2021). Mapping the spatial distribution of the dengue vector *Aedes aegypti* and predicting its abundance in northeastern Thailand using machine-learning approach. *One Health*, 13(December). <https://doi.org/10.1016/j.onehlt.2021.100358>

- Raizada, S., Mala, S., & Shankar, A. (2020). Vector borne disease outbreak prediction by machine learning. In *Proceedings of the International Conference on Smart Technologies in Computing, Electrical and Electronics, ICSTCEE 2020*, (213–218). <https://doi.org/10.1109/ICSTCEE49637.2020.9277286>
- Rana, S. K., Boruah, A. N., Biswas, S. K., Chakraborty, M., & Purkayastha, B. (2022). Dengue Fever Prediction using Machine Learning Analytics. In *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing*. <https://doi.org/10.1109/COM-IT-CON54601.2022.9850923>
- Salim, N. A. M., Wah, Y. B., Reeves, C., Smith, M., Yaacob, W. F. W., Mudin, R. N., Dapari, R., Sapri, N. N. F. F., & Haque, U. (2021). Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. *Scientific Reports*, 11(1), 1–9. <https://doi.org/10.1038/s41598-020-79193-2>
- Sarma, D., Hossain, S., Mittra, T., Bhuiya, M. A. M., Saha, I., & Chakma, R. (2020). Dengue Prediction using Machine Learning Algorithms. In *IEEE Region 10 Humanitarian Technology Conference, R10-HTC*. <https://doi.org/10.1109/R10-HTC49770.2020.9357035>
- Siddiq, A., Shukla, N., & Pradhan, B. (2021). Predicting Dengue Fever Transmission Using Machine Learning Methods. In *2021 IEEE International Conference on Industrial Engineering and Engineering Management, IEEM 2021*, 2007, (21–26). <https://doi.org/10.1109/IEEM50564.2021.9672977>
- Singh, K. D. (2021). Particle Swarm Optimization assisted Support Vector Machine based Diagnostic System for Dengue prediction at the early stage. In *Proceedings - 2021 3rd International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2021*, (844–848). <https://doi.org/10.1109/ICAC3N53548.2021.9725670>
- Souza, C., Maia, P., Stolerman, L. M., Rolla, V., & Velho, L. (2022). Predicting dengue outbreaks in Brazil with manifold learning on climate data (Formula presented). *Expert Systems with Applications*, 192(November 2020), 116324. <https://doi.org/10.1016/j.eswa.2021.116324>
- Srivastava, S., Soman, S., Rai, A., & Cheema, A. S. (2020). An online learning approach for dengue fever classification. In *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, (pp. 163–168). <https://doi.org/10.1109/CBMS49503.2020.00038>
- Sukama, Y., Hertono, G. F., Handari, B. D., & Aldila, D. (2020). Comparing activation functions in predicting dengue hemorrhagic fever cases in DKI Jakarta using u. In *AIP Conference Proceedings*, 2296. <https://doi.org/10.1063/5.0030456>
- Tanawi, I. N., Vito, V., Sarwinda, D., Tasman, H., & Hertono, G. F. (2021). Support Vector Regression for Predicting the Number of Dengue Incidents in DKI Jakarta. *Procedia Computer Science*, 179(2020), 747–753. <https://doi.org/10.1016/j.procs.2021.01.063>

- Xu, J., Xu, K., Li, Z., Meng, F., Tu, T., Xu, L., & Liu, Q. (2020). Forecast of dengue cases in 20 chinese cities based on the deep learning method. *International Journal of Environmental Research and Public Health*, 17(2). <https://doi.org/10.3390/ijerph17020453>
- Yavari Nejad, F., & Varathan, K. D. (2021). Identification of significant climatic risk factors and machine learning models in dengue outbreak prediction. *BMC Medical Informatics and Decision Making*, 21(1), 1–12. <https://doi.org/10.1186/s12911-021-01493-y>

Influência dos indicadores relacionados com governo eletrônico no Índice de Transparência Municipal (ITM) de Portugal

Kaio Vieira¹

kayoroberto@ua.pt

¹ Universidade de Aveiro, Aveiro, Portugal

DOI: 10.17013/risti.50.28-43

Resumo: Em que medida o Índice de Transparência Municipal (ITM) de Portugal é afetado pelos indicadores relacionados ao efeito esperado do governo eletrônico na transparência? Há inter-relacionamento entre eles? O presente artigo se propõe a responder a estas questões, utilizando-se do referencial teórico relacionado à transparência e ao governo eletrônico, que trata da importância da transparência, dos efeitos do governo eletrônico na transparência, do conteúdo dos sites governamentais e de como o governo eletrônico pode ajudar a melhorar a transparência. Por meio de procedimentos estatísticos, tratou-se de identificar a capacidade dos 15 indicadores selecionados para explicar o resultado do ITM, além de identificar similaridades entre os municípios portugueses e entre os indicadores. Os resultados mostraram que nos indicadores selecionados há poder de explicação considerável do ITM. Foi possível ainda identificar, dentre os 15, aqueles indicadores com maior poder de explicação do ITM bem como as similaridades entre os indicadores.

Palavras-chave: Transparência; Índice de transparência; Governo eletrônico; Portugal.

Influence of e-government related indicators on Portugal's Municipal Transparency Index (MTI)

Abstract: To what extent is Portugal's Municipal Transparency Index (MTI) affected by the indicators related to the expected effect of e-government on transparency? Is there an interrelationship between them? This paper proposes to answer these questions, using the theoretical framework related to transparency and e-government, which deals with the importance of transparency, the effects of e-government on transparency, the content of government websites and how e-government can help improve transparency. Through statistical procedures, we tried to identify the ability of the 15 selected indicators to explain the MTI score, and to identify similarities among the Portuguese municipalities and among the indicators. The results showed that in the selected indicators there is considerable explanatory power of the MTI. It was also possible to identify, among the 15, those indicators with greater power to explain the MTI as well as the similarities among the indicators.

Keywords: Transparency; Transparency index; Electronic government; Portugal.

1. Introdução

A transparência incorpora duas condições necessárias e suficientes – visibilidade e inferibilidade. Ocorre que, frequentemente, não se cumpre uma ou ambas as condições necessárias. Tal conceituação bidimensional permite graus contínuos de transparência, de baixa a excelente qualidade e todos os pontos intermediários. Desse conceito, advém uma premissa principal: transparência é informação registada. O princípio da visibilidade é justamente o motivo pelo qual as transcrições legislativas e os votos registados se tornaram a condição *sine qua non* do movimento de transparência parlamentar. Não há transparência processual sem registo; há apenas resultado (Michener & Bersch, 2013).

A transparência pode ser definida como “a disponibilidade de informações sobre uma organização ou ator que permite que atores externos monitorem o funcionamento interno ou o desempenho dessa organização”. Para tanto, no contexto das organizações públicas, a transparência é importante porque facilita a capacidade dos cidadãos de observar e compreender o que está acontecendo dentro de seu governo. Como tal, as tentativas de aumentar a transparência baseiam-se nos esforços do governo para divulgar publicamente informações relativas ao seu funcionamento interno (Porumbescu, 2017).

No contexto da Administração Pública, transparência significa a disponibilização de acesso das informações governamentais aos cidadãos. E à medida que há a percepção, cada vez mais recorrente, da importância da transparência, surge a necessidade de se criar instrumentos de medição da transparência.

Em Portugal, foi criado o Índice de Transparência Municipal (ITM) para este fim. Este, como o próprio nome indica, é um índice de aferição do grau de transparência municipal. Sua criação foi proposta pela Transparência e Integridade Associação Cívica (TIAC) de Portugal.

Os índices de transparência normalmente são compostos por indicadores. Nesse sentido, pretende-se identificar em que medida o ITM de Portugal é afetado pelos indicadores mais diretamente relacionados ao efeito esperado do governo eletrônico na transparência, que é o incentivo ao diálogo com o cidadão. Pretende-se ainda identificar se há inter-relacionamento entre esses indicadores.

O presente artigo se propõe a responder a estas questões, utilizando-se do referencial teórico relacionado à transparência e ao governo eletrônico.

Como desafio à pesquisa, pretende-se adotar um modelo de regressão linear com a finalidade de analisar a correlação entre os indicadores selecionados (variáveis independentes) e a variável índice de transparência (variável dependente). Realizou-se ainda uma análise de cluster, na busca de similaridades entre os municípios portugueses, bem como entre os indicadores selecionados.

Inicialmente são apresentados os principais tópicos relacionados ao ITM, sua conceituação, metodologia, seus objetivos e sua composição. Em seguida é apresentada a metodologia utilizada no presente trabalho. Na continuação, apresenta-se o referencial teórico utilizado para orientar o estudo, desde a definição da questão de pesquisa, a

seleção dos indicadores e a efetiva análise. Ao final são apresentados os resultados do estudo.

2. O ITM e seu referencial teórico

De acordo com o documento que apresenta a metodologia e faz a descrição dos indicadores do ITM¹, a ideia da TIAC foi desenvolver um índice que permitisse ao cidadão e ao decisor aferir o grau de transparência do seu município, sob a justificativa de que o Poder Local constitui uma pedra angular da democracia portuguesa.

No documento, afirma-se que o ITM “serve para criar critérios universais de aferição dos níveis de transparência nos municípios, através da análise de informação sobre a governação local disponibilizada nos seus websites”.

A elaboração de um ranking visa “criar pressão social e incentivos para as autoridades locais melhorarem as suas ferramentas de comunicação e interação com os cidadãos, com vista a uma governação mais aberta, responsável e participativa”.

Para a TIAC, como afirmado no documento, “o uso das Tecnologias de Informação e Comunicação não melhora por si só a qualidade da democracia, mas é um pequeno contributo para uma maior capacitação dos cidadãos na monitorização e participação”.

O conceito de transparência para o ITM é o seguinte: “providenciar informação relevante, fidedigna, atempada, inteligível e de fácil acesso sobre formato, desempenho e gestão do bem público”.

Nota-se a preocupação da TIAC em primar pela qualidade do índice, ao definir princípios, diretrizes e objetivos claros, como por exemplo, a preocupação com a melhoria da comunicação e interação entre governos locais e cidadãos, criação de um governo mais aberto, responsável e participativo.

O ITM é comparável a outros índices desenvolvidos por Capítulos Nacionais da TI, como TIEspanha ou TI-Eslováquia (da Cruz et al., 2016). Porém, possui as peculiaridades da realidade portuguesa.

O primeiro passo do projeto do ITM foi estabelecer uma equipe de pesquisa incluindo membros do TIAC e quatro instituições académicas diferentes e um grupo consultivo abrangente (AG) de especialistas para examinar e aprovar cada passo metodológico (da Cruz et al., 2016).

O ITM é composto por um conjunto de 76 indicadores agrupados nas sete dimensões discriminadas a seguir:

- A. Informações organizacionais, composição social e funcionamento do município (órgãos executivos e deliberativos) (18 indicadores);
- B. Planos e Relatórios (13 indicadores);
- C. Impostos, taxas, tarifas, preços e regulamentos (5 indicadores);

¹ Disponível em: https://transparencia.pt/wp-content/uploads/2017/11/ITM_Apresentacao_e_Indicadores_2017.pdf.

- D. Relação com a sociedade (8 indicadores);
- E. Contratação Pública (10 indicadores);
- F. Transparência econômico-financeira (12 indicadores);
- G. Transparência na área do Urbanismo (10 indicadores).

O ITM não valoriza todas as dimensões de forma igual. Nesse sentido, foram atribuídos pesos às dimensões, conforme figura a seguir:

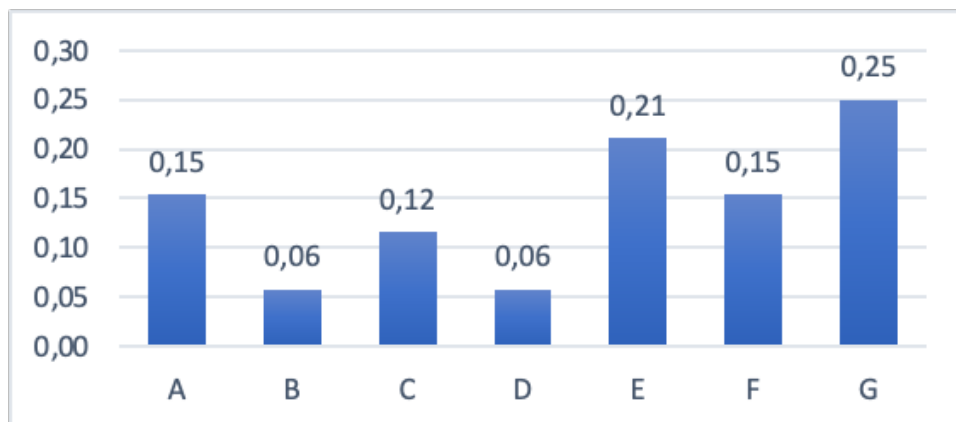


Figura 1 – Pesos das dimensões. Fonte: Documento que apresenta a metodologia e descreve os indicadores do ITM.

Os pesos adotados para o ITM português são resultado das opiniões dos membros do AG, mas a metodologia do índice pode ser replicada em outros contextos. Tais pesos variam de acordo com a composição do AG ou grupo decisório (da Cruz et al., 2016).

Os resultados do ITM têm sido publicados anualmente desde 2013, sendo que a última edição foi publicada em 2017, totalizando cinco edições. Para o presente trabalho, foram utilizados os resultados do ITM 2017. Nas tabelas seguintes são apresentadas as listas dos municípios que obtiveram os melhores índices e dos que obtiveram os piores resultados:

ID	Municípios	ITM	Ranking
13	Alfândega da Fé	90,66	1
285	Vila do Bispo	90,66	1
292	Vila Nova de Cerveira	90,25	3
31	Arcos de Valdevez	86,95	4
234	Santa Marta de Penaguião	86,54	5
257	Sintra	84,89	6
182	Oliveira do Hospital	84,07	7
261	Tábua	83,79	8

ID	Municípios	ITM	Ranking
78	Cinfães	83,24	9
124	Leiria	82,83	10

Fonte: Resultado ITM 2017.

Tabela 1 – Melhores resultados do ITM 2017

ID	Municípios	ITM	Ranking
303	Vila Viçosa	24,86	299
103	Freixo de Espada à Cinta	23,76	300
83	Corvo	23,63	301
18	Almeida	23,35	302
250	Serpa	22,11	303
100	Figueira de Castelo Rodrigo	21,56	304
163	Mora	14,70	305
165	Moura	13,60	306
191	Pedrógão Grande	10,71	307
212	Porto Moniz	10,17	308

Fonte: Resultado ITM 2017.

Tabela 2 – Piores resultados do ITM 2017

A literatura fornece informações sobre a criação de outras escalas ou índices de avaliação de transparência, como a Escala Brasil Transparente (EBT), uma metodologia utilizada para medir a transparência pública em estados e municípios brasileiros (Silva e Bruni, 2019), “The fiscal transparency index” e “The e-democracy index” (EDI) (Manes Rossi et al., 2018) ou o Government Openness Index (GOI) (Park & Oh, 2019), este último para aplicação em países desenvolvidos, segundo os autores.

Na opinião de Da Cruz e outros (2016), a maioria dos índices apresentados na literatura sofre de duas limitações principais: vários índices são baseados em modelos simples de avaliação aditiva com pesos iguais para todos os indicadores; numerosos índices se baseiam em inquéritos aos municípios, o que acarreta tanto problemas de auto seleção das respostas como uma representação imprecisa do nível real de transparência devido ao viés positivo introduzido pelos inquiridos (da Cruz et al., 2016).

O ITM busca evitar esse problema usando uma abordagem participativa para recuperar os coeficientes de ponderação de um processo iterativo com um grupo de tomada de decisão e evita as dificuldades associadas a um índice baseado em pesquisas. Por se basear em sites municipais, não há necessidade de atribuir uma pontuação mínima aos municípios que não responderam (da Cruz et al., 2016).

Nota-se, portanto, que o ITM é baseado na avaliação de sites municipais. Dessa forma, espera-se que estes sites propiciem transparência, sejam interativos e contenham as características abordadas em cada um dos 76 indicadores do índice.

Quanto à interatividade, percebe-se que o uso crescente da Web 2.0, da mídia social e das TIC móveis e sem fio pelos cidadãos pode impactar fortemente a forma como os serviços públicos são prestados e como os processos de engajamento dos cidadãos são realizados (Liu & Yuan, 2015).

Cada vez mais se pretende alcançar melhor interatividade. Dessa forma, as plataformas de governo aberto têm se tornado um marco tecnológico essencial para o empoderamento dos cidadãos (Maier-Rabler & Huber, 2011).

Importante para o presente estudo é o que afirma Pina et al. (2007) acerca do efeito do governo eletrônico na transparência. Para os autores, os benefícios só virão se o conteúdo dos sites for centrado no cidadão. Afirmam ser importante que os governos fortaleçam a interatividade de seus sites, pois foi identificado que poucos sites mostram sinais claros de abertura que incentivam o diálogo cidadão. Nesse sentido, recriar a estrutura organizacional existente não contribui para transformar o governo e não ajuda os cidadãos não familiarizados com a organização a navegar em um site. Deve-se fortalecer a interatividade dos sites governamentais (Pina et al., 2007).

O governo eletrônico tem um enorme potencial para aproximar os governos dos cidadãos e, portanto, para aumentar a confiança do público nos governos. A contribuição dos sites para promover a transparência, prestação de contas e abertura e para alterar a relação burocrática entre governo e cidadãos, governos e formuladores de políticas ocorrerá se for fortalecida a sua interatividade (Pina et al., 2009).

São basicamente esses os valores almejados pelo ITM, promoção da transparência, aproximação e diálogo com o cidadão, prestação de contas, dentre outras. Resumidamente, os efeitos do governo eletrônico na transparência são os seguintes: Benefícios virão se os conteúdos dos sites forem centrados no cidadão; a interatividade dos sites deve ser fortalecida; deve-se incentivar o diálogo com o cidadão; somente recriar a estrutura existente não contribui para transformar o governo. (Pina et al., 2007, 2009).

Os diferentes aspetos da interação cidadão-setor público podem ser colocados em três diferentes grupos, que também representam diferentes estágios de adoção do governo eletrônico: Fornecimento e disseminação de informações; Transações e prestação de serviços; Participação e colaboração (Andersen et al., 2010).

São esses, portanto, os três estágios de adoção do governo eletrônico e os quais devem os sites governamentais buscar alcançar, ou seja, disponibilizar a informação, permitir transações e prestar contas aos cidadãos e permitir a participação e colaboração dos cidadãos no processo de tomada de decisões.

O autor cita estudos de Chadwick e May (2003) e Torres, Pina e Royo (2005) para afirmar que, apesar da possibilidade de interação, as dimensões de participação e consulta em iniciativas de governo eletrônico permanecem mínimas. Já ao citar estudo de Welch e Pandey (2006), que examinou a relação entre a burocracia e uso da internet, afirma que

o uso da internet está associado à redução na burocracia. O estudo conclui que o estágio “outdoor” do governo eletrônico está bem desenvolvido. Embora o governo eletrônico tenha elementos que o diferenciam do uso anterior das TIC na administração pública, a pesquisa sobre governo eletrônico ainda não demonstrou uma mudança substancial na natureza e nas direções dos impactos na Administração Pública (Andersen et al., 2010).

De forma transversal, os conceitos de transparência e de governo aberto também foram relacionados como de interesse pelos pesquisadores, tanto na perspectiva de identificar os efeitos do governo eletrônico na transparência, como no desenvolvimento de um índice de abertura do governo e para analisar a transparência e a participação democrática nos governos locais (Manes Rossi et al., 2018), (Park & Oh, 2019), (Pina et al., 2007), (Pina et al., 2009).

Em termos de uso das mídias sociais, existem vários papéis importantes que elas podem desempenhar como meio de governo eletrônico colaborativo para ajudar a melhorar a transparência do governo e aumentar o combate à corrupção (John Carlo Bertot et al., 2012): aumentar e melhorar o acesso ao público às informações governamentais, oferecendo informações via internet por meio de múltiplos canais interativos dinâmicos; interagir com o público e abordar interesses e preocupações específicas dos cidadãos; alcançar populações que, de outra forma, não encontrariam as informações do governo; servir como veículos de informação e comunicação para denunciantes divulgarem informações confidenciais; complementar ou substituir informações corruptas ou deficientes por jornalismo cidadão para melhor informar o público sobre as condições.

Baseado nesse referencial teórico é que foram selecionados os 15 indicadores do ITM, dentre os 76 existentes, para o presente estudo.

3. Metodologia

Dos 76 indicadores que compõem o ITM, foram selecionados aqueles mais diretamente relacionados ao efeito esperado do governo eletrônico na transparência, que é o incentivo ao diálogo com o cidadão.

A seleção é composta de 15 indicadores, todos voltados a aferição da disponibilização de informações, publicação de conteúdos considerados importantes ao cidadão, motores de busca, redes sociais, serviços de pedidos de informação, espaço para reclamação e outros. O quadro a seguir apresenta os indicadores selecionados:

Indicador	
1	A.1.3. Publicação dos endereços eletrônicos dos membros do Executivo
2	A.2.3. Publicação de avisos de abertura de procedimentos concursais de recrutamento
3	A.3.1. Informação geral sobre os diferentes órgãos autárquicos e suas funções
4	A.3.6. Publicação dos endereços eletrônicos gerais dos diferentes órgãos autárquicos (CM + AM) e Juntas de Freguesia
5	B.4. Publicação de um relatório com informação sobre o volume e tipo de reclamações e/ou sugestões por unidade/serviço
6	C.4. Publicação do Boletim Municipal

Indicador	
7	D.1. Motor de busca no site do Município
8	D.2. Link(s) para redes sociais com atividade
9	D.4. Serviço de pedidos de informação que possibilite ao cidadão acompanhar o procedimento administrativo online
10	D.5. Provedor do Município: Publicação do endereço eletrónico/contacto
11	D.8. Espaço para Reclamações / Sugestões
12	E.3.3. Publicação dos pareceres, vistos e relatórios de auditoria das entidades de fiscalização
13	F.1.1. Orçamento do Município
14	F.2.3. Relatório de Gestão
15	G.2.4. Publicação dos resultados da discussão pública dos planos Municipais de Ordenamento do Território

Fonte: Elaboração própria.

Quadro 1 – Indicadores selecionados

Selecionados os indicadores, realizou-se uma análise de regressão, tendo como variável dependente o ITM de cada município e variáveis independentes os indicadores selecionados. Também foi realizada uma análise de cluster, visando identificar possíveis similaridades entre os municípios portugueses, com base na classificação de cada um dos indicadores selecionados. Da mesma forma, essa análise buscou identificar possíveis similaridades entre os próprios indicadores. O trabalho foi conduzido utilizando-se como apoio as ferramentas “Excel” e “Geoda”.

Buscou-se identificar a importância do papel do governo eletrônico, descrita no referencial teórico, no resultado do ITM.

4. Resultado

Em uma primeira análise dos resultados do ITM 2017, observa-se que a escala do ITM considera como aceitável o score de 36 a 63 e como bom, a partir de 64. 74 municípios atingiram score acima de 64, o equivalente a 24% dos municípios portugueses. 161 municípios atingiram score entre 36 e 63, equivalente a 52% e os demais, totalizando 73 municípios atingiram score abaixo de 36, considerado insuficiente conforme escala do ITM.

Conforme se observa no mapa, há 62 municípios portugueses com índice na escala entre 66 e 90, score considerado bom, conforme documento de apresentação da metodologia e dos indicadores do ITM. Nota-se que, embora se veja uma leve concentração da cor mais escura ao sul, um pouco ao norte e levemente inclinada ao centro-oeste, sua distribuição parece não seguir um padrão, ou seja, não é uniforme. Parece não haver, portanto, uma relação direta entre a localização geográfica do município e o resultado do ITM.

O gráfico a seguir mostra a dispersão dos resultados do ITM 2017 com base no índice de Moran. O índice de Moran é uma ferramenta estatística para testar a autocorrelação espacial. O objetivo é verificar se a relação de vizinhança interfere no resultado do ITM,

ou seja, se o município com ITM bom pode influenciar de alguma forma seus vizinhos e esses passam a atingir bons scores de ITM.

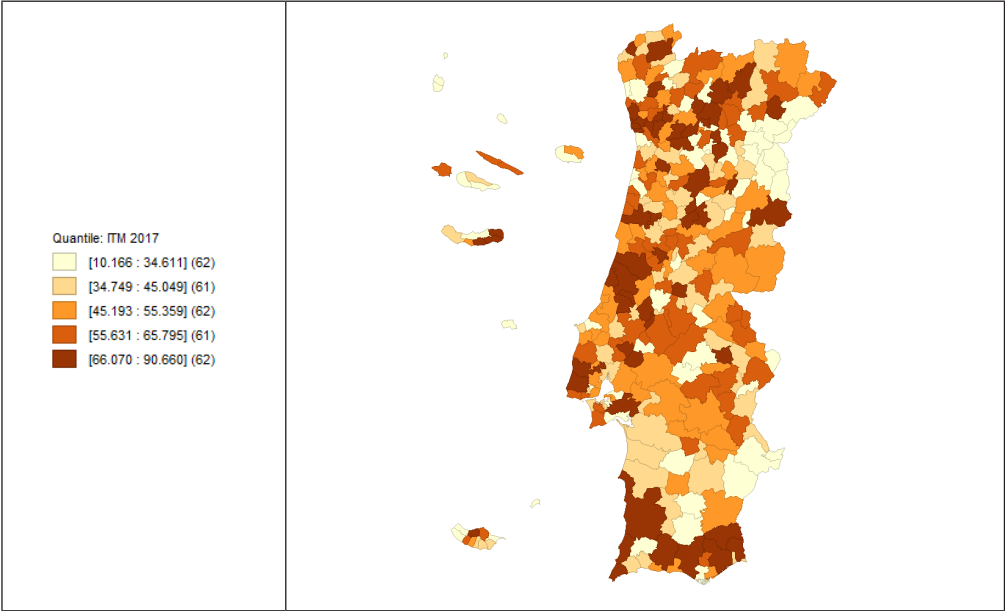


Figura 2 – Distribuição Geográfica do ITM 2017 - Portugal.

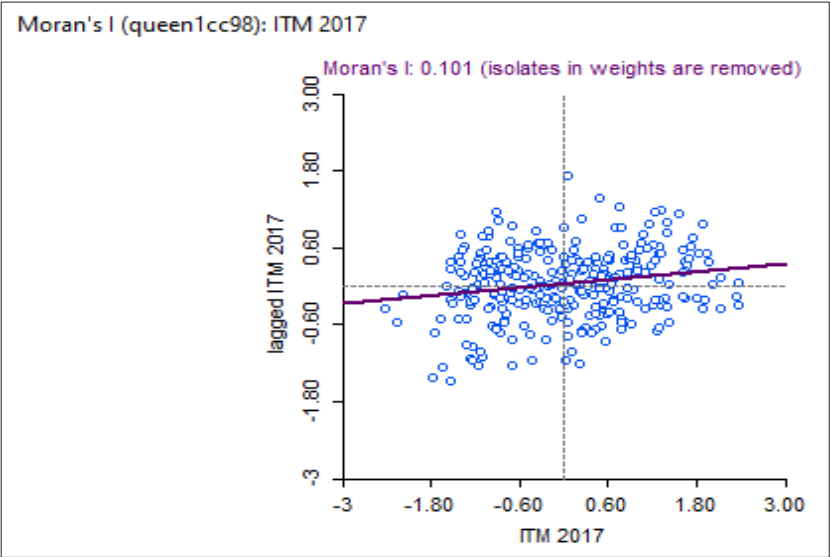


Figura 3 – Dispersão – Índice Moran do ITM 2017 - Portugal.

O resultado apresentou um valor baixo, de 0,101 e percebe-se não haver concentração em nenhum dos quadrantes, significando baixa correlação entre o ITM e a localização geográfica. Na representação da dispersão no mapa a seguir pode-se ver que para 243 municípios a localização geográfica não é significativa para o resultado do ITM. Ou seja, não importa o resultado do ITM do município vizinho para o seu próprio resultado. Confirma-se, assim, a informação de não haver relação direta entre a localização geográfica do município e o resultado do ITM.

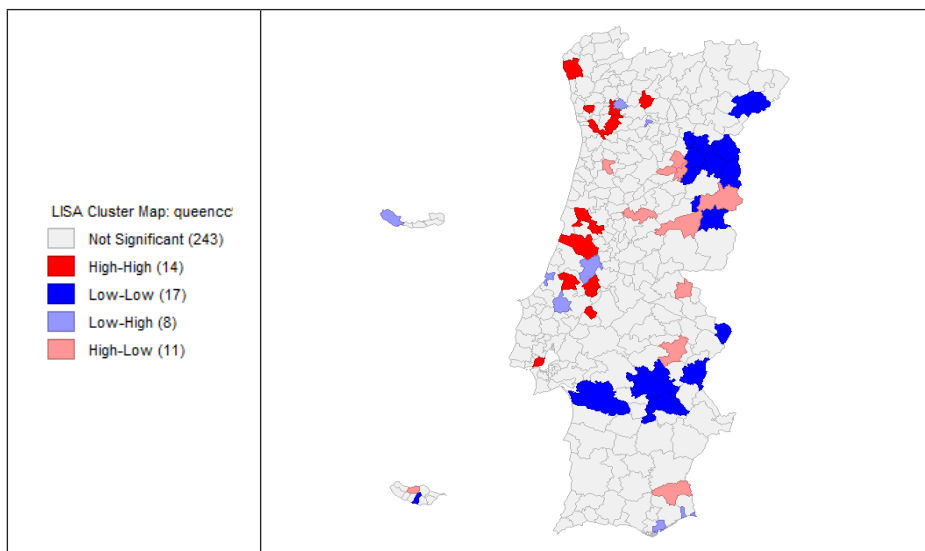


Figura 4 – Representação da Dispersão no mapa - Índice Moran do ITM 2017 - Portugal.

Em 14 municípios, os da cor vermelha, o resultado apresentou uma relação high-high, ou seja, ITM alto para ambos os vizinhos. Em 17 municípios, a relação foi de ITM baixo para ambos. Nesses dois casos há uma correlação direta. Nos casos low-high (8) e high-low (11), a relação é inversa.

A conclusão é que o Índice de Moran mostrou não haver correlação entre o ITM de um município e o ITM dos municípios vizinhos.

Partindo-se para a análise de regressão, sendo que a variável dependente é o ITM e as variáveis independentes são os indicadores selecionados, apresenta-se os resultados a seguir:

O resultado R-Quadrado = 0,644755 nos informa que há uma linearidade entre os indicadores e o ITM. Ou seja, os indicadores selecionados explicam em 64% o ITM.

<i>Estatística de regressão</i>	
R múltiplo	0,802966
R-Quadrado	0,644755
R-quadrado ajustado	0,626506
Erro padrão	10,33107
Observações	308

	Coefficientes	Erro padrão	Stat t	valor-P
Interseção	-76,1794	9,594372	-7,94001	4,39E-14
A.1.3	4,592269	1,308799	3,508765	0,000521
A.2.3	6,596331	1,884951	3,49947	0,000539
A.3.1	7,845099	1,401664	5,596989	5,03E-08
A.3.6	1,178722	1,36404	0,86414	0,38822
B.4	8,340623	1,889092	4,41515	1,42E-05
C.4	2,393589	1,248302	1,917477	0,056153
D.1	3,064457	2,539299	1,206812	0,228481
D.2	2,722877	2,55341	1,066369	0,287138
D.4	4,13784	1,29942	3,184374	0,001608
D.5	6,558992	4,15264	1,579475	0,11531
D.8	5,371521	1,342066	4,002425	7,96E-05
E.3.3	16,16786	2,529952	6,39058	6,52E-10
F.1.1	5,225999	2,883173	1,812586	0,070923
F.2.3	7,525646	2,627097	2,864625	0,004478
G.2.4	5,522606	1,471592	3,752811	0,000211

Fonte: Elaboração própria.

Tabela 3 – Resumo dos Resultados

Já na coluna dos coeficientes, os valores representam o fator multiplicativo no ITM para cada unidade de aumento o indicador. O resultado apresentado mostra que com exceção dos indicadores A.3.6, C.4 e D.2, todos exercem considerável influência no ITM, porém, os indicadores que mais influenciam o ITM são os seguintes:

- A.3.1. Informação geral sobre os diferentes órgãos autárquicos e suas funções;
- B.4. Publicação de um relatório com informação sobre o volume e tipo de reclamações e/ou sugestões;
- E.3.3. Publicação dos pareceres, vistos e relatórios de auditoria das entidades de fiscalização;
- F.2.3. Relatório de Gestão.

Os efeitos do governo eletrônico na transparência, identificados por Pina et. Al. (2007, 2009), ou seja, com conteúdo dos sites centrados no cidadão, fortalecimento da interatividade, incentivo do diálogo com o cidadão e não repetição da estrutura existente estão presentes em todos os indicadores.

Dando seguimento às análises, visando identificar similaridades entre as variáveis independentes, bem como entre os municípios, realizou-se uma análise de clusters, cujos resultados são apresentados a seguir:

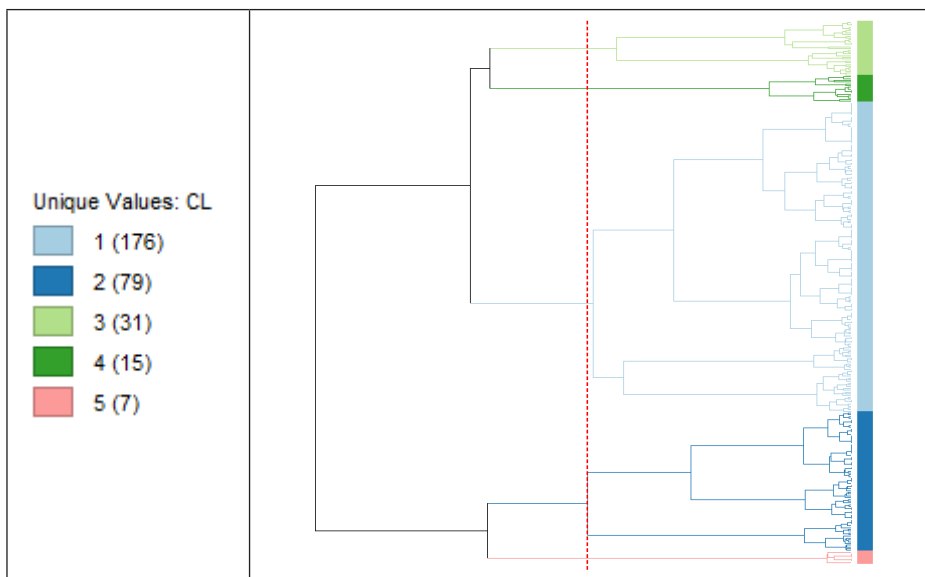


Figura 5 – Dendrograma dos municípios.

No dendrograma nota-se a existência de cinco clusters, de modo que o primeiro, C1, abrange a maioria dos municípios portugueses (176), seguido do C2, com 79 municípios. No mapa a seguir, pode-se observar a subdivisão dos municípios nos clusters.

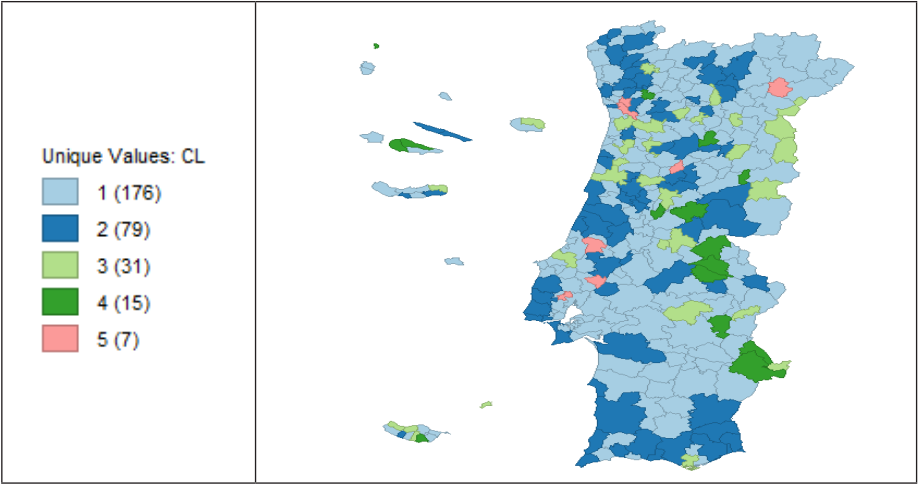


Figura 6 – Mapa dos clusters formados.

Quanto aos indicadores selecionados, apresenta-se a seguir o resultado da análise de cluster:

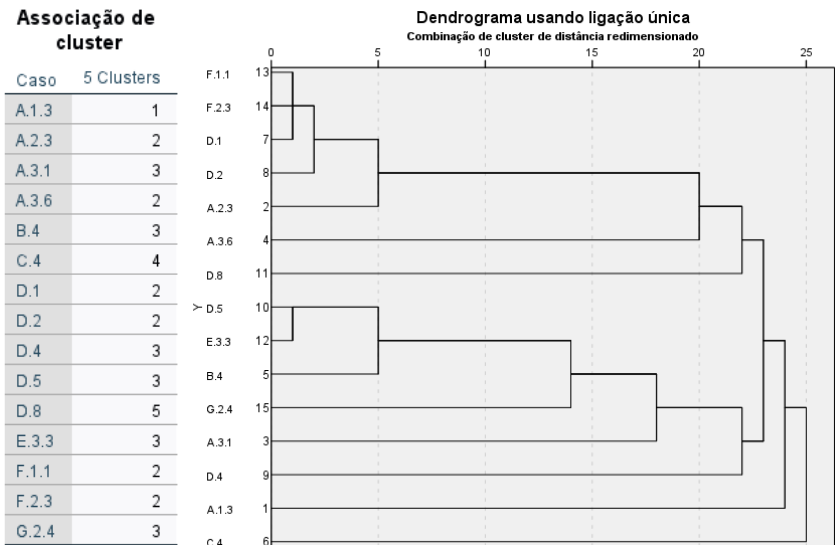


Figura 7 – Cluster dos indicadores.

A subdivisão dos indicadores em cinco clusters apresentou o seguinte resultado:

Cluster	Indicador
1	A.1.3. Publicação dos endereços eletrônicos dos membros do Executivo
2	A.2.3. Publicação de avisos de abertura de procedimentos concursais de recrutamento A.3.6. Publicação dos endereços eletrônicos gerais dos diferentes órgãos autárquicos (CM + AM) e Juntas de Freguesia D.1. Motor de busca no site do Município D.2. Link(s) para redes sociais com atividade F.1.1. Orçamento do Município F.2.3. Relatório de Gestão
3	A.3.1. Informação geral sobre os diferentes órgãos autárquicos e suas funções B.4. Publicação de um relatório com informação sobre o volume e tipo de reclamações e/ou sugestões por unidade/serviço D.4. Serviço de pedidos de informação que possibilite ao cidadão acompanhar o procedimento administrativo online D.5. Provedor do Município: Publicação do endereço eletrónico/contacto E.3.3. Publicação dos pareceres, vistos e relatórios de auditoria das entidades de fiscalização G.2.4. Publicação dos resultados da discussão pública dos planos Municipais de Ordenamento do Território
4	C.4. Publicação do Boletim Municipal
5	D.8. Espaço para Reclamações / Sugestões

Fonte: Elaboração própria.

Quadro 2 – Subdivisão dos indicadores nos clusters

Observa-se, desde já, a relevância do cluster 3 que abriga três dos quatro indicadores que mais influenciam o ITM: A.3.1, B.4 e E.3.3. Além disso, os efeitos do governo eletrônico que ajudam a melhorar a transparência, identificados na literatura de referência, estão todos representados nos indicadores:

- Conteúdo centrado no cidadão (A.2.3, D.4, D.8);
- Fortalecimento da interatividade dos sites/Incentivo ao diálogo/Interagir com o público (D.1, D.2, D.4, D.8);
- Oferecer informações via internet/alcançar a população/complementar informação (A.1.3, A.2.3, A.3.1, A.3.6, B.4, C.4, D.5, E.3.3, F.1.1, F.2.3, G.2.4);
- Veículo de informação para denunciantes (D.4, D.8).

6. Conclusão

Os resultados mostraram que os 15 indicadores selecionados, aqueles mais relacionados ao tratamento dos efeitos do governo eletrônico na transparência, guardam relação de linearidade com o ITM, possuindo capacidade preditiva sobre este. A regressão alcançou o percentual de 64% do ITM.

Os indicadores que menos exercem influência sobre o ITM são os seguintes:

- A.3.6. Publicação dos endereços eletrônicos gerais dos diferentes órgãos autárquicos (CM + AM) e Juntas de Freguesia;
- C.4. Publicação do Boletim Municipal;
- D.2. Link(s) para redes sociais com atividade.

Já os indicadores que mais exercem influência sobre o ITM são os seguintes:

- A.3.1. Informação geral sobre os diferentes órgãos autárquicos e suas funções;
- B.4. Publicação de um relatório com informação sobre o volume e tipo de reclamações e/ou sugestões;
- E.3.3. Publicação dos pareceres, vistos e relatórios de auditoria das entidades de fiscalização;
- F.2.3. Relatório de Gestão.

Quanto ao teste de autocorrelação espacial, que teve por objetivo verificar se a relação de vizinhança entre os municípios portugueses interfere no resultado do ITM, a conclusão foi que não há correlação entre o ITM de um município e o ITM dos municípios vizinhos.

Foi verificado, por fim, que os efeitos do governo eletrônico que ajudam a melhorar a transparência estão todos representados nos indicadores. Portanto, os indicadores selecionados tem o potencial de avaliar se o conteúdo dos sites governamentais são centrados no cidadão, se tem havido fortalecimento da interatividade dos sites governamentais, se há incentivo ao diálogo com o cidadão, à interação com o público, ao oferecimento de informações via internet, com alcance da população que de outra forma não teria acesso, e se tem havido a complementação de informação bem como se criado um veículo de informação para denunciantes.

Referências

- Andersen, K. N., Henriksen, H. Z., Medaglia, R., Danziger, J. N., Sannarnes, M. K., & Enemærke, M. (2010). Fads and facts of E-government: A review of impacts of E-government (2003-2009). *International Journal of Public Administration*, 33(11), 564–579. <https://doi.org/10.1080/01900692.2010.517724>
- Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2012). Promoting transparency and accountability through ICTs, social media, and collaborative e-government. *Transforming Government: People, Process and Policy*, 6(1), 78–91. <https://doi.org/10.1108/17506161211214831>
- Da Cruz, N. F., Tavares, A. F., Marques, R. C., Jorge, S., & de Sousa, L. (2016). Measuring Local Government Transparency. *Public Management Review*, 18(6), 866–893. <https://doi.org/10.1080/14719037.2015.1051572>
- Manes Rossi, F., Brusca, I., & Aversano, N. (2018). Financial Sustainability as a Driver for Transparency and E-Democracy: A Comparative Study in Italian and Spanish Local Governments. *International Journal of Public Administration*, 41(1), 22–33. <https://doi.org/10.1080/01900692.2016.1242623>
- Maier-Rabler, U., & Huber, S. (2011). “Open”: The changing relation between citizens, public administration, and political authority: Towards researching the Public Value of Open Government. *EJournal of EDemocracy and Open Government*, 3(2), 182–191. <https://doi.org/10.29379/jedem.v3i2.66>
- Michener, G., & Bersch, K. (2013). Identifying transparency. *Information Polity*, 18(3), 233–242. <https://doi.org/10.3233/IP-130299>

- Park, E. G., & Oh, W. (2019). Developing a Government Openness Index: The case of developing countries. *Information Development*, 35(1), 121–134. <https://doi.org/10.1177/0266666917731946>
- Pina, V., Torres, L., & Royo, S. (2007). Are ICTs improving transparency and accountability in the EU regional and local governments? An empirical study. *Public Administration*, 85(2), 449–472. <https://doi.org/10.1111/J.1467-9299.2007.00654.X>
- Pina, V., Torres, L., & Royo, S. (2009). E-government evolution in EU local governments: A comparative perspective. *Online Information Review*, 33(6), 1137–1168. <https://doi.org/10.1108/14684520911011052>
- Porumbescu, G. (2017). Linking Transparency to Trust in Government and Voice. *American Review of Public Administration*, 47(5), 520–537. <https://doi.org/10.1177/0275074015607301>
- Silva, W. A. O., Bruni, A. L. (2019). Variáveis socioeconômicas determinantes para a transparência pública nos municípios brasileiros. *Revista de Administração Pública*, 53(2), 415–431. <http://dx.doi.org/10.1590/0034-761220170383>

Efeitos da realidade aumentada na aprendizagem das ciências naturais: uma revisão sistemática da literatura

António Faria¹, Guilhermina Lobato Miranda¹

antonio.faria@campus.ul.pt; gmiranda@ie.ulisboa.pt

¹ Instituto de Educação, Universidade de Lisboa, Alameda da Universidade, 1649-013 Lisboa, Portugal

DOI: 10.17013/risti.50.44-57

Resumo: Este artigo apresenta os resultados de uma revisão sistemática da literatura sobre os efeitos da utilização da realidade aumentada (RA) na aprendizagem dos estudantes de ciências (Ciências Naturais e Biologia) do 3.º ciclo do ensino básico e do ensino secundário. Foi utilizada a metodologia PRISMA, tendo sido selecionados 10 artigos de revistas indexadas, num período de tempo entre 2010 e 2022. As bases de dados consultadas foram *Scopus*, *ScienceDirect*, *Springer Link*, *EBSCO* e *Web of Science*. A maioria dos artigos selecionados (n=8), após a aplicação dos critérios de inclusão e exclusão, utilizou uma metodologia experimental. Os resultados indicam que a RA teve efeitos positivos na aprendizagem e motivação dos estudantes, bem como noutras variáveis como as competências visuoespaciais e o envolvimento dos estudantes nas tarefas de aprendizagem.

Palavras-chave: realidade aumentada; ensino das ciências; desempenho académico.

Effects of using augmented reality on students learning: a systematic review

Abstract: This article presents the results of a systematic review of the literature regarding the effects of using augmented reality (AR) on the learning of science students (Natural Sciences and Biology) in lower and high secondary education. The PRISMA methodology was used. Ten articles were selected from indexed journals, in a period of time between 2010 and 2022. The databases consulted were Scopus, ScienceDirect, Springer Link, EBSCO, and Web of Science databases. Most articles selected (n=8), after applying the inclusion and exclusion criteria, used an experimental methodology. The results indicate that AR had positive effects on students' learning and motivation, as well as on other variables such as visuospatial skills and student involvement in tasks.

Keywords: augmented reality; science education; academic achievement.

1. Introdução

A integração da realidade aumentada (RA) em contextos de trabalho e aprendizagem surge em meados dos anos 90 através engenheiro Tom Caudell (Karagozlu, 2018 e Singh, 2019). A RA pode ser integrada num contexto mais amplo, que inicialmente se definiu como *virtuality continuum* (Milgram & Kishino, 1994), mas atualmente foi revisto para o quadro teórico de realidade mista (Skarabez et al., 2021). Nesta revisão Skarabez et al. (2021) referem “*mixed reality (MR) environment is one in which real world and virtual world objects are stimulated and presented together within a single perception. That is, when a user simultaneously perceives both real and virtual content...*” (p.4).

O uso da RA é muito diversificado e podemos destacar, entre outros, os contextos militar, industrial, médico, museológico, entretenimento e navegação (Singh, 2019). A evolução dos sistemas móveis, o facto de serem *gadgets* acessíveis, aliado ao desenvolvimento constante de aplicações (app) de RA, tem favorecido a sua integração em contextos de aprendizagem formal e informal (Nincarean et al., 2013; Zheng, 2015). A RA torna visíveis aspetos microscópicos, bem como a manipulação simulada de situações potencialmente perigosas (Page et al., 2021; Say & Pan, 2017).

Os sistemas que geram a RA podem funcionar com ou sem o uso de marcadores (Karagozlu, 2018; Weng et al., 2020). Em contexto de aprendizagem com integração da RA menos imersiva, é importante acautelar o conjugado uso de app, com os dispositivos mais apropriados, nomeadamente *smartphone* ou *tablet* (Parong, 2022, p. 503). Numa análise de diversas revisões de literatura e estudos meta-analíticos, destaca-se que sistemas pouco imersivos, como os criados a partir de *smartphone* se mostram facilitadores da aprendizagem, quando são implicados os princípios da Teoria Cognitiva da Aprendizagem Multimédia (Parong, 2022, pp. 504-505). Os sistemas baseados em marcadores incluem o uso de imagens e/ou códigos QR (*quick response*). Existem três componentes fundamentais: informação do marcador, uma câmara para capturar essa imagem e o produto 3D que é formado e apresentado. Por outro lado, aplicações sem o uso de marcadores requerem um Sistema de Posicionamento Global (GPS) e uma bússola ou um sistema de monitorização composto por um dispositivo de reconhecimento de imagem através da câmara (Karagozlu, 2018; Sahin et al., 2019; Hung et al., 2016; Weg et al. 2020), aspetos que os *smartphones* atuais integram. Naranjo et al. (2021) destacam a importância da RA na educação, nomeadamente pelas imagens produzidas e que, ao fornecerem feedback, orientam os alunos na resolução de problemas e na tomada de decisões.

A análise dos estudos realizados foi orientada pelas seguintes questões:

1. Quais são os efeitos do uso da RA na aprendizagem de Ciências (Biologia e Ciências Naturais) de alunos do 3.º ciclo do ensino básico e do ensino secundário?
2. Existem outras variáveis que expliquem o desempenho positivo dos alunos que utilizaram a RA como parte das atividades de ensino e aprendizagem?

2. Materiais e Métodos

2.1. Recolha de dados

Foi realizada uma revisão sistemática da literatura seguindo a metodologia PRISMA (Page et al., 2021). O período considerado foi entre 2010 e 2022. As bases de dados consultadas foram *Scopus*, *ScienceDirect*, *Springer Link*, *EBSCO* e *Web of Science*. Usámos os termos realidade aumentada (*augmented reality/ realidad aumentada/ réalité augmentée*), e ensino das ciências (*science teaching/ enseñanza de las ciencias/ enseignement des sciences*) e excluímos os termos realidade virtual (*virtual reality/ realidad virtual/ réalité virtuelle*) e universidade (*university/ universidad/ université*) conforme Tabela 1.

Base de dados	Forma da pesquisa	Registos
Scopus	ALL (“augmented reality” AND “science teaching”) AND NOT (“virtual reality”) AND NOT (“university”)) AND PUBYEAR > 2010	78
	ALL (“realidad aumentada” AND “enseñanza de las ciencias”) AND NOT (“realidad virtual”) AND NOT (“universidad”)) AND PUBYEAR > 2010	12
	ALL (“réalité augmentée” AND “enseignement des sciences”) AND NOT (“réalité virtuelle”) AND NOT (“université”)) AND PUBYEAR > 2010	0
ScienceDirect	‘augmented reality’ AND ‘science%teaching’ NOT ‘virtual reality’ NOT ‘university’	142
	‘realidad aumentada’ + ‘enseñanza%ciencias’ - ‘realidad virtual’ - ‘universidad’	14
	‘réalité augmentée’ + ‘enseignement%sciences’ - ‘réalité virtuelle’ - ‘université’	8
Springer Link	(‘augmented reality’ AND ‘science teaching’) AND NOT (‘virtual reality’ OR ‘university’)	394
	(‘realidad aumentada’ AND ‘enseñanza de las ciencias’) AND NOT (‘realidad virtual’ OR ‘universidad’)	0
	(‘réalité augmentée’ AND ‘enseignement des sciences’) AND NOT (‘réalité virtuelle’ OR ‘université’)	0
EBSCO	“augmented reality” AND “science teaching” NOT “virtual reality” NOT “university”	54
	“realidad aumentada” AND “enseñanza de las ciencias” NOT “realidad virtual” NOT “universidad”	8
	“réalité augmentée” AND “enseignement des sciences” NOT “réalité virtuelle” NOT “université”	0
Web of Science	((((ALL=(“Augmented reality”)) AND ALL=(“Science teaching”)) NOT ALL=(“Virtual reality”)) NOT ALL=(“university”)) AND PY=(2010-2022)	9
	((((ALL=(“realidad aumentada”)) AND ALL=(“enseñanza de las ciencias”)) NOT ALL=(“realidad virtual”)) NOT ALL=(“universidad”)) AND PY=(2010-2022)	0
	((((ALL=(“réalité augmentée”)) AND ALL=(“enseignement des sciences”)) NOT ALL=(“réalité virtuelle”)) NOT ALL=(“université”)) AND PY=(2010-2022)	0

Tabela 1 – Bases de dados e operadores num primeiro processo de revisão sistemática da literatura

2.2. Resultados da pesquisa

Das cinco bases de dados consultadas: *Scopus*, *ScienceDirect*, *Springer Link*, *EBSCO* e *Web of Science*, foram devolvidas um total de 719 artigos, aos quais adicionamos cinco do *Scholar Google*, por se considerarem igualmente relevantes, resultando num total de 724 artigos, conforme Figura 1.

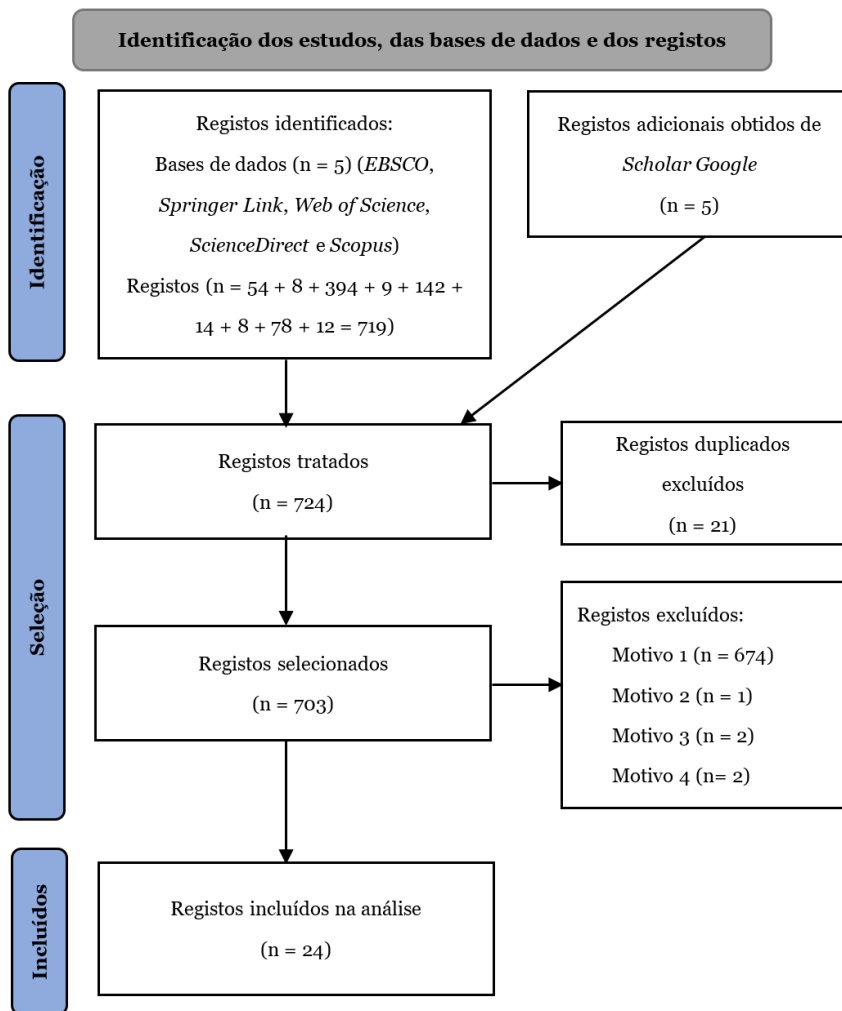


Figura 1 – Aplicação da metodologia PRISMA na revisão sistemática da literatura

2.3. Critérios de inclusão e exclusão

Na seleção dos artigos procedemos primeiramente à retirada de 2.90% (n = 21) por serem publicações duplicadas, tendo-se seguido a exclusão de 93.74% (n = 674) das

publicações após leitura do título, das palavras-chave e do resumo, pois não integravam referências conjugadas a RA e CN (ou Biologia e/ou Geologia) e Ensino Básico ou Ensino Secundário (equivalente ISCED 2 e 3) (motivo 1); 0.14% ($n = 1$) por não apresentar a estratégia de ensino no trabalho com os alunos (motivo 2); 0.28% ($n = 2$) por incluírem a realidade virtual no estudo (motivo 3) e, por último, 0.28% ($n = 2$), uma meta-análise que não incluía estratégias educativas e integrava o ensino universitário e a outra uma revisão da literatura sobre a indústria alimentar (tema específico que não faz parte dos programas). Para este artigo, foram selecionadas 10 publicações, por serem as que apresentaram resultados na aprendizagem dos alunos do ensino básico e secundário, nas disciplinas de Ciências Naturais e/ou Biologia (Tabela 2).

Autor(es) e ano	Metodologia do Estudo	País	N (GC + GE)	Contexto geral	Duração	app e sistema (com/sem marcador)	Idade
Gopalan et al. (2016)	Experimental	Malásia	140 (70 + 70)	Sala de aula	Pré-teste 20 minutos; Uma semana de trabalho; pós-teste 20 minutos	Aplicação eStar*, computador (com marcador)	13 e 14 anos
Hung et al. (2016)	Experimental	Taiwan	72 (24 + 24) + 24)	Sala de aula	Nd	Aplicação construída, óculos de RA (com marcador)	12 e 13 anos
Say et al. (2017)	Métodos mistos explanatório	Turquia	53 (27 + 26)	Sala de aula	Um mês; pré e pós-teste	Aplicação construída, tablet (com marcador)	12 e 13 anos
Karagozlu (2018)	Experimental	Não definido	147 (70 + 77)	Sala de aula	14 semanas de conhecimento pré e pós-teste	Aurasma app*, smartphone (com marcador)	12 e 13 anos
López (2018)	Quase-Experimental	Colômbia	13 (0 + 13)	Sequência didática de aprendizagem	Nd	Arloon Anatomy AR app, smartphone (com marcador)	10 e 11 anos
Sahin et al. (2019)	Quase-Experimental	Turquia	100 (50 + 50)	Sala de aula	Quatro semanas (16 horas); um teste de avaliação	app não definida, computador (com marcador)	12 e 13 anos

Autor(es) e ano	Metodologia do Estudo	País	N (GC + GE)	Contexto geral	Duração	app e sistema (com/sem marcador)	Idade
Gnidovec et al. (2020)	Quase-Experimental	Eslovénia	51 (0 + 51)	Sala de aula	junho (pré-teste, ensino e pós-teste), setembro (pós-teste e questionário)	Mirage AR9 app*, <i>smartphone</i> (com marcador)	13 e 14 anos
Petrov et al. (2020)	Estudo de caso	Bulgária	80 (0 + 80)	Sala de aula	Um ano letivo; pré e pós-teste	Aplicação zSpace, computador e óculos 3D (sem marcador)	Nd (<i>High school</i>)
Weng et al. (2020)	Quase-Experimental	Indonésia	68 (34 + 34)	Sala de aula	90 minutos (organizados em 20 + 5 + 45 + 20); pré e pós-teste	aplicativo construído e <i>smartphone</i> (com marcador)	14 e 15 anos
López-Cortés et al. (2021)	Pré-Experimental	Chile	162 (0 + 162)	Sala de aula	Nd	Aplicação Mitotic Division, <i>smartphone</i> ou <i>tablet</i> (com marcador)	15 e 16 anos

Nota: GC - grupo controlo; GE - grupo experimental; * app descontinuada; Nd – não definido.

Tabela 2 – Metodologia do estudo, país, amostra, duração e sistema de RA utilizado

Estes 10 artigos utilizaram diferentes *designs* de investigação: quase-experimental (4 estudos), experimental (3 estudos), pré-experimental (1 estudo), métodos mistos (1 estudo) e estudo de caso (1 estudo), como ilustrado na Figura 2.

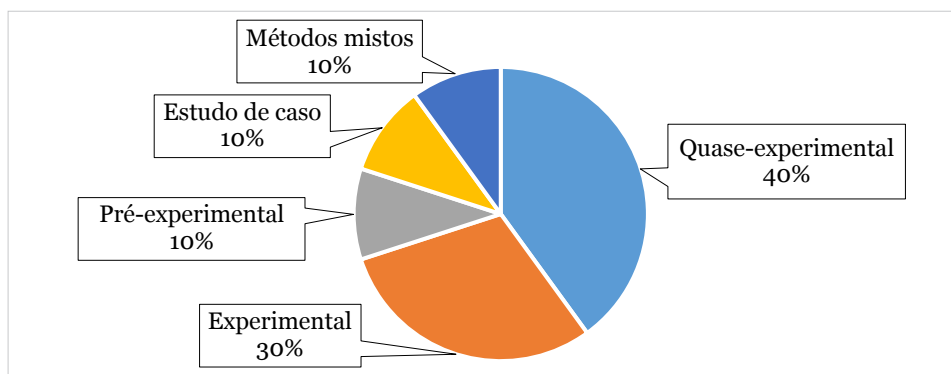


Figura 2 – Metodologia dos estudos analisados.

Com exceção de um artigo (Petrov et al, 2020) em que os autores não explicitam a idade dos alunos, verificámos um maior número de estudos na faixa etária 12-13 anos (4 estudos), sendo seguida pelas idades 13-14 anos (2 estudos) e, por fim as idades 10-11 anos, 14-15 anos e 15-16 anos (1 estudo cada), como ilustrado na Figura 3. Podemos encontrar em Rosser (1994, citado por Hung et al., 2017) a referência ao desenvolvimento das capacidades de visualização mental que ocorrem até ao final da infância.

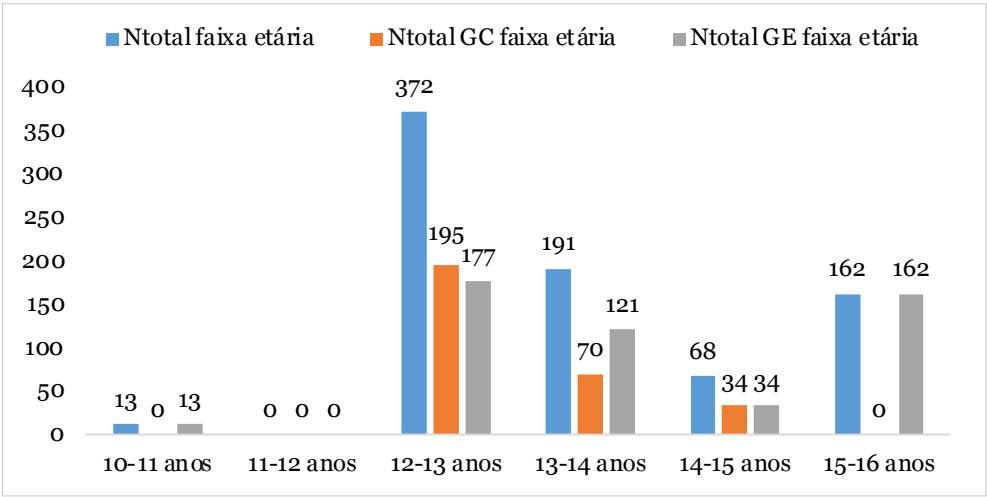


Figura 3 – Número de alunos envolvidos nos estudos analisados, por faixa etária.

Procedeu-se à descrição das questões de pesquisa relacionadas à aprendizagem ou ao desempenho dos alunos e às conclusões dos autores, conforme Tabela 3.

Autor(es) (ano)	Questões de investigação ou Objetivo do estudo	Instrumentos de recolha	Conclusões
Gopalan et al. (2016)	Os alunos que utilizam o eSTAR têm, eventualmente, uma melhor pontuação no pós-teste e que resulta numa melhor aprendizagem, em comparação com aqueles que usaram a aprendizagem convencional de ciências.	Pré e Pós-teste de conhecimentos.	... através da utilização da aplicação eSTAR, os alunos do grupo experimental têm uma melhor pontuação no pós-teste e eventualmente um melhor desempenho da aprendizagem, quando comparados com os que seguiram uma aprendizagem convencional de ciências.

Autor(es) (ano)	Questões de investigação ou Objetivo do estudo	Instrumentos de recolha	Conclusões
Hung et al. (2016)	Pode a RA melhorar o desempenho da aprendizagem dos alunos em comparação com os materiais didáticos convencionais - livros ilustrados e interações físicas? Pode a RA criar mais interesse/motivação nos alunos do que os materiais didáticos convencionais - livros ilustrados e interações físicas?	Teste de erro – <i>error test</i> (avaliação imediata) e teste de retenção – <i>retention test</i> (avaliação posterior).	Os resultados das análises podem responder à primeira pergunta de pesquisa: o livro gráfico com RA oferece uma maneira prática e de “mãos na massa” para as crianças explorarem e aprenderem sobre as bactérias, e a RA melhora o desempenho da aprendizagem dos alunos numa medida semelhante aos materiais didáticos convencionais, como livros ilustrados e interações físicas. Os resultados mostraram que, afinal, a RA produziu benefícios educacionais semelhantes aos materiais didáticos tradicionais.
Say et al. (2017)	... o estudo procurou investigar o impacto da RA nas atitudes e no sucesso dos alunos.	Pré e Pós-teste de conhecimentos.	Verificou-se que os níveis de aproveitamento académico dos alunos aumentam mais com a educação proporcionada pelos cartões de RA. (...) De todos ..., pensa-se que a RA seja um suporte eficaz no ensino e deve ser utilizado.
Karagozlu (2018)	O objetivo deste estudo é determinar o impacto de uma aplicação de realidade aumentada concebida para o ensino de ciências do 7º ano, sobre o desempenho e as competências de resolução de problemas dos alunos	Pré e Pós-teste de conhecimentos.	... as pontuações de desempenho pós-teste dos alunos no final do semestre mostraram que houve aumento nas pontuações de desempenho dos alunos em ambos os grupos, mas que o aumento no grupo experimental, que foi apoiado pela tecnologia RA, foi maior.
López (2018)	Avaliar o processo de apropriação alcançado pelos alunos durante a implementação da sequência didática.	Pré e Pós-teste de conhecimentos.	Os resultados da sequência didática... mostraram que este tipo de estratégia tem, de facto, um impacto positivo nos processos de ensino-aprendizagem. Por exemplo, um dos desafios no ensino do corpo humano e da sua anatomia é a impossibilidade de realmente observar as características e funcionamento, de modo que há uma desconexão entre a teoria e a compreensão prática. (...) Nesse sentido, confirmou-se a hipótese que sugeria que as TIC e o uso de ferramentas como a Realidade Aumentada ajudaram a enfrentar essa rutura na aprendizagem.

Autor(es) (ano)	Questões de investigação ou Objetivo do estudo	Instrumentos de recolha	Conclusões
Sahin et al. (2019)	Existe uma diferença significativa entre o desempenho académico dos alunos do ensino secundário que utilizam aplicações de RA e as que utilizam métodos convencionais?	Teste de desempenho do curso de Ciências	Observou-se uma diferença significativa entre o grupo experimental e o grupo controlo no desempenho académico, sendo que os alunos que tiveram as aulas de ciências com RA obtiveram pontuações mais altas num teste de desempenho, no curso de ciências, do que aqueles que aprenderam por meio de métodos convencionais. ... Em conclusão, é evidente que a RA afeta positivamente o desempenho académico dos alunos.
Gnidovec et al. (2020)	Como os materiais educativos baseados em RA ajudam os alunos a entender o sistema circulatório humano?	Pré e Pós-teste de conhecimentos e follow-up de conhecimentos após férias escolares.	A atividade instrucional neste estudo melhorou significativamente o conhecimento dos alunos sobre a anatomia básica e fisiologia do coração humano. Além disso, o pós-teste tardio mostrou que a mudança conceitual alcançada na compreensão dos alunos foi estável durante um período de dois meses. Os resultados... confirmam a primeira questão de investigação, que os materiais educativos baseados em RA ajudaram os alunos a compreender a complexidade do sistema circulatório humano ...
Petrov et al. (2020)	Este artigo tem como objetivo explorar o impacto de uma ferramenta de RA no desempenho de aprendizagem dos alunos.	Pré e Pós-teste de conhecimentos.	O estudo mostrou uma melhoria substancial na compreensão dos alunos sobre o assunto estudado. (...) Com base na análise dos dados sobre a eficácia da aprendizagem através de ambientes aumentados e dos testes escritos, pode-se afirmar que a integração da RA na referida unidade teve efeitos significativos na aprendizagem dos conteúdos.
Weng et al. (2020)	Qual é o efeito do uso de RA num livro de biologia nos resultados da aprendizagem dos alunos?	Pré e Pós-teste em aula e entrevista a três alunos de cada grupo	O estudo descobriu que a tecnologia de RA melhora os resultados de aprendizagem no nível de análise na taxonomia de Bloom, mas não melhora significativamente os resultados de aprendizagem nos níveis de memorização e compreensão. Assim, a tecnologia de RA pode ter o potencial de ser útil no desenvolvimento das habilidades cognitivas de ordem superior dos alunos.

Autor(es) (ano)	Questões de investigação ou Objetivo do estudo	Instrumentos de recolha	Conclusões
López- Cortés et. al. (2021)	Avaliar as representações dos estudantes acerca do fenómeno mitótico através da análise de explicações sobre uma situação quotidiana (cicatrização de uma ferida). Detetar alterações nos níveis de representação dos estudantes que participam no estudo ao trabalhar com uma sequência que promove a visualização através do uso de realidade aumentada.	Pré e Pós teste com a avaliação por rubrica do desempenho	... possibilita inferir que a realidade aumentada permite que os alunos observem um fenómeno abstrato e descontextualizado de diferentes pontos de vista, favorecendo a seleção e utilização de esquemas mentais anteriores para processar novas informações, otimizando a memória de trabalho na busca de novas conceções. (...) É por isso que consideramos que o recurso como um todo (sequência e app) oferece oportunidades de aprendizagem eficazes, ativas e ubíquas (Cabero et al., 2019), geradoras de representações mais complexas e robustas na explicação de um fenómeno complexo da biologia, como a divisão mitótica.

Tabela 3 – Questões de investigação (ou objetivos) presentes nos artigos e resultados relacionados do uso da RA na aprendizagem/desempenho dos alunos

3. Resultados

A análise da Tabela 3 mostra que 9 dos 10 estudos analisados indicaram melhoria nos resultados/desempenho académico nos alunos que utilizaram a RA em contexto de aprendizagem, quando comparados com o desempenho dos alunos que utilizaram o método convencional. O estudo desenvolvido por Weng et al. (2020) refere que não houve diferença nos resultados/desempenho académico dos alunos, nos níveis de memorização e compreensão, mas apenas no nível de análise, tendo como referencial a taxonomia de Bloom (1956) e destacando o potencial da RA no desenvolvimento dos processos cognitivos de ordem superior. Um outro estudo de Hung et al. (2016) refere que não se verificou diferença no desempenho académico dos alunos, mas que os resultados obtidos são consonantes com a literatura ao considerarem vantajosa a RA na melhor perceção de conceitos ou processos 3D (visuoespaciais). Ressaltamos o estudo de Gnidovec et al. (2020) que apresenta como limitação o facto de não existir um grupo de controlo. No estudo de Karagozlu (2018) a autora aponta que a RA favoreceu a resolução de problemas e o autocontrolo dos estudantes.

4. Discussão

O número de estudos desenvolvidos em meio educativo não superior com integração da RA tem vindo a aumentar mas, conforme referem Carrilho e Vera (2022), a RA é ainda uma tecnologia inovadora e emergente. Estes autores mencionam Toledo e Garcia (2017) que destacam que a RA possibilita trazer para a aula a tridimensionalidade e a interatividade que são aspetos únicos e importantes para a aprendizagem de certas matérias. Crespo e Marín (2020, citam Martin et al., 2011, e Martin-Gutierrez, 2017)

para destacarem os desafios educativos que a RA introduz nos processos de ensino e aprendizagem. Referem que melhora a motivação dos estudantes, que se constitui como uma oportunidade de desenvolver novas estratégias de ensino e de aprendizagem e de estar a tornar-se comum em dispositivos acessíveis.

Com exceção de um estudo (Petrov et al., 2020) que utilizou um sistema de computador, que os autores apontam como uma limitação devido ao alto custo do sistema, em todos os outros foram utilizados *smartphones* ou *tablets*, aspeto que torna o uso da RA mais acessível.

Na nossa análise encontramos estudos que utilizaram aplicativos criados para conteúdos ensinados e analisados na investigação, mas que já não existem no mercado como disponíveis, o que não permite a replicação da atividade com novos grupos de alunos. Outras aplicações continuam a existir, mas apenas uma parte da sua utilização pode ser explorada gratuitamente (López, 2018) e outras continuam a existir e permitem a exploração plena e livre (López-Cortés et al., 2021).

Os autores são unânimes em indicar que os alunos estão mais ligados à atividade da aula quando a RA está presente. Este resultado aparece em todas as investigações realizadas previamente a este artigo e em outras revisões sistemáticas da literatura onde foi usada a RA.

Salientamos que o impacto positivo da RA na aprendizagem dos alunos é referido através de uma maior proximidade entre o conhecimento teórico e dimensões não visíveis da realidade (Nincarean et al., 2013; Say et al., 2017; López, 2018; Gnidovec et al., 2020; López-Cortés et al., 2021). Este aspeto ajuda os alunos a aprender estruturas anatómicas internas (López, 2018), aspetos complexos (Gnidovec et al., 2020), a melhorar o desenvolvimento de habilidades cognitivas de ordem superior (Weng et al., 2020), a otimizar a memória de trabalho, gerando representações mentais mais complexas e robustas (López-Cortés et al., 2021) e a apoiar a aprendizagem significativa (Petrov et al., 2020).

Em síntese podemos referir:

1. Em resposta à nossa Q1, ser unânime nos estudos analisados que a aprendizagem foi favorecida ou que ocorreu um melhor desempenho dos alunos, ainda que Hung et al. (2016) refiram que não se verificou diferença no desempenho académico dos estudantes que integraram o seu estudo. Contudo, entendemos destacar o estudo de Weng et al. (2020) que detalham o potencial da RA no desenvolvimento dos processos cognitivos de ordem superior, de acordo com a taxonomia de Bloom (1956);
2. Em resposta à nossa Q2, podemos listar os diversos aspetos que os autores destacam como vantagens na integração da RA em contextos de ensino e aprendizagem, sem que a ordem em que apresentamos os estudos traduza a sua maior ou menor importância. A RA:
 - 2.1. Permite uma maior proximidade ao conhecimento teórico e a dimensões não visíveis dos fenómenos em estudo (Nincarean et al., 2013; Say et al., 2017; López, 2018; Gnidovec et al., 2020; López-Cortés et al., 2021);

- 2.2. Melhora a perceção de aspetos complexos dos conteúdos em análise (Gnidovec et al., 2020; López-Cortés et al., 2021);
- 2.3. Motiva os alunos, promove o seu bem-estar e uma aprendizagem mais aprazível (Hung et al., 2016; Say et al., 2017);
- 2.4. Fornece feedback imediato (López-Cortés et al., 2021)
- 2.5. Permite desenvolver competências viso-espaciais (Say et al., 2017);
- 2.6. Possibilita a compreensão prática do não visível (Lópes, 2018);
- 2.7. Apoia a aprendizagem significativa (Petrov et al., 2020);
- 2.8. Pode ser usada de forma gratuita (López, 2018).

5. Conclusão

É comum a todos os estudos analisados que a RA se revela, no processo de ensino e aprendizagem das Ciências Naturais e/ou Biologia, como um aspeto inovador e motivador, promovendo maior atenção e envolvimento dos alunos. Destaca-se também como um meio complementar à utilização de outras ferramentas, nomeadamente no ensino das ciências (STEM), dada a possibilidade de tornar visível o não visível. Os aspetos mencionados convergem, na opinião dos autores dos estudos, a levar a uma maior motivação, envolvimento e entusiasmo na aprendizagem, conduzindo os alunos a sentirem uma maior confiança no que estão a realizar.

Referências

- Bloom, B. S. (Ed.) (1956). *Taxonomy of educational objectives, Handbook 1: Cognitive domain*. New York: Longmans.
- Carrilho, J., & Vera, W. (2022). Realidad aumentada en aplicaciones móviles educativas. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (E47), 77-94. <http://www.risti.xyz/issues/ristie47.pdf>
- Crespo, V., & Marín, L. (2020). La realidade aumentada y el aprendizaje basado en problemas como estrategia que soportan los Living Lab. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (E32), 572-582. <http://www.risti.xyz/issues/ristie32.pdf>
- Gnidovec, T., Žemlja, M., Dolenec, A., & Torkar, G. (2020). Using augmented reality and the structure-behavior-function model to teach lower secondary school students about the human circulatory system. *Journal of Science Education and Technology*, 29, 774-784. <https://doi.org/10.1007/s10956-020-09850-8>
- Gopalan, V., Zulkifli, A., & Bakar, J. (2016). A learning performance study between the conventional approach and augmented reality textbook among secondary school students. *AIP Conference Proceedings*, 1761(1), 020029-1-020039-5. <https://doi.org/10.1063/1.4960879>

- Hung, Y.-H., Chen, C.-H., & Huang, S.-W. (2016). Applying augmented reality to enhance learning: a study of different teaching materials. *Journal of Computer Assisted Learning*, 33, 252-266. <https://doi.org/10.1111/jcal.12173>
- Karagozlu, D. (2018). Determination of the impact of augmented reality application on the success and problem-solving skills of students. *Quality & Quantity*, 52, 2393-2402. <https://doi.org/10.1007/s11135-017-0674-5>
- López, J. (2018). Apropiación de la realidad aumentada como apoyo a la enseñanza de las ciencias naturales en educación básica primaria. *Revista Boletín Redipe*, 7(12), 144-157. <https://revista.redipe.org/index.php/1/article/view/655>
- López-Cortés, F., Moreno, E., Palmas-Rojas, C., & Rubilar, C. (2021). Niveles de representación externa de estudiantes de Educación Secundaria acerca de la división celular mitótica: una experiencia con realidad aumentada. *Pixel-Bit Revista de Medios y Educación*, 62, 7-37. <https://doi.org/10.12795/pixelbit.84491>
- Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, 77(12), 1321-1329. https://web.cs.wpi.edu/~gogo/courses/cs525H_2010f/papers/Milgram_IEICE_1994.pdf
- Naranjo, J., Robalino-López, A., Alarcon-Ortiz, A., Peralvo, A., Romero, R., & Garcia, M. (2021). Sistema de realidade aumentada para la enseñanza de matemática en tiempos de COVID-19. RISTI - *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E42), 530-541. <http://www.risti.xyz/issues/ristie42.pdf>
- Nincarean, D., Ali, M. B., Halim, N. D. A., & Rahman, M. H. A. (2013). Mobile augmented reality: The potential for education. *Procedia-Social and Behavioral Sciences*, 103, 657-664. <https://doi.org/10.1016/j.sbspro.2013.10.385>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Aki, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372(71), 1-9. <http://dx.doi.org/10.1136/bmj.n71>
- Parong, J. (2021). Multimedia Learning in Virtual and Mixed Reality. In R. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (pp. 498-509). Cambridge University Press. <https://doi.org/10.1017/9781108894333.051>
- Petrov, P., & Atanasova, T. V. (2020). The effect of augmented reality on students' learning performance in stem education. *Information*, 11(4), 209-219. <https://doi.org/10.3390/info11040209>
- Sahin, D., & Yilmaz, R. M. (2020). The effect of Augmented Reality Technology on middle school students' achievements and attitudes towards science education. *Computers & Education*, 144(103710), 1-11. <https://doi.org/10.1016/j.compedu.2019.103710>

- Say, S., & Pan, V. (2017). The effect of instruction with augmented reality astronomy cards on 7th grade students' attitudes towards astronomy and academic achievement. *The Turkish Online Journal of Educational Technology, (Special Issue for INTE 2017)*, 295-301. <https://files.eric.ed.gov/fulltext/ED600996.pdf>
- Singh, A. (2019). Virtual reality, augmented reality and mixed reality: Human-computer interaction. Publicação independente
- Skarbez, R., Smith, M., & Whitton, M. C. (2021). Revisiting Milgram and Kishino's Reality-Virtuality Continuum. *Frontiers in Virtual Reality*, 2(647997), 1-8. <https://doi.org/10.3389/frvir.2021.647997>
- Zheng, S. (2015) Research on Mobile Learning Based on Augmented Reality. *Open Journal of Social Sciences*, 3, 179-182. <http://dx.doi.org/10.4236/jss.2015.312019>
- Weng, C., Otanga, S., Christianito, S., & Chu, R. (2020). Enhancing students' biology learning by using augmented reality as a learning supplement. *Journal of Educational Computing Research*, 58(4), 747-770. <https://doi.org/10.1177/0735633119884213>

Sistemas Automáticos de Calificación de Tareas de Programación: Revisión y Crítica

Carlos Cares¹, Ricardo Gacitúa¹

carlos.cares@ceisufro.cl; ricardo.gacitua@ufrontera.cl

¹ Departamento de Ciencias de Computación e Informática, Universidad de La Frontera, Temuco 4811230, Chile

DOI: 10.17013/risti.50.58-72

Resumen: La corrección de las tareas de programación de computadoras es una actividad crítica y de alto consumo de tiempo para los académicos. Como respuesta a este problema, se han propuesto diversos sistemas de retroalimentación y calificación automática. Este artículo presenta una revisión de sistemas y enfoques de corrección y calificación automatizada de tareas de programación con el objetivo de ofrecer un mapa general de las funciones de estos sistemas y sus desafíos. Los resultados muestran que existe una importante diferencia entre la industria y la academia, así como una carencia en marcos de evaluación para estas herramientas.

Palabras-clave: Programación; enseñanza; calificación automática; retroalimentación; resultados de aprendizaje.

Automatic Grading Systems for Programming Assignments: Review and Critique

Abstract: Reviewing programming assignments has become a time-consuming and critical activity for academics. Various feedback and automatic grading systems have been proposed in response to this problem. This article presents a review of various systems and automated feedback approaches for grading programming assignments to provide an overview of the functions of these systems and their challenges. The results show a difference between the tools proposed by the academy and those proposed by the industry, as well as a lack of evaluation frameworks,

Keywords: Programming; teaching; automatic grading; feedback; learning outcomes.

1. Introducción

La programación de computadores se trata de componer y organizar un conjunto de instrucciones en un lenguaje de programación con el objetivo de resolver un problema de proceso de información. El desarrollo de esta habilidad es frecuentemente percibido como complejo por los estudiantes dado que la habilidad es compuesta y jerárquica (Ismail et al., 2010).

Por otra parte, la habilidad de programación presenta alta demanda. Algunas universidades norteamericanas han doblado sus matrículas y se reconoce una férrea competencia entre la industria y las universidades por el contrato de doctores en informática. A este respecto se ha llegado a hablar de una demanda “desbordada”, caracterizada además por otros síntomas como loterías de cupos y largas listas de espera para los cursos de ciencias de computación (Singer, 2019).

De esta manera, los cursos masivos en línea (MOOC -Massive Open Online Courses-) han comenzado a tener una alta relevancia, no sólo por la masividad de la respuesta, sino también debido a la pandemia de Covid-19, que ha aumentado y fortalecido la percepción de satisfacción por un proceso de enseñanza en línea (Bianchi et al., 2022; Blinkin, 2022).

En este contexto, la enseñanza de la programación de computadores requiere por parte de los profesores una alta dedicación a la corrección y calificación de las soluciones de programación de los estudiantes, lo que constituye una barrera relevante para la escalabilidad de los cursos. A su vez, los tiempos de demora impactan negativamente en la percepción de los estudiantes respecto de la calidad del curso (Ismail, 2021). Debido a esta necesidad histórica, pero además acentuada por la pandemia y aumento de la demanda, es que la asistencia automática en la corrección y calificación se hace relevante en la enseñanza de la programación (Daradoumis, 2013; Aldriye, 2019).

Adicionalmente, un tercer elemento es el desarrollo de las habilidades metacognitivas, lo cual requiere conciencia respecto de “qué es lo que sé”, y del “cómo aprendo” (Armstrong, 1989). Las habilidades metacognitivas conforman parte de los objetivos de la enseñanza y, en particular, la corrección automática de programas, la identificación de errores y de desvíos en términos de métricas de programación, constituyen habilitadores para el autodiagnóstico estudiantil.

Un sistema de corrección automática (y calificación) de tareas de programación (CAP) podría incluir el reconocer qué programa respuesta contiene errores, si logra pasar todas las situaciones de prueba definidas, si la solución responde a todos los requerimientos de la tarea asignada, así como también si ofrece realimentación que habilite el aprendizaje. Los requerimientos para el producto podrían, por lo tanto, ser variados y complejos, y la forma de desarrollar y mostrar los resultados también. Poy y Gonzales-Aguilar (2014), muestran diversas debilidades del sistema como objetivos pedagógicos no alcanzados, y certificación a personas que no realizan las actividades interactivas. Ante la diversidad de logros y debilidades se ha llegado a afirmar que el problema siempre se ha abordado parcialmente (Hass, Yuan & Li, 2019).

El objetivo de este trabajo de investigación es analizar los sistemas automatizados de calificación, mostrar sus tipos y las herramientas existentes, tanto las propuestas por la academia como aquellas provenientes de la industria. Asimismo, establecemos una crítica técnico-pedagógica basada en las carencias que hemos percibido, y proponemos oportunidades de mejora.

La estructura del artículo es la que sigue, la sección 2 presenta los fundamentos básicos para caracterizar los sistemas de calificación automática de tareas de programación de computadoras. La sección 3 presenta la metodología general de la revisión y la base del enfoque de la investigación. La sección 4 presenta y caracteriza los resultados del estudio.

La sección 5 presenta una breve discusión respecto de las debilidades encontradas, así como del trabajo futuro.

2. Antecedentes

Debido al aumento en el uso de tecnologías digitales y la masificación de los cursos masivos en línea, se han introducido diversos cambios que afectan los procesos de enseñanza-aprendizaje. Uno de estos cambios se refiere a la evaluación automática de tareas.

Evaluar tareas de programación de computadores manualmente conlleva una serie de problemas, entre ellos la inexactitud de la evaluación (Douce, 2005), así como el esfuerzo y el tiempo requerido, especialmente en cursos con gran cantidad de estudiantes. La evaluación automática se ha convertido en un método que suple estos problemas permitiendo no sólo calificar los ejercicios de programación de los estudiantes, sino también entregar información sobre la calidad de las soluciones.

La naturaleza de la programación puede considerar constructores y lógicas diferentes para su solución. En efecto, diferentes constructores y lógicas pueden producir el mismo resultado bajo un desarrollo diferente en forma y fondo. Adicionalmente, el tipo de problema pedido podría fácilmente caer en categorías donde no existen límites finitos para comprobar una solución satisfactoria (Budd & Angluin, 1982). Esta situación teórica agrega dificultad para determinar la corrección de una tarea de programación, haciendo el proceso lento y, además, proclive a errores.

Hass, Yuan & Li. (2019) han resumido las ventajas y las influencias entre el proceso de enseñanza-aprendizaje y el trabajo de la evaluación automatizada. En términos de enseñanza-aprendizaje la autora reconoce que (a) la evaluación no sesgada, (b) la realimentación es oportuna, y (c) existe posibilidad de autoaprendizaje. A su vez, en relación a las tareas de evaluación reconoce una (a) mejor disponibilidad, (b) reducción de la carga de trabajo, (c) mejor distribución y asignación de tareas, (d) un sistema verificable de entrega de tareas, (e) posibilidad de análisis de código y verificación de manera previa a la entrega por parte de estudiantes además de la asistencia a profesores. En este artículo hemos separado las ventajas reconocidas diferenciando las perspectivas del estudiante y del docente, y hemos agregado aquellas ventajas relacionadas al almacenamiento de los elementos, como por ejemplo el reuso de comentarios y casos de prueba. La Figura 1 ilustra el resultado.

Desde el punto de vista de la ingeniería del software, un programa es correcto si produce la salida esperada para cada entrada posible. Sin embargo, las posibilidades son más amplias, de este modo mientras algunos sistemas sólo detectan errores en una solución, otros permiten explicar estos errores e incluso ofrecer sugerencias de posibles mejoras (Striwe, 2009).

Los errores lógicos, también llamados semánticos, se refieren a aquellos errores que, aún cuando el compilador no arroje errores, provocan que la salida del programa no sea la esperada. Esto puede deberse a diferentes factores: mala comprensión del problema, errores lógico-matemáticos, desborde de las representaciones numéricas finitas, entre otros.



Figura 1 – Ventajas de la revisión automática de programas de computadora. Una extensión a la propuesta de Hass, Yuan & Li (2019)

Los errores de tiempo de ejecución aparecen cuando el programa está funcionando y lleva a que el programa deje de ejecutarse. Por ejemplo, debido a una comprobación de límites de la pila de sucesiones de invocaciones, excepciones de punteros a direcciones de memoria inaccesibles o inexistentes, fuentes de datos no encontradas o no disponibles, entre varios otros factores.

En términos generales, una herramienta automatizada de calificación puede considerar dos enfoques principales: (1) un análisis estático, donde el objetivo es analizar cómo está representada la solución en el correspondiente lenguaje de programación y (2) un análisis dinámico, donde el objetivo es analizar cómo se comporta la solución.

De este modo, el análisis estático funciona sin que se requiera la ejecución del código fuente, y sin embargo, se requiere del código para poder analizarlo. En el análisis dinámico, el código fuente se compila y se ejecuta mediante casos de prueba con diferentes datos de entrada para producir determinados datos de salida. Por definición, los sistemas capaces de realizar análisis dinámicos de código pueden realizar también análisis estáticos. Si la funcionalidad abarca estos dos tipos de análisis entonces son denominados sistemas híbridos.

3. Metodología

La verificación automática de programas de computador para propósitos de enseñanza académica continúa siendo un desafío, puesto que es posible evaluar programas simples automáticamente pero, para tareas más complejas, aún presentan limitaciones. Cuando se trata de la comprobación dinámica de errores, la situación es aún peor. Hass, Yuan & Li (2019) reconocen que para algunos casos se dispone de prototipos de investigación rudimentarios y sus prestaciones no se han probado a fondo ni han resultado satisfactorias. En la mayoría de los casos, sólo se permite un lenguaje de programación o un número limitado de tipos de trabajo.

Considerando, por tanto, la falta de claridad acerca de las características efectivas de estas herramientas automatizadas de calificación, y la carencia de evidencia que muestre el real uso de estas herramientas en la práctica, parece razonable establecer una caracterización de diferentes sistemas de calificación de estudiantes y presentar evidencia de su uso.

Ante la existencia de herramientas de carácter comercial y otras herramientas que son prototipos académicos, se utiliza una estrategia de revisión cualitativa de la literatura, la cual aborda la realidad desde una perspectiva holística e intenta comprenderla o describirla sin recurrir para ello a formular hipótesis, establecer medidas objetivas, o controlar exhaustivamente todas las variables, pero sí corresponde a una metodología de tradición crítica, exponiendo los hechos bajo calificaciones sociales (Sandín-Esteban, capítulo 7, 2003). Por el mismo motivo, este estudio no tiene pretensiones predictivas sino de generar los fundamentos y comunicación de las cualidades a considerar.

El objetivo de este trabajo de investigación es entonces describir las características principales de los sistemas automatizados de calificación de tareas de programación y recoger evidencia sobre el nivel de interés, como medida indirecta, de su uso. En término de las fuentes de datos, se recurre a bases de datos académicas y comunidades de usuarios de Internet para determinar las herramientas que son consideradas para su posterior análisis. En la Figura 2 ilustramos las fases del trabajo: los círculos son procesos y los rectángulos las entradas y salidas de información.

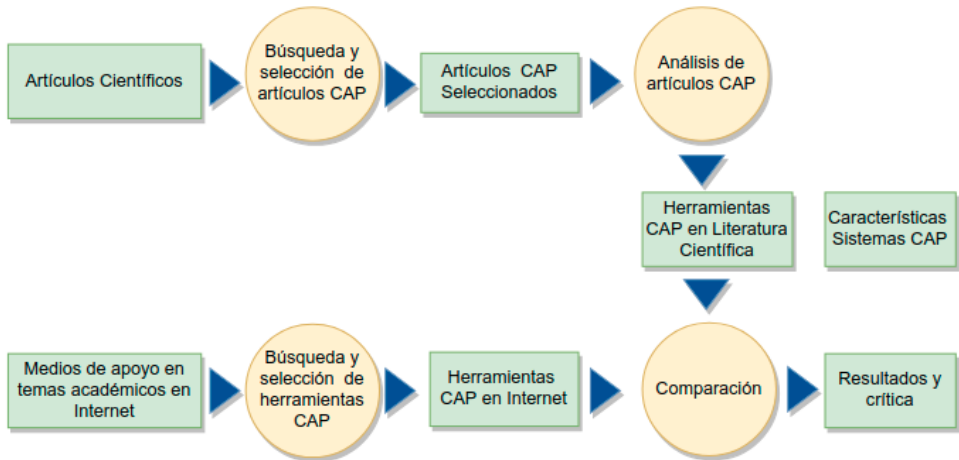


Figura 2 – Fases seguidas para generar la comparación y crítica

El protocolo descrito en la Figura 2 tiene por objeto permitir la descripción de características de las herramientas mencionadas en la literatura y determinar evidencias de uso. Adicionalmente, y considerando que Storey et al. (2014) plantean que existe un conocimiento colectivo en las redes sociales y comunidades de usuarios. Así, recopilamos evidencia por este medio. En este caso, se consideran dos fuentes de información, los foros de la red social Researchgate (<https://www.researchgate.net>), la cual es una red de

académicos y en cuyos foros es posible seguir el hilo de una conversación, de modo que las interacciones conforman una medida del interés en el tópico. La segunda fuente en Internet considerada es Youtube (<https://www.youtube.com/>), la cual es una red para compartir vídeos. En el caso de esta segunda red, es visible la cantidad de veces que un vídeo se ha activado, constituyendo así una medida de la popularidad del material disponible (Maraza-Quispe, 2020).

4. Ejecución del Estudio

4.1. Búsqueda y Selección de Artículos

La primera fase de nuestra revisión cualitativa de literatura comprende la definición de las fuentes de información. Considerando la existencia de algunas revisiones sistemáticas publicadas en la literatura, las cuales listan sistemas automatizados de calificación de tareas de programación, se considera como primera fuente de información este tipo de metaanálisis. La cadena de búsqueda se construyó a partir de la Tabla 1. Los términos dentro de una fila se conectan con “OR” y los términos dentro de la columna se conectan con “AND”.

Nº	Términos de Búsqueda
1	Revisión, Mapeo. Metaanálisis, estudio de casos, ensayo
2	Calificación, Evaluación
3	<i>Tarea, Asignación</i>
4	Programación

Tabla 1 – Términos de Búsqueda en la Revisión Bibliográfica

La cadena de búsqueda se aplicó en las bibliotecas digitales de IEEEExplore, ACM, y en la base de datos Scopus. Debido al bajo número de artículos ampliamos a una búsqueda general en Google Scholar. Estas son bases de datos conocidas y contienen información actualizada de artículos de revistas y conferencias disponibles a cualquier lector. Se revisaron las entradas para su adecuación e inclusión en el estudio. Se consideró, como criterio de inclusión, sólo revisiones sistemáticas, mapeos sistemáticos y metaanálisis. Los años considerados fueron aquellos entre 2015 y 2021. El resultado de la búsqueda permitió identificar cinco artículos, los cuales son presentados en la primera columna en la Tabla 2.

4.2. Análisis de los Estudios de Revisión

En el contenido de estos estudios se identificaron 52 sistemas de graduación automática o semi-automática. Sin embargo, en cinco casos, el sistema se identificó sólo por medio de los nombres de sus autores, es posible asumir entonces que estos productos no tienen un nombre o identificación comercial, o al menos no la tenían en el momento de la publicación. En 47 de estos estudios sí hubo un nombre de producto identificado. Las coincidencias las hemos resumido en la Tabla 2. De este modo es posible ver, por

ejemplo, que el producto “Web-Cat” es analizado en 4 de los 5 artículos seleccionados. El sistema denominado “OL Judge” es analizado en 3 de los 5 artículos. El resto de las coincidencias corresponden al mínimo, es decir, sólo 2 artículos mencionando el mismo producto. Los productos sobre los cuales coinciden los autores son sólo 8 de los 52. Cualitativamente podemos afirmar que tenemos una baja coincidencia entre los autores. Parte de esta diferencia podría estar en los años de los artículos, lo que indicaría que varios productos no tuvieron continuidad.

Nº	Referencia								
		Autograder	AutoLEP	Boss	Mooshak	OL judge	Web-Cat	ASSYST	ASys
P1	Souza, Felizardo, & Barbosa (2016)	☒	☒	☒	☒	☒	☒	☐	☐
P2	Gupta and Gupta (2018)	☒	☒	☒	☒	☒	☒	☒	☐
P3	Aldriye , Alkhalaf & Alkhalaf (2019)	☐	☐	☐	☐	☐	☒	☒	☒
P4	Hass, Yuan & Li (2019)	☐	☐	☐	☐	☐	☒	☐	☐
P5	Ismail and Lakulu (2021)	☐	☐	☐	☐	☒	☐	☐	☒

Tabla 2 – Coincidencia en los productos analizados

4.3.Características de los Productos de Calificación Automática

En estos cinco estudios seleccionados se describen los productos de calificación automática productos de su análisis comparativo. En estas descripciones algunos categorizan las funcionalidades y otros simplemente describen los productos. El supuesto que realizamos es que las características descritas son aquellas que resultan relevantes para los académicos que realizan los estudios, de este modo, asumimos que las características no nombradas no se consideran relevantes.

La aplicación automática de pruebas es una característica transversalmente reconocida, pero, al mismo tiempo, la generación de estos casos de prueba no es considerada una característica para evaluar.

Otro tipo de características, clasificadas generalmente (pero no siempre) como análisis estáticos, se trata de patrones de comparación o métricas de código relacionadas a orientación a objetos (clases y métodos), a métricas de código, o calificación de patrones estructurales, por ejemplo que lo esperado como respuesta es un algoritmo con máximo una condición dentro de un ciclo. Otros elementos comunes a calificar son comentarios y estilos de codificación. En la Tabla 3 hemos resumido un conjunto de características reconocidas por los estudios.

4.4. Medios de Apoyo a la Academia

La relevancia de los medios de comunicación digital y de las redes sociales en particular, conforman un recurso para compartir dudas, y obtener respuestas. Storey et al. (2014) han reconocido la relevancia de blogs, foros, y otras redes sociales como fuente de información práctica, masiva y accesible. En la red Researchgate encontramos un foro con la pregunta: “What tools do you use for automated grading of assignments that involve programming?”. Las respuestas van desde el año 2013 hasta el año 2017. Siete herramientas son mencionadas en las diferentes respuestas. Nos dimos a la tarea de buscarlas para verificar que aún existían. Sólo en un caso la herramienta no fue posible de recuperar. El resultado de los hallazgos los mostramos en la Tabla 4.

Funcionalidades Reconocidas		Souza, Felizardo, & Barbosa (2016)	Gupta and Gupta (2018)	Aldriye , Alkhalaf & Alkhalaf	Hass, Yuan & Li (2019)	Ismail and Lakulu (2020b)
Pruebas		☐	☒	☒	☒	☒
	Errores	☒	☒	☒	☒	☐
	Pruebas de software	☒	☒	☒	☒	☒
	Pruebas unitarias (cuestionarios)	☒	☒	☒	☐	☐
	Competencia (comparativa)	☒	☐	☒	☐	☐
	Generación de casos de prueba	☐	☐	☐	☐	☒
Rendimiento		☒	☐	☒	☒	☐
Patrones		☐	☐	☒	☒	☐
	Clases / Funciones esperadas	☐	☐	☒	☒	☐
	Estructura de Programas	☐	☐	☒	☐	☒
	Estructuras similares (plagio)	☒	☐	☐	☐	☒
	Estilo de programación	☒	☒	☐	☒	☐
Métricas		☐	☐	☐	☒	☐
	Código (lineas, complejidad)	☒	☐	☐	☒	☐
	Comentarios	☒	☒	☐	☒	☐
Retro-alimentación		☒	☒	☒	☒	☒
	Corrección inmediata	☒	☐	☒	☒	☐
	Oportunidad de corregir	☒	☐	☐	☒	☐
	Entre estudiantes	☐	☐	☒	☐	☐
	Profesor	☒	☒	☐	☐	☒
Historial		☒	☐	☒	☒	☐
	Gestión recursos profesor	☒	☐	☒	☒	☐
	Revisión detallada estudiante	☒	☒	☐	☐	☐

Tabla 3 – Características comunes en los productos CAP

Año	Respuestas a uso de calificación automática de programas
2013	Web-CAT
2015	Stepic (encontrado como Stepik)
2015	Fitchfork (no encontrado)
2016	Inginious (encontrado como ECI Inginious)
2016	Autolab

Año	Respuestas a uso de calificación automática de programas
2016	Autogradr (encontrado en csforall.org)
2016	Stepik
2017	Web-CAT

Tabla 4 – Respuestas de productos CAP en Researchgate

Una segunda fuente muy común de material educativo, particularmente en la oferta de vídeos, es Youtube (youtube.com). Si bien en esta red los vídeos son de todo tipo, también ha sido reconocida como una fuente habitual de material documental (Maraza-Quispe et al. 2020). En Youtube se hizo la búsqueda de vídeos tutoriales de los programas identificados y, además, se realizó una búsqueda por las palabras claves previamente identificadas. Para cada producto se buscaron los primeros cinco vídeos más vistos y sumamos la cantidad de visualizaciones. En la Tabla 5, hemos resumido estos resultados. En el caso del producto Stepik, hemos dejado dos celdas en amarillo debido a que dichos vídeos están en ruso. Todo el resto está en inglés. No se encontró nada en español.

4.5.Comparación

Hemos realizado una revisión de estudios comparativos de herramientas de calificación automática de tareas de programación. A su vez, hemos buscado en redes de asistencia académica como Researchgate y otra de fuentes documentales como Youtube, sobre recomendaciones y tutoriales.

Herramienta	Vídeo 1	Vídeo 2	Vídeo 3	Vídeo 4	Vídeo 5	Total	Porcentaje
Gradescope	69,021	14,030	12,512	12,227	7,455	115,245	59.6%
Nbgrader	16,861	1,739	1,679	1,177	1,142	22,598	11.7%
Github Classroom	7,839	6,376	2,932	2,345	1,376	20,868	10.8%
Uva Online Judge	4,850	3,706	3,082	2,605	2,200	16,443	8.5%
Web-Cat	2,017	2,006	1,886	546	444	6,899	3.6%
Codegrade	2,559	751	684	167	125	4,286	2.2%
Autolab	866	859	424			2,149	1.1%
Mooshak	767	371	364	339	284	2,125	1.1%
Stepik	829	447	280	223	177	1,956	1.0%
ECI Ingenious	727	34	25	23	11	820	0.4%

Tabla 5 – Visitas en Youtube de tutoriales de productos CAP

Lo primero a destacar es que, dentro de los estudios, existe una baja coincidencia de las herramientas a considerar como parte del estudio. En todos los casos, las fuentes

de herramientas fueron otros estudios académicos donde se proponían herramientas CAP más que búsquedas en el mercado o repositorios de software. De los cinco estudios seleccionados, por ejemplo, no hay ninguna herramienta CAP que haya sido considerada en todos los estudios. Sólo una (Web-Cat) fue considerada en cuatro de los estudios seleccionados.

4.6.Comparación

Hemos realizado una revisión de estudios comparativos de herramientas de calificación automática de tareas de programación. A su vez, hemos buscado en redes de asistencia académica como Researchgate y otra de fuentes documentales como Youtube, sobre recomendaciones y tutoriales.

Lo primero a destacar es que, dentro de los estudios, existe una baja coincidencia de las herramientas a considerar como parte del estudio. En todos los casos, las fuentes de herramientas fueron otros estudios académicos donde se proponían herramientas CAP más que búsquedas en el mercado o repositorios de software. De los cinco estudios seleccionados, por ejemplo, no hay ninguna herramienta CAP que haya sido considerada en todos los estudios. Sólo una (Web-Cat) fue considerada en cuatro de los estudios seleccionados. La moda (valor más repetido) ha sido dos, es decir, lo habitual es que, entre los estudios académicos, la coincidencia es de sólo dos herramientas. En los estudios, existe sólo un 50% de coincidencia en las recomendaciones de foro, es decir, de las ocho herramientas CAP recomendadas en Researchgate, sólo cuatro habían sido mencionadas entre las 52 encontradas. En la Tabla 6 hemos resumido las coincidencias coloreando de verde las coincidencias entre tres fuentes y en amarillo las coincidencias entre dos.

Herramienta	Estudios (Total 5)	Recomendaciones (Total 8)	Visitas Youtube (Total 193,389)
Autograder	2/5	-	
AutoLEP	2/5	-	
ASys	2/5	-	
Boss	2/5	-	
Mooshak	2/5	-	2,125 (1%)
OL Judge	3/5	-	16,443 (9%)
Web-Cat	4/5	2/8	6,899 (7%)
ASSYST	2/5	-	
INGinious	1/5	1/8	820 (<1%)
Stepik	-	2/8	
Autolab	-	1/8	2,149 (2%)
Gradescope	-	1/8	115,245 (60%)
NBgrader	-	-	22,589 (23%)
Github Classroom	-	-	20,868 (21%)
Codegrade	-	-	4,286 (2%)

Tabla 6 – Coincidencias de herramientas

4.7. Crítica

El primer comentario que aparece obvio ante los resultados obtenidos es que estamos en frente de un área de estudio donde los objetos de estudio, es decir las herramientas de calificación automática, evolucionan rápidamente, y, como todo producto de software, se han movido los últimos años desde aplicaciones de escritorio a plataformas web, y también han evolucionado hacia el concepto de software como servicio. Sin embargo, muchas de las descripciones no hacen referencia básica a las características de los productos en cuanto a su plataforma tecnológica, ni forma de distribución o modelo de negocio si es que existe.

Adicionalmente, como parte de los hallazgos, varias de las herramientas mencionadas en los estudios no fueron encontradas en la revisión. Esto quiere decir que muy probablemente estas herramientas no tenían una versión para distribuir ni siquiera a nivel de prototipo experimental, menos como un producto o servicio. De este modo, el considerar un marco de referencia para la evolución desde prototipo a producto debería ser un requisito en la descripción de estos productos, pues no resulta comparable una idea inicial expresada en un prototipo a un producto maduro y con un equipo de desarrollo de software de tiempo completo. En este sentido, la propuesta original de la NASA para la evolución de los prototipos, que distingue nueve niveles, de TRL-1 a TRL-9, podría conformar una referencia conocida con este propósito (Olechowski, Eppinger & Joglekar, 2015).

Estos elementos particulares que hemos distinguido aquí, es decir, el tipo de tecnología de base, el modelo de negocio de la oferta del producto, y el estado de progreso, estabilidad o desarrollo del producto en sí, constituyen características de calidad de un producto de software. El conjunto de estas características, sus subcaracterísticas y métricas conforman lo que se denomina un modelo de calidad (Carvallo, Franch & Quer, 2006). Los modelos de calidad de un producto de software deberían estar basado en alguna referencia, como por ejemplo el estándar ISO/IEC 25000, que ha alcanzado un alto nivel de madurez como para considerarlo como referencia (Nistala, Nori & Reddy, 2019).

En este sentido, la Tabla 3, se representa una primera aproximación a la dimensión de calidad de la ISO/IEC 25010 denominada “Adecuación Funcional”. En esta dimensión de calidad, se debe considerar la calidad de completo, la calidad de correcto, y la calidad de pertinente. En particular, la calidad de pertinente, debería poder definirse y posteriormente calificarse sólo en la medida del correspondiente objetivo pedagógico. Selby (2015) caracteriza, dentro de la taxonomía de Bloom, las habilidades que pueden ser evaluadas mediante la realización de un programa, éstas van desde la comprensión básica de instrucciones de programación (comprensión) hasta alta competencias cognitivas como la capacidad de síntesis y evaluación. Es decir, el objetivo pedagógico, que podría establecerse en relación a la taxonomía de Bloom, cubre un amplio espectro de competencias. De este modo, las funcionalidades de las herramientas deberían poder mapearse a los resultados de aprendizaje que conforman el contexto pedagógico. En todos los estudios considerados, aparece que la competencia de programación como una competencia atómica (indivisible). Sin embargo, como se ha mostrado con anterioridad (Fuller et al. 2007), la programación manifiesta ciertos niveles de competencia diferenciada, y por lo tanto, es esperable que exista una relación entre el nivel de

competencia a evaluar y las funcionalidades de las herramientas. Esta división no es sólo teórica, sino que se ha llevado a la práctica con éxito en una prueba de programación concreta (Costa y Miranda, 2017). Estas posibles funcionalidades y formas de evaluación deberían ser parte de un modelo de calidad.

Otra dimensión relevante en un modelo de calidad es la seguridad. La aplicación de calificación automática debería tener suficiente seguridad como para que un estudiante no pueda, a pesar que lo intente, verificar las respuestas de otros estudiantes o los ejemplos de respuestas correctas provistos si existiesen. Una característica no funcional, como la seguridad, puede tomar forma en funciones no observables (como la encriptación), o puede tomar formas observables, como la implementación de una particular política de actualización y propiedad de las claves. Estas alternativas de calidad deberían también formar parte de las subcaracterísticas de calidad en un modelo de calidad difundido y compartido como el propuesto.

5. Conclusiones

En este artículo, hemos revisado los trabajos sobre programas que realizan una calificación automática de las tareas de programación de computadores. Hemos buscado estudios de revisión de estos tipos de programas y hemos encontrado cinco. Entre los cinco estudios, que van entre 2016 y 2021, se nombran un total de 52 aplicaciones. Alternativamente, hemos buscado en redes sociales para encontrar otras aplicaciones usadas. Hemos encontrado en la red académica Researchgate un foro de recomendaciones sobre aplicaciones de este tipo. La tercera fuente analizada fueron los vídeos tutoriales en Youtube.

El principal hallazgo es la baja coincidencia entre los tipos de estudios y las menciones en las redes sociales. En el caso de los estudios, además, hay una baja coincidencia entre ellos. La principal conclusión de estos hallazgos es que los estudios académicos consideran como fuente otros estudios académicos y productos de los cuales no existe antecedentes como para poder afirmar que hayan alcanzado algún nivel de madurez tecnológica.

Un segundo hallazgo es que las características que los estudios destacan, y las funcionalidades analizadas, no tiene un nivel de detalle y tampoco una referencia que permita poder establecer una comparación repetible y usable también por otras herramientas del mismo tipo.

En el contexto anterior, hemos establecido una crítica a esta área de estudios. La crítica tiene un eje pedagógico, y un eje tecnológico. El eje pedagógico dice relación con la baja conceptualización de la habilidad de programación y su relación a un contexto de resultados de aprendizaje bajo alguna gradualidad de progreso ampliamente aceptada, como por ejemplo la taxonomía Bloom. El segundo elemento de la crítica es el eje tecnológico. Hemos formulado que el conjunto de características a evaluar debería estar guiado por algún estándar de calidad, y hemos afirmado que el ISO/IEC 25.000 es adecuado

La principal limitación de este estudio, es que no hemos intentado levantar un catastro contemporáneo de las herramientas y servicios de calificación automática de tareas

de programación. Por lo tanto, el estudio se basa en descripciones de terceros, y cuando revisamos y resumimos las características, es un esfuerzo de síntesis sobre las características que otros han reconocido.

Con la limitación descrita, la contribución que se realiza, es una contribución al cómo se puede progresar en esta área de estudio, esto incluye (1) establecer una relación teórica entre el progreso cognitivo y habilidades de programación de computadores de los estudiantes, y las funcionalidades de la aplicación y (2) establecer un modelo de calidad que incluya no sólo los aspectos funcionales sino una gama de atributos de calidad ya reconocidos en estándares de software.

Referencias

- Aldriye, H., Alkhalaf, A., & Alkhalaf, M. (2019). Automated grading systems for programming assignments: A literature review. *International Journal of Advanced Computer Science and Applications*, 10(3). <https://doi.org/10.14569/IJACSA.2019.0100328>
- Armstrong, A. M. (1989). The development of self-regulation skills through the modeling and structuring of computer programming. *Educational Technology Research and Development*, 37(2), 69-76. <https://doi.org/10.1007/BF02298291>
- Bianchi, I. S., Daehn, C. M., Dávila, G. A., Tovma, N., & Shurenov, N. (2022). Cursos online Abertos Massivos (MOOCs) como potencializadores do conhecimento em Instituições de Educação Superior. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (48), 59-73. <https://doi.org/10.17013/risti.48.59-73>
- Blinkin, C. (2022, January 1). Dozens of U.S. colleges switching back to online classes as COVID-19 cases soar. *Global news*. <https://globalnews.ca/news/8483197/covid-us-colleges-online-classes/>
- Budd, T. A., & Angluin, D. (1982). Two notions of correctness and their relation to testing. *Acta informatica*, 18(1), 31-45. <https://doi.org/10.1007/BF00625279>
- Carvalho, J. P., Franch, X., & Quer, C. (2006). Managing non-technical requirements in COTS components selection. In *14th IEEE Int. Requirements Engineering Conf. (RE'06)* (pp. 323-326). IEEE. <https://doi.org/10.1109/RE.2006.40>
- Costa, J. M., & Miranda, G. L. (2017). Desenvolvimento e validação de uma prova de avaliação das competências iniciais de programação. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (25), 66-81. <https://doi.org/10.17013/risti.25.66-81>
- Daradoumis, T., Bassi, R., Xhafa, F., & Caballé, S. (2013). A review on massive e-learning (MOOC) design, delivery and assessment. In *2013 8th Int. Conf. on P2P, Parallel, Grid, Cloud and Internet Computing* (pp. 208-213). IEEE. <https://doi.org/doi.10.1109/3PGCIC.2013.37>
- Douce, C., Livingstone, D., & Orwell, J. (2005). Automatic test-based assessment of programming: A review. *Journal on Educational Resources in Computing*, 5(3), 4. <https://doi.org/10.1145/1163405.1163409>

- Fuller, U., Johnson, C. G., Ahoniemi, T., Cukierman, D., Hernán Losada, I., & Jackova, J., et al. (2007). Developing a computer science specific learning taxonomy. In Proc. of the 7th Conf. on Innovation and Technology in Computer Science Education (pp. 152-170). United Kingdom: ACM. <https://doi.org/10.1145/1345375.1345438>
- Gupta, S., & Anamika G.. (2017). E-Assessment Tools for Programming Languages: A Review. In Proc. of the 1st Int. Conf. on Information Technology and Knowledge Management (ICITKM'2017), New Delhi, India,(pp. 65-70).
- Hass, B., Yuan, C., & Li, Z. (2019). On the automatic assessment of learning outcomes in programming techniques. In 2019 IEEE 14th Int. Conf. on Intelligent Systems and Knowledge Engineering (ISKE) (pp. 274-278). IEEE. <https://doi.org/10.1109/ISKE47853.2019.9170370>
- Ismail, M. H., & Lakulu, M. M. (2021). A Critical Review on Recent Proposed Automated Programming Assessment Tool. Turkish Journal of Computer and Mathematics Education, 12(3), 884-894. <https://doi.org/10.17762/TURCOMAT.V12I3.799>
- Ismail, M. N., Ngah, N. A., & Umar, I. N. (2010). Instructional strategy in the teaching of computer programming: a need assessment analyses. The Turkish online Journal of Educational Technology, 9(2), 125-131.
- Maraza-Quispe, B., Alejandro-Oviedo, O., Fernández-Gambarini, W., Cisneros-Chavez, B., & Choquehuanca-Quispe, W. (2020). Análisis de YouTube como herramienta de investigación documental en estudiantes de educación superior. Publicaciones, 50(2), 133-147, <https://doi.org/10.30827/publicaciones.v50i2.13949>.
- Nistala, P., Nori, K. V., & Reddy, R. (2019, May). Software quality models: A systematic mapping study. In 2019 IEEE/ACM Int. Conf. on Software and System Processes (ICSSP) (pp. 125-134). IEEE. <https://doi.org/10.1109/ICSSP.2019.00025>.
- Olechowski, A., Eppinger, S. D., & Joglekar, N. (2015). Technology readiness levels at 40: A study of state-of-the-art use, challenges, and opportunities. In 2015 Portland Int. Conf. on management of engineering and technology (pp. 2084-2094). IEEE. <https://doi.org/10.1109/PICMET.2015.7273196>.
- Poy, R., & Gonzales-Aguilar, A. (2014). Factores de éxito de los MOOC: algunas consideraciones críticas. RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação, (E1), 105-118. <http://doi.org/10.4304/risti.e1.105-118>
- Sandín-Esteban, M.P. (2003). Investigación Cualitativa en Investigación en Educación: Fundamentos y Tradiciones. Madrid; Mc Graw and Hill.
- Selby, C. C. (2015). Relationships: computational thinking, pedagogy of programming, and Bloom's Taxonomy. In Proc. of the workshop in primary and secondary computing education (pp. 80-87). <http://doi.org/10.1145/2818314.2818315>.
- Souza, D. M., Felizardo, K. R., & Barbosa, E. F. (2016). A systematic literature review of assessment tools for programming assignments. In 2016 IEEE 29th Int. Conf. on Software Engineering Education and Training (CSEET) (pp. 147-156). IEEE. <http://doi.org/10.1109/CSEET.2016.48>

- Striewe, M., & Goedicke, M. (2009). Effekte automatischer Bewertungen für Programmieraufgaben in Übungs-und Prüfungssituationen. DeLFI 2009-7. Tagung der Fachgruppe E-Learning der Gesellschaft für Informatik eV.
- Storey, M.A., Singer, L., Cleary, B., Figueira Filho, F., Zagalsky, A., (2014). The (r) evolution of social media in software engineering. In Proc. of Future of Software Engineering (FOSE), 100-116. <https://doi.org/10.1145/2593882.2593887>.

Modelo de tendencias para series complejas: caso de estudio tarifas hoteleras

Abraham Briones Juárez¹, Judith Alejandra Velázquez Castro²

abrahambriones2003@gmail.com; judithalejandra666@gmail.com

¹ Universidad Autónoma del Estado de Hidalgo, ICEA, Turismo, 42160, Hidalgo México.

² Universidad Autónoma del Estado de Hidalgo, ICEA, Turismo, 42160, Hidalgo México.

DOI: 10.17013/risti.50.73-90

Resumen: El análisis riguroso de datos en turismo es cada vez más común a raíz de las afectaciones sufridas por el covid-19, lo que ha contribuido a identificar trayectorias a partir de datos históricos. En este artículo, se propone una combinación de la teoría de la complejidad y la suavización exponencial para conocer los patrones que ayuden a estimar la estacionalidad y tendencia en las tarifas hoteleras. Para ello, se fragmenta la serie en particiones para explorar la similitud y se aplican cinco métodos de trazado. Los resultados muestran que el gráfico de longitud R , es el único que despliega anti-persistencia, con una saturación en el tiempo $n=8$, este dato se usa como ciclo de longitud del modelo aditivo de Holt-Winters cuyos parámetros ($a=0.36$, $b=0.78$, $y=0.67$), minimizan el error presentado. Se concluye que la combinación de métodos de análisis numérico permite el ajuste estacional en los datos tarifarios.

Palabras-clave: Tarifas hoteleras, anti persistencia, minimización.

Trend Model for complex series: case study hotel rates

Abstract: In The rigorous analysis of data in tourism is increasingly common as a result of the effects suffered by covid-19, which has contributed to identifying trajectories based on historical data. In this article, a combination of complexity theory and exponential smoothing is proposed to understand the patterns that help to estimate the seasonality and trend in hotel rates. To do this, the series is fragmented into partitions to explore the similarity and five plotting methods are applied. The results show that the graph of the length R , is the only one that displays anti-persistence, with a saturation in time $n=8$, this data is used as the length cycle of the additive Holt-Winters model whose parameters ($a=0.36$, $b=0.78$, $y=0.67$), minimize the error presented. It is concluded that the combination of numerical analysis methods allows the seasonal adjustment in the tariff data.

Keywords: Hotel rates, anti-persistence, minimization.

1. Introducción

El turismo constituye una fuente de riqueza y desarrollo, pese a su vulnerabilidad por situaciones de conflictos sociales o emergencias sanitarias como el covid-19. La recuperación que muestra este sector, denota su resiliencia (Mendoza et al., 2021) y capacidad de adaptación, lo que en el mercado global se evidencia con el margen de estimación de ingresos que alcanzó 1.9 billones de dólares en el año 2021 a nivel global (NWTO, 2021). Sin embargo, la industria de la hospitalidad a nivel local puede presentar adversidades para incrementar la ocupación de cuartos. La ciudad de México por ejemplo reportó un margen de ocupación del 62% en periodo pre-pandémico (DATATUR, 2019), mientras que en tiempos de pandemia este dato oscilaba alrededor del 30% en el año 2021 (Camila, 2022). Por tanto, el éxito de la administración de estas empresas depende en gran medida de los medios analíticos para mejorar la previsión de datos, es decir trabajar sobre la precisión de las predicciones (Liu et al., 2014), lo que demanda un esfuerzo para desarrollar nuevos procedimientos matemáticos con la intención de identificar parámetros que se adapten a las necesidades del presente y con ello propicia el crecimiento y buen funcionamiento de las empresas al tiempo que se utilizan tecnología y sistemas de información que mejoren la toma de decisiones en la industria (Patiño et al., 2020).

La venta de servicios de hospedaje se realiza a través de la configuración tarifaria, lo que significa que los clientes reciben ciertos productos y servicios de acuerdo a los tipos de segmento a los que pertenecen (Liu et al., 2022), ejemplo de ello lo constituyen las condiciones particulares que integran el servicio (Somphong et al., 2022) como la ubicación del cuarto, el acondicionamiento, las amenidades y otros bienes de consumo. No obstante, el comportamiento de las tarifas hoteleras ha sido poco estudiado.

El análisis de datos en hotelería supone incluir aquellos aspectos que hacen variar las cifras de manera dinámica para determinar patrones a partir de la complejidad de los fenómenos. Los fractales denotan el comportamiento por puntos de cambio, que a su vez se pueden explicar por leyes de potencia. La teoría de la complejidad ha sido de gran utilidad para explicar situaciones en distintas ramas del conocimiento como la economía (Kachhia, 2023), la física (Hou et al., 2021), la mecánica (Peng, & Li, 2023), las ciencias sociales (Blanco, 2019) y en el turismo (Briones et al., 2011) por su gran capacidad para detectar patrones en series temporales.

Debido a que se examinará una serie que aparentemente es aleatoria se busca la repetición por escalas y a distintos niveles. Así, la compatibilidad estadística en fractales auto-afines, se presenta por un conjunto que muestra ser invariante por escalamiento. En este trabajo se propone desarrollar un análisis de tendencias futuras de series complejas en tarifas para determinar la estimación optimizada de parámetros. Para ello, se desarrollan las siguientes fases: 1) se recaban los datos con el sistema de ingresos de un hotel de la Ciudad de México. 2) Se realizan particiones de los valores por desviaciones estándar y se da a conocer el comportamiento de la serie a distintas ventanas de orden "n". 3) Se estiman los exponentes de Hurst y se presentan las ilustraciones log-log por método, para indicar el más apropiado que genera punto de estacionalidad y antipersistencia. 4) Se prueba la longitud del ciclo obtenida en el modelo Holt-Winters y se obtienen los parámetros optimizados. 5) Se presentan la gráfica de ajuste con los valores utilizados,

y la predicción de los primeros 8 datos de la serie. 6) Se vierten las conclusiones y recomendaciones finales.

2. Revisión de la bibliografía

El análisis de datos se encuentra ligado al uso tecnológico en empresas de hotelería no solamente para apoyar en las decisiones, sino para optimizar y maximizar los beneficios (Pimentel et al., 2019) ya que la información generada desde la gestión constituye una serie de datos que se gesta desde una estructura compleja (Hernández et al., 2021).

Algunos de los temas centrales en el manejo de cifras en la industria de la hospitalidad son los aspectos financieros (Ozdemir et al., 2021), la calidad de servicios (Nilashi et al., 2021) y el uso de los recursos (Iranmanesh et al., 2022). Sin embargo el análisis de datos toma relevancia significativa para revelar tendencias de comportamiento y con ello mejor la generación de fuentes de información confiables que direccionan la planeación de acciones en el servicio (Pan et al., 2012). Derivado de la gran necesidad de la industria turística se han empleado distintas herramientas para comprender, explicar y predecir la tendencia de datos en estas empresas, como se muestra en la tabla 1, que sintetiza algunos de los esfuerzos para adecuar modelos de previsión con base en datos recolectados.

Técnica	Explicación	Autores
<i>Modelos de de explicación automático</i>	<i>Combinación de modelos dinámicos de pronostico</i>	<i>(Pereira & Cerqueira, 2022)</i>
<i>Modelo de memoria de atención a corto y largo plazo</i>	<i>Usa datos de la demanda con características adicionales proporcionadas por una empresa de turismo.</i>	<i>(Kaya et al., 2022)</i>
<i>Modelo de aprendizaje profundo.</i>	<i>Usa correlaciones espaciales y temporales por medio de algoritmo bayesiano</i>	<i>(Huang & Zheng, 2021)</i>
<i>Modelo de previsión de cancelaciones en hotelería</i>	<i>Usa técnicas de redes neuronales optimizadas con algoritmos genéticos</i>	<i>(Sánchez-Medina & C-Sánchez, 2020)</i>
<i>Modelo de precios de las habitaciones</i>	<i>Usa modelo de promedio móvil, la máquina de Boltzman, el modelo de máquina de vector y el sistema de interferencia difusa de red adaptativa.</i>	<i>(Al Shehhi & Karathanasopoulos, 2020)</i>
<i>Modelo de pronóstico de series temporales</i>	<i>Usa el espacio de estado estacional</i>	<i>(Pereira, 2016)</i>
<i>Modelo de serie temporal con fuente de datos big data</i>	<i>Usa múltiples fuentes de big data</i>	<i>(Pan & Yang, 2017)</i>
<i>Modelo de redes neuronales</i>	<i>Redes con revisión de puntaje</i>	<i>(Chang et al., 2021)</i>

Tabla 1 – Métodos de previsión de datos en empresas turísticas

Pese, a que se ha incrementado el número y tipo de técnicas numéricas en la hospitalidad, existe muy poca evidencia sobre el análisis de tarifas hoteleras. En este sentido Kim

et al., (2020) desarrollaron un estudio de tarifas y la discrepancia entre los canales de distribución con la intención de conocer los criterios sobre los cuales se calculan los precios. Por su parte Lee (2011), aborda el cambio en las tarifas y subraya que el número de turistas y el desempeño económico se relacionan de forma positiva con su incremento. De igual manera Lim & Ok (2022), determinaron como las tarifas promocionales influían en las percepciones de los clientes. En contraste, la previsión de tarifas muestra poco interés por parte de los académicos y los analistas del turismo y la hotelería. Con el propósito de abonar a las herramientas de analíticas y dado que la pandemia de covid-19 ha incentivado el uso de la teoría compleja para realizar mediciones en la dinámica temporal de series de datos (Castillo & Melin, 2020) y debido a que su uso ha sido recurrente dado a que también describe la trayectoria de la enfermedad (Easwaramoorthy et al., 2021). En este trabajo se abordan los datos bajo la teoría de la complejidad, la cual ha sido utilizada en turismo para dar explicar la dinámica de datos a partir del dinamismo del sector (Baggio & Sainaghi, 2011;Baggio & Sainaghi, 2016;Baggio, 2014).

La intención de refinar los modelos de pronóstico en la industria recae en el hecho de constituir herramientas que apoyen en la toma de decisiones gerenciales en la industria (Teixeira & Gunter, 2023). En tal sentido, han sido de gran utilidad las técnicas Box-Jenkins en su forma SARIMA y ARMA (Chuwang & Chen, 2022); los modelos de series de tendencias estructurales para identificar variables y dentar una mejora en la previsión de datos (Rashad, 2022); la combinación de técnicas de inteligencia artificial y enfoques convencionales que dan pie a modelos híbridos (Ouassou & Taya, 2022) y el uso de arquitecturas de redes neuronales con caracterización de periodos (Nguyen et al., 2022).

Estos esfuerzos tienen la intención de constituir métodos técnicas y herramientas que inducen modelos aplicables a la industria. (Gunter, 2021) especialmente en tiempos de pandemia en los que se vio afectado el flujo de turistas y de manera directa las comunidades residentes en los que se enclavan los atractivos turísticos (Godovkykh et al., 2021).

3. Metodología

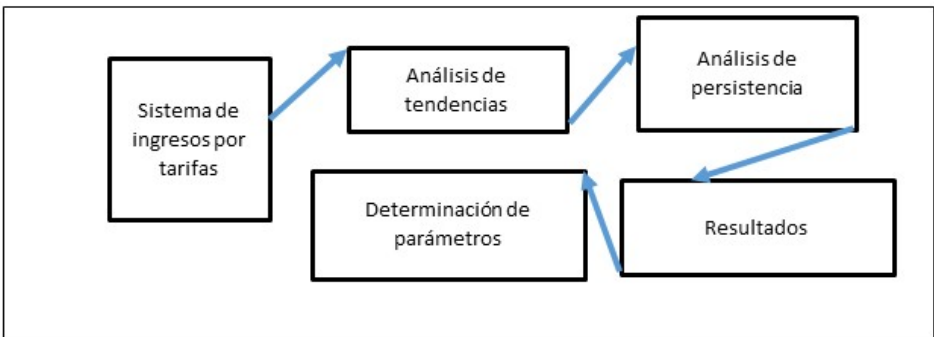


Figura 1 – Metodología

A continuación se presenta la metodología que sigue este trabajo en la figura 1, en la cual se puede percibir las fases que sigue la investigación. Cabe señalar que el cálculo de los parámetros ayudó en el desarrollo de datos futuros a partir del reconocimiento de las tendencias en los datos. Bajo esta consideración, enseguida se hace la descripción de los apartados para comprender el proceso que se siguió en este artículo.

3.1. Sistema de ingresos por tarifa

En esta fase primero se compilaron los datos a partir del sistema central de información del hotel (PMS) Property management system por sus siglas en idioma inglés, en un hotel categoría cinco estrellas. En esta empresa se asigna una clave a cada tipo de tarifa para poder otorgarla a los clientes. El gerente de recepción gestiona el otorgamiento de las tarifas según corresponda y tiene la facultad junto con la gerencia en turno de integrar servicios adicionales como llamadas telefónicas, alimentos, amenidades y promociones. Estos aspectos en general representan relaciones numéricas que cambian el total de ingresos por cada tarifa y la suma total de los datos que se observan.

El sistema central de información permite obtener datos por códigos y conceptos desglosados. El concentrado de la participación, su definición y el tipo de tarifa aparece en la tabla 2, que además muestra el porcentaje de participación de cada tipo de tarifa. Partiendo de la base anterior, se toma los datos de entrada del modelo, mismos que se constituyen con la sumatoria de las tarifas registradas en cada una de las observaciones que contabilizaron 974 de ellas.

Tipo de tarifa	Concepto	Porcentaje de representación
Corporaciones	Tarifa de descuento por ocupaciones	7%
Regular	Se brinda en recepción para clientes sin reservaciones	12%
Contratos	Se ofrece a partir de garantizar cierto número de estancias al año	19%
Gobierno	Tarifa preferencial a miembros diplomáticos y otros gobernantes	12%
Promocional	Se determina por la temporalidad	5%
Paquete	Incluye elementos de paquete	4%
Descuento	Se ofrece en temporada baja y fines de semana	6%

Tabla 2 – Tipología de tarifas del hotel.

3.2. Análisis de tendencias

Con el conjunto de datos se procedió a desarrollar el gráfico 2, para observar la dinámica a través del tiempo de los datos históricos de la serie y analizar su comportamiento. A primera vista, la serie parece aleatoria. Sin embargo bajo un análisis cualitativo se observan los siguientes aspectos:

Se observan oscilaciones periódicas, sin embargo es difícil precisar el número cíclico que describa la tendencia y la estacionalidad de la serie.

Se muestra que en el mes de abril del año 2022, no contabilizan ingresos por tarifas ya que el establecimiento permaneció cerrado en estas fechas debido a la contingencia sanitaria por covid-19.

Los datos mantienen una tendencia ascendente.

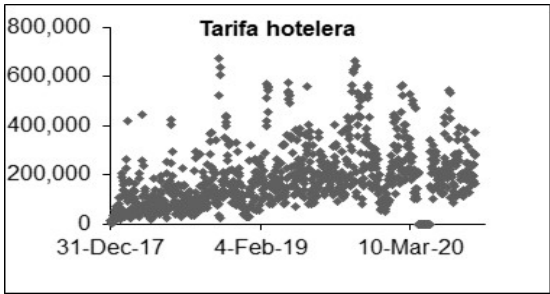


Figura 2 – Serie temporal de tarifas hoteleras

Para detectar la existencia de patrones repetidos se realiza una fragmentación de la serie en ventanas a partir de las desviaciones estándar de los datos; es decir, para la ventana 1, se calculó (sigma) del dato 1 al dato 2 (σ_1,σ_2), y así de manera sucesiva hasta completar 100 datos (σ_n,σ_{n+1}). Lo mismo se realizó con las ventanas sucesivas ampliando este cálculo a partir del dato 110 hasta 500, con conjuntos de 10 datos ($\sigma_{101},\sigma_{111}$).y de 500 a 1000 con conjuntos de 50 datos. La tabla 3, muestra el procedimiento y las ventanas que se toman como muestra para analizar las tendencias.

Fecha	Datos	Ingresos por Tarifas			
01/01/2018	1	13,377.99	n=10		
02/01/2018	2	5,851.44	17.0315625	n=20	
03/01/2018	3	3,003.84	61.036875	0.375507812	n=30
04/01/2018	4	3,588.48	207.6428125	0.033203125	7.69043E-05
05/01/2018	5	3,669.12	197.1746875	0.255791016	0.000500488

Tabla 3 – Ventanas de muestra de la serie tarifas hoteleras.

Se tomó una muestra de 3 observaciones con la misma distancia de n , para visualizar si existe similitud, el origen de la muestra depende de las ventanas analizadas y de la comprobación de las tendencias del comportamiento gráfico en los datos. Para este caso se comprueba similitud en las fluctuaciones de las ventanas seleccionadas al observar las figuras 3, 4 y 5 respectivamente.

En la comparación de la las figuras 2 y 3, se preservan los picos de ascenso en los datos máximos.

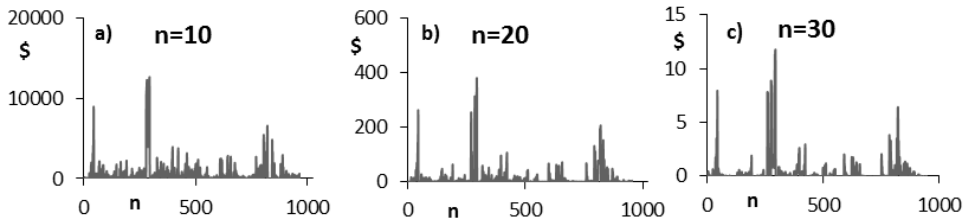


Figura 3 – Ventanas de tamaño a) $n=10$, b) $n=20$ y c) $n=30$.

Con las figuras 3, se puede apreciar que a diferentes valores de ventana n , se guarda una similitud palpable en las tendencias de la serie. Esta partición de datos responde a la necesidad de visualizar el cambio relativo en cada iteración para valorar la pertinencia de desarrollar un análisis que induzca la identificación de parámetros a partir de patrones similares a distintas recursiones.

3.3. Análisis de persistencia

Con la certeza de que se puede encontrar una base de comportamiento que describa la serie, se realiza un análisis de persistencia de los datos a través del exponente de Hurts o exponente H . Lo anterior, para determinar si existe estacionalidad. Para ello, se utiliza el software Benoit 1.31, dado que la serie muestra aleatoriedad a primera vista, se desarrollan las gráficas log-log del exponente con distintos métodos de tratamiento auto-afín. Primero se empleó el método R/S o rango reescalado, el cual arroja un valor H , que mide la intensidad de dependencia a largo plazo de las series temporales.

$$\frac{R}{S} = \frac{R(n)}{S(n)} = \frac{\max\{0, r_t : t = 1, 2, 3, \dots, n\} - \min\{0, r_t : t = 1, 2, 3, \dots, n\}}{\sqrt{S^2(n)}} \quad (1), \text{ con}$$

$$r_k = \sum_{t=1}^k X_t - \frac{k}{n} \sum_{t=1}^n X_t \quad (2), \text{ y}$$

$$S(n) = \left[\frac{1}{n} \sum_{t=1}^n \left(X_t - \frac{1}{n} \sum_{t=1}^n X_t \right)^2 \right]^{\frac{1}{2}} \quad (3)$$

Donde: $X = \{X_t; t = 1, 2, 3, \dots, n\}$, y $R(n)$ es normalizada por la desviación estándar $S(n)$. Así, el rango rescalado se define por n^H , donde $n \rightarrow \infty$, en una serie con traza auto-afín.

En la figura 4, se observa una estacionalidad cercana al 0.5 por lo tanto, bajo este método la serie sigue siendo aleatoria y no se observan persistencias que favorezcan la previsión de datos, esto debido a que los exponentes H, vierte estacionalidad alrededor del punto medio =0.5.

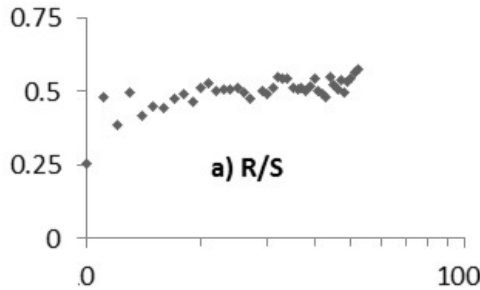


Figura 4 – Estacionalidad aleatoria del método R/S.

En segundo lugar se utilizó la longitud de rugosidad o método R/L que se mide como el valor de la raíz cuadrada media del residuo en una tendencia lineal ajustada a los puntos de muestra en una ventana de longitud w , entonces la rugosidad de la raíz cuadrada media se calcula como sigue:

$$RL(w) = \frac{1}{n_w} \sum_{i=1}^{w_n} \sqrt{\frac{1}{m_{i-2}} \sum_{j \in w_i} (z_j - z)^2} \quad (4)$$

Donde n_w es el número total de ventanas de longitud w , m_i es el número de puntos en la ventana, w_i, z_j es el residual en la tendencia y z es la media residual en la ventana w_i .

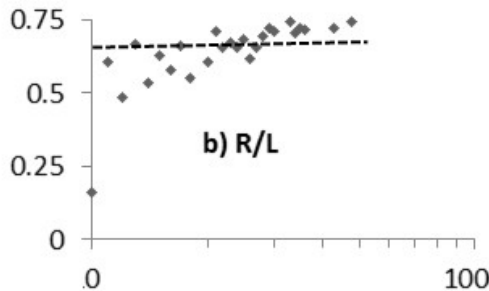


Figura 5 – Comportamiento estacionario del exponente H con el método R/L

En la figura 5, se observa que con el método R/L se estacionaliza la serie en el dato 8, esto se presenta con valores de H alrededor de 0.70, lo que significa que existe persistencia traducida en la visualización del exponente H por encima del promedio, este gráfico ilustra un posible parámetro para precisar el comportamiento de la serie.

De la misma forma el análisis incluyo el Método de espectro de potencia P/S, el cual se relaciona con el exponente H a través de “ b ” como la pendiente negativa cuya forma matemática se constituye por:

$$D_s = \frac{5-b}{2} \quad (5)$$

Con este análisis se observa en la figura 6, que a partir del dato 22 la fórmula arroja datos cero para continuar con tendencia alcista. Por tanto, los valores del exponente H , no presentan estacionalidad alguna.

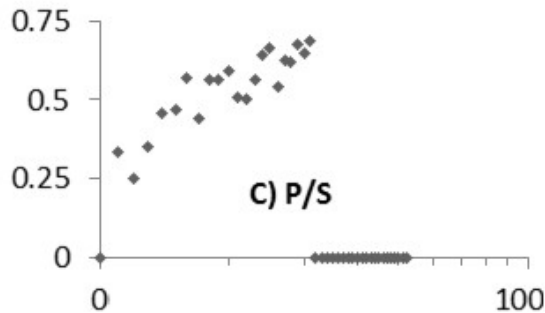


Figura 6 – Tendencia alcista de H con el método P/S.

El siguiente método a valorar fue el variograma o V/R definido como el cuadrado medio del incremento de puntos, estimado de la siguiente manera:

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^n [V(x_i) - Vx_{i+h}]^2 \quad (6)$$

Donde h , es la distancia de seguimiento (distancia entre dos puntos sucesivos), $\gamma(h)$ es el variograma a distancia de seguimiento h (el número de pares a una distancia h de retraso) y $V(x_i)$ son los valores de la muestra en la ubicación x_i , entonces las distribuciones del fractal se caracterizan por un modelo de variograma de la siguiente forma:

$$\gamma(h) = \gamma_0 h^{2H} \quad (7)$$

Donde H , es el exponente de Hurst. Los métodos de rugosidad-longitud y variograma se basan en el comportamiento de la escala de la desviación estándar $S_n \propto (\tau)^{H_n}$ y la semi-varianza $Var \propto \tau^{2H}$.

Este método muestra en la figura 7, datos oscilando alrededor $H = 0.25$ y con datos ceros después de la observación 27, lo que denota incapacidad para identificar previsiones.

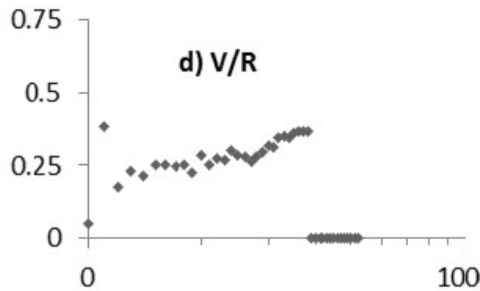


Figura 7 – Estacionalidad anti-persistente del método de variograma

Finalmente se valoró el método de veletas el cual se asocia con un coeficiente:

$$W \propto a^{H_n + 1/2} \quad (8)$$

Donde a es el un parámetro que mide la onda.

En este último análisis en la figura 8, solo aparecen los datos por debajo de 0.75, sin mostrar estabilidad ni punto de saturación.

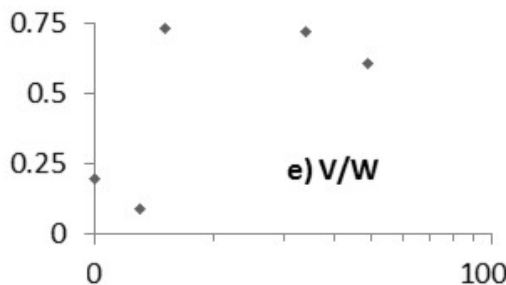


Figura 8 – Dispersión de datos por el método de veletas.

3.4. Resultados

Con las fases desarrolladas hasta ahora se presentan a continuación los resultados observados.

La manera en la que están constituidas las tarifas hoteleras y la flexibilidad que tiene la gerencia para modificarlas además de la injerencia de fenómenos externos como el impacto del covid-19 generan comportamiento caótico en la serie difícil de caracterizar.

La serie muestra auto-afinidad de tendencias al ser observada mediante el análisis de desviaciones estándar.

La serie de datos de las tarifas hoteleras, solo se estaciona en el dato $H = 8$ en el método de longitud de rugosidad o L/R , en que además se muestra persistencia entre los datos, por tanto se tiene posibilidad de estimar parámetros que expliquen el comportamiento de los datos.

Dado que el dato de saturación se puede interpretar como un ciclo de longitud de la serie se realiza una prueba para estimar parámetros en el del modelo aditivo de Holt-Winters. Lo anterior, debido a que la estacionalidad es constante, lo cual se realiza en el siguiente apartado.

3.5. Determinación de parámetros optimizados

Los parámetros principales del método de Holt-Winters están constituidos de la siguiente forma:

α = Constante que actúa como atenuante para el promedio de datos ($0 < \alpha < 1$)

β = Constante que actúa como atenuante de las tendencias estimadas ($0 < \alpha < 1$)

γ = Constante de atenuación que actúa en la estacionalidad ($0 < \gamma < 1$)

Para calcular el valor atenuado A_t , se usa la siguiente relación:

$$A_t = \alpha \frac{Y_t}{S_{t-L}} + (1 - \alpha)(A_{t-1} + T_{t-1}) \quad (9)$$

Donde

Y_t = Son las observaciones

S_t = Estacionalidad

L = Longitud de estacionalidad

T_t = Tendencia

Para calcular la tendencia T_t , se utiliza la siguiente relación:

$$T_t = \beta(A_t - A_{t-1}) + (1 - \beta)T_{t-1} \tag{10}$$

Para calcular el valor pronosticado Y_{t+p} se utiliza la siguiente fórmula:

$$Y_{t+p} = (A_t + pT_t)S_{t-L+p} \tag{11}$$

Donde p = Periodos a pronosticar

Con la primer parte del análisis se toma $L= 8$ asumiendo con ello que cada 8 datos se repite el ciclo de comportamiento de los datos, con tendencia creciente excepto en el periodo de cierre por covid-19 de acuerdo con la figura 2. Para medir la adaptabilidad de los datos se desarrollan los siguientes procesos:

Se utilizan parámetros iniciales para el modelo.

Se realiza una optimización con la herramienta de solver de EXCELL.

Se realiza una optimización con el software winQSB.

Se vierten las conclusiones.

En la tabla 4, se muestran los parámetros iniciales y los optimizados por técnica analítica.

Parámetros	Iniciales	Optimizados	
<i>L</i>	8	<i>Solver</i>	<i>WINQSB</i>
<i>Alfa (α)</i>	0.1	0.43433065	0.36
<i>Beta (β)</i>	0.2	0.83944777	0.78
<i>Gamma (γ)</i>	0.7	0.25185388	0.67

Tabla 4 – Parámetros del modelo

A partir de la información de la tabla 2, se prueban los parámetros iniciales obtenidos con la herramienta solver de Excell y los obtenidos con el software winQSB. Estos últimos remarcados en negritas en la tabla ya que minimizan el error del modelo de acuerdo con los siguientes coeficientes obtenidos.

De la misma manera, se vierten los coeficientes de errores mínimos en la tabla 5, la cual muestra: la desviación absoluta media (MAD) “Meand absolute desviation”, el error medio cuadrático (MSE) “mean squared error” y la media absoluta por error (MAPE).

Coefficientes de error	Dato
<i>MAD</i>	7866.99
<i>MSE</i>	8.43
<i>MAPE</i>	42.52

Tabla 5 – Coeficientes de error mínimo del modelo

En el mismo sentido se procedió a calcular la predicción de los primeros 8 datos de la serie cuyos resultados se ilustran en la tabla 6 y en la gráfica 9.

Primera serie pronosticada	Dato		
$S(1)$	159642.575	$S(5)$	137442.295
$S(2)$	115612.975	$S(6)$	115525.905
$S(3)$	144109.575	$S(7)$	129278.515
$S(4)$	138241.935	$S(8)$	130956.785

Tabla 6 – Primera predicción de la serie

Finalmente se presenta el comparativo del ajuste de los parámetros encontrados en la figura 9, comparados con los datos históricos de la serie. En ella se observa que el uso de los parámetros mantiene un ajuste considerable con la serie original, lo cual muestra las bondades del modelo para predecir datos en el futuro. Cabe señalar que el pronóstico se puede calcular por bloques de 8 datos como aparece en la figura 9 para obtener datos futuros que contengan la relación interna de la serie original.

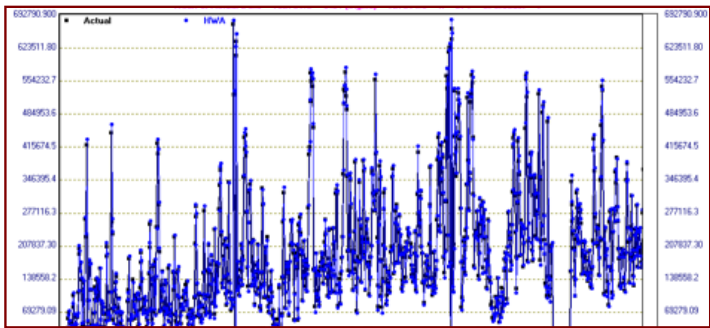


Figura 9 – Ajuste del comportamiento de la base de datos tarifas hoteleras, actual y modelo aditivo de Holt-Winters.



Figura 10 – Pronóstico de los primeros 8 datos de la serie.

4. Conclusiones

La previsión de datos en turismo y hotelería es una de los aspectos importantes dado que la emergencia de covid-19 ha generado impactos directos en la forma de estimar valores que apoyen los cursos de acción tomados por la dirección de las empresas. El turismo en este sentido se comporta de manera compleja dado que los datos analizados pueden contener información sin registro de ventas por tarifas hoteleras, tal es el caso del mes de abril del 2020 de la serie analizada.

Para poder realizar un modelo de previsión de datos en este trabajo se propuso la revisión de las tendencias de la serie para visualizar comportamientos afines a distintos valores de n , el uso de cinco métodos de análisis de persistencia a través del coeficiente H o de

Hurts para conocer la estacionalidad y el punto de cambio de sentido en las grafías en escalas logarítmicas. El resultado utilizado en el modelo aditivo de Holt-Winters parte del gráfico de longitud de rumorosidad L/R con una estacionalidad en el tiempo $n=8$,

este dato se usa como ciclo de longitud del modelo aditivo de Holt-Winters cuyos parámetros ($a=0.36$, $b=0.78$, $y=0.67$), minimizan el error presentado. Con este artículo se concluye que la combinación de herramientas matemáticas resulta pertinente para estimar el comportamiento de la serie de las tarifas hoteleras y describir la dinámica interna de los datos.

Estos resultados se pueden utilizar para mejorar la toma de decisiones en la gestión de servicios, a través de la ampliación del modelo para idéntica escenarios futuros del comportamiento de las tarifas que a su vez representen el comportamiento de la demanda para programar recursos físicos humanos y materiales en la gestión de servicios. Sin embargo este trabajo se limita solamente a encontrar la relación de parámetros que minimizan el error en la serie. Para futuras investigaciones se recomienda trabajar con los rendimientos logarítmicos de la serie y probar métodos que sean compatible con los resultados encontrados.

Referencias

- Al Shehhi, M., & Karathanasopoulos, A. (2020). Forecasting hotel room prices in selected GCC cities using deep learning. *Journal of Hospitality and Tourism Management*, 42, 40-50. <https://doi.org/10.1016/j.jhtm.2019.11.003>
- Blanco Rivero, J. J. (2019). The fractal geometry of Luhmann's sociological theory or debugging systems theory. *Technological Forecasting and Social Change*, 146, 31-40. <https://doi.org/10.1016/j.techfore.2019.05.020>
- Baggio, R. (2014). Complex tourism systems: a visibility graph approach. *Kybernetes*, 43(3-4), 445-461. <https://doi.org/10.1108/K-12-2013-0266>
- Baggio, R., & Sainaghi, R. (2011). Complex and chaotic tourism systems: towards a quantitative approach. *International Journal of Contemporary Hospitality Management*, 23(6), 840-861. <https://doi.org/10.1108/09596111111153501>

- Baggio, R., & Sainaghi, R. (2016). Mapping time series into networks as a tool to assess the complex dynamics of tourism systems. *Tourism Management*, 54, 23-33. <https://doi.org/10.1016/j.tourman.2015.10.008>
- Briones-Juarez, A., Padilla, R. T., Morales-Matamoros, O., Badillo-Pintilde, I. (2011). Fractal behavior of the sales restaurant in a great tourism hotel. *International Journal of Physical Sciences*, 6(10), 2422-2431. <https://doi.org/10.5897/IJPS11.193>
- Camila, A. (2022, enero 6). *Hoteles capitalinos no alcanzaron los niveles de ocupación de prepandemia en 2021*. El Economista. <https://www.eleconomista.com.mx/estados/Hoteles-capitalinos-no-alcanzaron-los-niveles-de-ocupacion-de-prepandemia-en-2021-20220106-0047.html>
- Castillo, O., & Melin, P. (2020). Forecasting of COVID-19 time series for countries in the world based on a hybrid approach combining the fractal dimension and fuzzy logic. *Chaos, Solitons, and Fractals*, 140, 110242. <https://doi.org/10.1016/j.chaos.2020.110242>
- Chang, Y.-M., Chen, C.-H., Lai, J.-P., Lin, Y.-L., & Pai, P.-F. (2021). Forecasting Hotel Room Occupancy Using Long Short-Term Memory Networks with Sentiment Analysis and Scores of Customer Online Reviews. *Applied Sciences*, 11(21), 10291. <https://doi.org/10.3390/app112110291>
- Chuwang, D. D., & Chen, W. (2022). Forecasting Daily and Weekly Passenger Demand for Urban Rail Transit Stations Based on a Time Series Model Approach. *Forecasting*, 4(4), 904-924. <https://doi.org/10.3390/forecast4040049>
- DATATUR. (2019). *Porcentaje de ocupación hotelera*. <https://datatur.sectur.gob.mx/SitePages/Inicio.aspx>
- Easwaramoorthy, D., Gowrisankar, A., Manimaran, A., Nandhini, S., Rondoni, L., & Banerjee, S. (2021). An exploration of fractal-based prognostic model and comparative analysis for second wave of COVID-19 diffusion. *Nonlinear Dynamics*, 106(2), 1375-1395. <https://doi.org/10.1007/s11071-021-06865-7>
- Godovych, M., Ridderstaat, J., Baker, C., & Fyall, A. (2021). COVID-19 and Tourism: Analyzing the Effects of COVID-19 Statistics and Media Coverage on Attitudes toward Tourism. *Forecasting*, 3(4), 870-883. <https://doi.org/10.3390/forecast3040053>
- Gunter, U. (2021). Improving Hotel Room Demand Forecasts for Vienna across Hotel Classes and Forecast Horizons: Single Models and Combination Techniques Based on Encompassing Tests. *Forecasting*, 3(4), 884-919. <https://doi.org/10.3390/forecast3040054>
- Hernández Rangel, M., Nieto-Malpica, J., & Quiñónez Guagua, O. (2021). Sistema de gestión de calidad y comunicación en el sector del turismo: Una aproximación al estado del conocimiento. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (E39), 368-378.
- Huang, L., & Zheng, W. (2021). Novel deep learning approach for forecasting daily hotel demand with agglomeration effect. *International Journal of Hospitality Management*, 98, 103038. <https://doi.org/10.1016/j.ijhm.2021.103038>

- Hou, K., Qian, H., Zhang, Y., Qu, W., Ren, W., & Wang, H. (2021). Relationship between fractal characteristics of grain-size and physical properties: Insights from a typical loess profile of the loess Plateau. *CATENA*, 207, 105653. <https://doi.org/10.1016/j.catena.2021.105653>
- Iranmanesh, M., Ghobakhloo, M., Nilashi, M., Tseng, M.-L., Yadegaridehkordi, E., & Leung, N. (2022). Applications of disruptive digital technologies in hotel industry: A systematic review. *International Journal of Hospitality Management*, 107, 103304. <https://doi.org/10.1016/j.ijhm.2022.103304>
- Kachhia, K. B. (2023). Chaos in fractional order financial model with fractal–fractional derivatives. *Partial Differential Equations in Applied Mathematics*, 7, 100502. <https://doi.org/10.1016/j.padiff.2023.100502>
- Kaya, K., Yılmaz, Y., Yaslan, Y., Ögüdücü, S. G., & Çıngı, F. (2022). Demand forecasting model using hotel clustering findings for hospitality industry. *Information Processing & Management*, 59(1), 102816. <https://doi.org/10.1016/j.ipm.2021.102816>
- Kim, W. G., McGinley, S., Choi, H.-M., Luberto, E., & Li, J. (2020). How does room rate and rate dispersion in U.S. hotels fluctuate? *Journal of Hospitality and Tourism Management*, 44, 227-237. <https://doi.org/10.1016/j.jhtm.2020.06.014>
- Lee, C. G. (2011). The determinants of hotel room rates: Another visit with Singapore's data. *International Journal of Hospitality Management*, 30(3), 756-758. <https://doi.org/10.1016/j.ijhm.2010.09.010>
- Lim, S., & Ok, C. (2022). A percentage-off discount versus free surcharge: The impact of promotion type on hotel consumers' responses. *Tourism Management*, 91, 104504. <https://doi.org/10.1016/j.tourman.2022.104504>
- Liu, P., Wu, L., & Li, X. (2022). What can hotels learn from the last recovery? Examining hotel occupancy rate and the guest experience. *International Journal of Hospitality Management*, 103, 103200. <https://doi.org/10.1016/j.ijhm.2022.103200>
- Liu, X., Peng, H., Bai, Y., Zhu, Y., & Liao, L. (2014). Tourism Flows Prediction based on an Improved Grey GM(1,1) Model. *Procedia - Social and Behavioral Sciences*, 138, 767-775. <https://doi.org/10.1016/j.sbspro.2014.07.256>
- Mendoza, Á. G. F., Arce, A. Z., & Holguín, J. S. V. (2021). Turismo En Tiempo De Pandemias. Covid-19 En Latinoamérica1. *Turismo y Sociedad*, 29, 129-155.
- Nguyen, L. Q., Fernandes, P. O., & Teixeira, J. P. (2022). Analyzing and Forecasting Tourism Demand in Vietnam with Artificial Neural Networks. *Forecasting*, 4(1), 36-50. <https://doi.org/10.3390/forecast4010003>
- Nilashi, M., Abumalloh, R. A., Almulihi, A., Alrizq, M., Alghamdi, A., Ismail, M. Y., Bashar, A., Zogaan, W. A., & Asadi, S. (2021). Big social data analysis for impact of food quality on travelers' satisfaction in eco-friendly hotels. *ICT Express*, 9(2), 182-188. <https://doi.org/10.1016/j.ict.2021.11.006>

- NWTO. (2021). *UNWTO World Tourism Barometer | Global Tourism Statistics*. <https://www.unwto.org/es/barometro-del-turismo-mundial-de-la-omt>
- Ouassou, E. H., & Taya, H. (2022). Forecasting Regional Tourism Demand in Morocco from Traditional and AI-Based Methods to Ensemble Modeling. *Forecasting*, 4(2), 420-437. <https://doi.org/10.3390/forecast4020024>
- Ozdemir, O., Dogru, T., Kizildag, M., Mody, M., & Suess, C. (2021). Quantifying the economic impact of COVID-19 on the U.S. hotel industry: Examination of hotel segments and operational structures. *Tourism Management Perspectives*, 39, 100864. <https://doi.org/10.1016/j.tmp.2021.100864>
- Pan, B., Chenguang Wu, D., & Song, H. (2012). Forecasting hotel room demand using search engine data. *Journal of Hospitality and Tourism Technology*, 3(3), 196-210. <https://doi.org/10.1108/17579881211264486>
- Pan, B., & Yang, Y. (2017). Forecasting Destination Weekly Hotel Occupancy with Big Data. *Journal of Travel Research*, 56(7), 957-970. <https://doi.org/10.1177/0047287516669050>
- Patiño, J., Mardones, R., Gonzalez, J., Romero-Riaño, E. , Palacios-Moya, L.(2020). Desarrollo de la investigación en sistemas de información para el turismo: una revisión centrada en visualización científica RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação, (E36), 387-398.
- Peng, D., & Li, X. (2023). Fractal contact analysis for transversely isotropic piezoelectric materials: Theoretical and numerical predictions. *Tribology International*, 181, 108323. <https://doi.org/10.1016/j.triboint.2023.108323>
- Pereira, L. N. (2016). An introduction to helpful forecasting methods for hotel revenue management. *International Journal of Hospitality Management*, 58, 13-23. <https://doi.org/10.1016/j.ijhm.2016.07.003>
- Pereira, L. N., & Cerqueira, V. (2022). Forecasting hotel demand for revenue management using machine learning regression methods. *Current Issues in Tourism*, 25(17), 2733-2750. <https://doi.org/10.1080/13683500.2021.1999397>
- Pimentel, V., Aizezikali, A., & Baker, T. (2019). Hotel revenue management: Benefits of simultaneous overbooking and allocation problem formulation in price optimization. *Computers & Industrial Engineering*, 137, 106073. <https://doi.org/10.1016/j.cie.2019.106073>
- Rashad, A. S. (2022). The Power of Travel Search Data in Forecasting the Tourism Demand in Dubai. *Forecasting*, 4(3), 674-684. <https://doi.org/10.3390/forecast4030036>
- Rocha, Á. (2012). Framework for a Global Quality Evaluation of a Website. *Online Information Review*, 36(3), 374-382. <https://www.abc.com>
- Sánchez-Medina, A. J., & C-Sánchez, E. (2020). Using machine learning and big data for efficient forecasting of hotel booking cancellations. *International Journal of Hospitality Management*, 89, 102546. <https://doi.org/10.1016/j.ijhm.2020.102546>

- Somphong, C., Udo, K., Ritphring, S., & Shirakawa, H. (2022). An estimate of the value of the beachfront with respect to the hotel room rates in Thailand. *Ocean & Coastal Management*, 226, 106272. <https://doi.org/10.1016/j.ocecoaman.2022.106272>
- Teixeira, J. P., & Gunter, U. (2023). Editorial for Special Issue: "Tourism Forecasting: Time-Series Analysis of World and Regional Data. *Forecasting*, 5(1), 210-212. <https://doi.org/10.3390/forecast5010011>

Divergencia en los niveles de madurez de e-gobierno en México

Oscar Yahev Carrera Mora¹

ocarrera@uv.mx

¹ Universidad Veracruzana, Facultad de Negocios y Tecnologías Campus Ixtaczoquitlán. 94720, Ixtaczoquitlán, Veracruz, México

DOI: 10.17013/risti.50.91-105

Resumen: La presente investigación busca de responder a dos ideas centrales, primero, demostrar que existen sitios web de e-gobierno en un mismo nivel de madurez con extremas divergencias; segundo, proponer las categorías que se desprenden de un mismo nivel de madurez de e-gobierno con la finalidad de conocer de una forma precisa el estatus de un sitio de e-gobierno a nivel municipal. El método utilizado fue el descriptivo, analizando los 2470 municipios que conforman a las 32 entidades de México. Los hallazgos demuestran una divergencia importante entre los sitios web ubicados en un mismo nivel de los modelos madurez, principalmente aquellos que se encuentran en el nivel interactivo, lo cual, sugiere que existen tipos de sitios web interactivos, unos con funciones asincrónicas y otros con sincrónicas. Concluyendo que existe un nivel de interacción básico y otro avanzado. Este análisis contribuye a la investigación ofreciendo avances teóricos relacionados con las etapas de e-gobierno.

Palabras clave: Administración pública; e-gobierno; Modelos de madurez; Sitios web

Divergence in e-government maturity levels in Mexico

Abstract: The present research seeks to respond to two central ideas: first, to demonstrate that there are e-government websites at the same level of maturity with extreme divergences; second, to propose the categories that arise from the same e-government level maturity to precisely know the status of an e-government site at the municipal level. The method used was descriptive, analyzing the 2,470 municipalities that make up the 32 Mexican states. The findings show a significant divergence between websites located at the same level of the maturity models, mainly those at the interactive level, which suggests that there are types of interactive websites, some with asynchronous functions and others with synchronous ones. Concluding that there is a basic and an advanced level of interaction. This analysis contributes to the research by offering theoretical advances related to the stages of e-government.

Keywords: Public administration; e-government; Maturity model; Website.

1. Introducción

Las estructuras presentadas en la literatura actual con respecto a los niveles de los modelos de madurez del e-gobierno presentan fases con una descripción que tipifica de forma general a cada sitio *web*. Esto permite ubicar el estado de desarrollo de una plataforma de e-gobierno, ayudando a clasificar estos sitios para poder entender su función y el cómo operan en cada nivel (Zahran, et al, 2015; Ore y Lozada 2017). Sin embargo, las investigaciones indican que la existencia de varios modelos de madurez de e-gobierno que han permitido evaluar sus avances, son criticables por su similitud, su excesiva simplicidad, su visión unilateral y las pocas características incorporadas en sus diferentes niveles (Patergiannaki y Pollalis, 2022). A partir de ese contexto, la evidencia científica ha demostrado que en los países sub-desarrollados la evolución del e-gobierno es limitada y en ocasiones tienen deficiencias porque solo buscan cumplir con los indicadores, haciendo poco útiles y eficientes estos sitios web (Carrera-Mora, et al, 2019). En ese sentido, se ha podido observar en la literatura que los niveles de los modelos de madurez que definidos, están acotados a niveles conceptualizados con características específicas, pero generales, sin dar sub-categorías. Por lo que se facilita con ello, la simulación del desarrollo real del e-gobierno en estas naciones.

Generalizar el desarrollo de estas plataformas de e-gobierno bajo el mismo nivel, limita el conocimiento del estado de estos mecanismos (Patergiannaki y Pollalis, 2022; Zahran, et al, 2015), ya que categorizar una plataforma de e-gobierno con funciones en extremo diferentes en un mismo nivel del modelo de madurez, no favorece al conocimiento preciso de su estatus. Y en países en desarrollo, es posible observar que la mayor parte de los sitios *web* de e-gobierno se encuentran en niveles primarios (presencia e interacción) de e-gobierno (UN, 2022). Por ello, es necesario contar con una descripción más precisa de la estructura de los sitios web de e-gobierno a través de la definición de más niveles y/o sub-niveles en los modelos de madurez que permitan el diseño de estrategias de e-gobierno adecuadas para cada necesidad.

Así mismo, en estos países, la polaridad entre un sitio web de e-gobierno y otro que están en el mismo nivel se acentúa principalmente en los gobiernos locales y/o municipales, no en aquellas plataformas que dan atención a nivel nacional, desde los gobiernos federales (Carrera, et al, 2018). Por otra parte, el nivel local y/o municipal tiene los servicios más cercanos y cotidianos para la ciudadanía (Vera-Revilla, et al, 2023), lo cual implica otorgar más y mejores servicios.

En consecuencia, el presente trabajo de investigación trata de responder a dos ideas centrales, primero, demostrar que existen sitios *web* de e-gobierno en un mismo nivel de los modelos de madurez con extremas divergencias; segundo, proponer las categorías que se desprenden de un mismo nivel de los modelos de madurez de e-gobierno con la finalidad de conocer de una forma más precisa el estatus de un sitio *web* de e-gobierno a nivel municipal.

2. Revisión de la literatura

Los modelos de madurez desempeñan un papel fundamental en el diseño de servicios sostenibles de e-gobierno, ya que evalúan el contexto de la implantación de proyectos

y proporcionan planes estratégicos adecuados para ejecutar los proyectos. Por ello, es fundamental graduar los alcances de los sitios *web* de e-gobierno a partir de lo que se conoce como modelos de madurez, porque permiten facilitar el conocimiento preciso de la estructura y funcionalidad de ellos. De hecho, Joshi y Shareeful (2018) explican que los proyectos de e-gobierno en los países en vías de desarrollo se enfrentan a muchos retos para ofrecer servicios sostenibles. Explicando que la mayoría de los estudios consideran que la falta de tecnología y las limitaciones en los presupuestos y los recursos humanos son los principales obstáculos para la implementación efectiva de los servicios de e-gobierno. Aunado a lo anterior, la falta de conocimiento preciso del estado de las plataformas genera mayores limitaciones en su desarrollo. En consecuencia, los modelos de madurez desempeñan un papel fundamental en el diseño de servicios sostenibles de e-gobierno, ya que evalúan el contexto de la implantación de esos proyectos que son específicos para cada caso en que se ejecutan.

En ese mismo sentido, diversos autores como Almuftah, *et al*, (2016), Almarabeh y Abuali (2010) y Chohan, *et al*, (2020) han explicado que los modelos de madurez de e-gobierno ayudan a diseñar, desarrollar e implementar servicios sostenibles y también pueden ofrecer la oportunidad de comprender los puntos fuertes y débiles de las aplicaciones de las Tecnologías de Información y Comunicación (TIC) gubernamentales en relación con los proyectos de e-gobierno que no tienen claridad de ello. Asimismo, el modelo de madurez de e-gobierno orienta a los tomadores de decisiones para controlar los procesos de desarrollo y mantenimiento de esos servicios, dando pautas de cómo hacer crecer una cultura de excelencia en la prestación y gestión de dichos servicios en un país. Por tal motivo, la literatura nos ofrece diversas alternativas para la medición del e-gobierno, a través de diferentes modelos que definen niveles o etapas de madurez de estos sitios *web*. De hecho, en estudios anteriores, varias investigaciones han presentado diversos modelos de madurez existentes para el desarrollo de su propio modelo según la necesidad específica de cada país. Por ejemplo, Layne y Lee (2001) describen cuatro etapas de desarrollo (Catálogo, transacción, integración vertical e integración horizontal) y Siau y Long (2005) amplían los modelos anteriores, describiendo cinco etapas del e-gobierno (Presencia *web*, interacción, transacción, transformación y democracia). Posteriormente Lee (2010), propuso el modelo sintetizado de desarrollo del e-gobierno siendo uno de los más citados por contener una meta síntesis de 12 modelos con sus etapas de desarrollo del e-gobierno, en donde, se presentan cinco etapas con dos sub-niveles: Presencia; Asimilación: Interacción e Integración; Reforma: Transacción y Racionalización; Morfología: Participación y Transformación; e-gobernanza: Involucramiento y Gestión de los procesos.

De este modo, es posible identificar de manera dominante en todos los modelos de madurez revisados, tres etapas y/o niveles base: Presencia, interacción y transacción (ver Tabla 1), que permiten simplificar y comprender la construcción de prácticamente todos los modelos de madurez de e-gobierno estudiados.

Esta forma de medir el avance del e-gobierno (por medio de un modelo de madurez), a pesar de ser un elemento básico e importante, comparado con los índices desarrollados por organismos internacionales como la UN (2020) a través de la UN e-government Survey que mide cuestiones como la infraestructura, el capital humano, la participación, los servicios en línea, etc., y el de la OCDE (2019) que también presenta el diseño de

un índice de gobierno digital que permite indagar sobre el avance en los países, no permiten visualizar claramente la situación de madurez del caso mexicano, porque estos índices consideran en su análisis, únicamente plataformas que cubren necesidades a nivel nacional, pero no a nivel municipal y/o local, las cuales requieren categorizarse en niveles de madurez para facilitar el conocimiento del avance de cada una de éstas, esto sin considerar el índice de usabilidad que se dará posteriormente con su implementación. Por lo cual, en un análisis primario de los portales *web* que ofrecen servicios en línea de la administración pública municipal, favorece en el conocimiento del estado en que se encuentra cada sitio web (al menos ofreciendo un panorama aproximado). Con ello, se puede establecer en un principio la utilidad que representarán estos mecanismos para la ciudadanía, pero de acuerdo con la visión de Lee (2010) existen tanto niveles altos como bajos en los niveles de los modelos de madurez del e-gobierno, a su vez, Müller y Skau (2015) explican que es necesario realizar la distinción entre factores de éxito de bajo y alto nivel para ayudar a los profesionales a modificar sus planes y tomar las decisiones correctas al digitalizar los servicios de la administración pública en los distintos modelos de madurez del e-gobierno.

Nivel	Descripción	Autores Base
Presencia	En este nivel, los sitios <i>web</i> de e-gobierno proporcionan únicamente información estática. Incluye publicaciones legislativas, avisos públicos, difusión de las actividades gubernamentales por Internet, información de contacto con las autoridades gubernamentales y portales electrónicos personalizados.	Almazan y Gil-Garcia (2008); Almuftah, et al, (2016); Group (2000); Long (2005); Moon (2002); Gartner IMCO (2014); Naser y Concha (2011); Lee (2010); Morales y Bayona (2020); Patergiannaki y Pollalis, (2022); Shareef, et al, (2010); Siau y Long (2005);
Interacción	Este nivel implica un proceso bidireccional de comunicación entre los ciudadanos y el gobierno a través del sitio <i>web</i> . Los usuarios pueden comunicarse con el gobierno de forma sincrónica a través de formularios, correo electrónico y/o asincrónica por chat u otros servicios de comunicación mensajería instantánea integrados o no en la plataforma <i>web</i> como WhatsApp, Messenger Facebook, entre otros.	
Transacción	En este nivel, los servicios y las transacciones financieras están disponibles en línea para el uso de los ciudadanos. El usuario puede realizar transacciones completas en línea, esto incluye pagos de servicios, de impuestos, etc., lo cual implica que el ciudadano no tenga que ir físicamente a una dependencia de gobierno a realizar el trámite.	

Fuente: Autor.

Tabla 1 – Niveles de madurez de e-gobierno

Los niveles bajos incluyen las etapas de presentación y asimilación y los niveles altos incluyen las etapas de reforma, transformación y gobernanza electrónica. Sin embargo, en muchos escenarios dentro del mismo nivel de asimilación, existen patrones que dan evidencia de un nivel alto y otro bajo dentro del mismo nivel, derivado de la aplicación de mecanismos que generan un diferencial de comunicación importante. Sin embargo, en el nivel de asimilación podemos encontrar dos grados: el de interacción y el de integración, cuyo objetivo es poder observar el grado tecnológico que posee una plataforma, pero no el grado de funcionalidad que estas tienen. De hecho, de los 12 modelos estudiados por

Lee (2010) y los verificados como los de Shareef, *et al*, (2010); Das, *et al*, (2016), etc., únicamente consideran una opción para el nivel interactivo, cuando la comunicación se puede dar de diversas formas (sincrónica o asincrónica) a través de los mecanismos tecnológicos actuales en función del procesamiento del servicio ofrecido (Suzuki y Romualdo-Suzuki, 2010).

En el modelo de comunicación síncrono, el servicio solicitado se procesa y concluye en la misma conexión (en tiempo real), el usuario recibe un mensaje con el resultado del procesamiento de la solicitud. En cambio, en el servicio de tipo asíncrono el procesamiento de la solicitud no se realiza en la misma conexión, pero el usuario debería recibir un acuse de recibo de la solicitud realizada. Esta diferencia no se presenta en los niveles de los modelos de madurez estudiados en la literatura actual (Almuftah, *et al*, 2016; Lee, 2010; Patergiannaki y Pollalis, 2022; Shareef, *et al*, 2010; Zahran, *et al*, 2015). Situación que lleva a plantear ¿Los niveles de madurez del e-gobierno actuales responden a las necesidades de diseño de los sitios web a nivel municipal y/o local en países emergentes?

En el caso mexicano los indicadores plantean que gobierno a nivel global en el año 2022 contaba con un índice de desarrollo de gobierno electrónico de 0.74 (ver Gráfico 1). Teniendo el índice más alto de todos los países de América Latina y el Caribe. A su vez, presenta un alto índice de desarrollo con un crecimiento anual sostenido de su desarrollo de e-gobierno. También ocupa el lugar 62 de 193 países con base en su índice. Asimismo, cuenta con un índice de servicios en línea (OSI) del 0.8245 (UN, 2022).

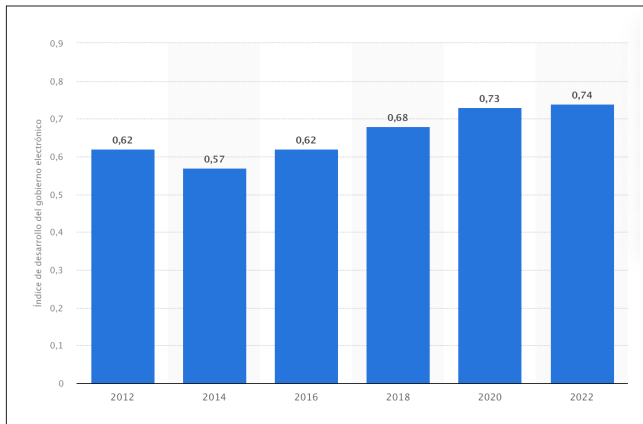


Gráfico 1 – Índice de desarrollo del e-gobierno en México de 2012 a 2022.
Fuente: Autor a partir de datos de la UN.

México en materia económica es uno de los países con un mayor PIB y equilibrio económico hasta 2021, según datos de Statista (2022). Sin embargo, las investigaciones indican lo contrario, presentando indicadores en un nivel alto, pero la realidad es otra, por ejemplo, un estudio realizado por Carrera, *et al*, (2018) de un estado en México con 212 municipios, demuestra que los indicadores dicen que el 100% de sus municipios cuenta con un sitio web de e-gobierno, pero la evidencia encontrada al realizar la revisión

demostró que de todos los portales declarados solo existía el 36% en la realidad. Por otra parte, se detectó esa polaridad entre sitios web de un mismo nivel de los modelos de madurez. Por lo cual, este caso presenta un escenario ideal de análisis que permite indagar sobre la realidad de estos servicios en línea y compararlo con el índice obtenido de la UN (2022) a través de la revisión de cada uno de los sitios web municipales existentes en México, ubicando con ello, el nivel correspondiente al modelo madurez que posee cada uno realmente.

Para lo anterior, se plantea la hipótesis siguiente: Existen sitios web de e-gobierno a nivel local en un mismo nivel de un modelo de madurez con extremas divergencias de los cuales se pueden desprender categorías o sub-niveles de un mismo nivel del modelo de madurez de e-gobierno, debido a, la funcionalidad dispar que poseen.

3. Metodología

3.1. Método

Se consideraron únicamente los servicios de e-gobierno a nivel municipal, debido a que estos son los más cercanos a la ciudadanía, así mismo, los sitios web de este nivel de gobierno representan un déficit mayor de desarrollo en países emergentes. Para su análisis se utilizó el método descriptivo a partir de una revisión a profundidad de los 2740 municipios que conforman las 32 entidades federativas de México. El proceso inicio buscando los URL de cada uno de los municipios para poder establecer, primero, si contaban o no con un sitio *web* de e-gobierno. Para decir que un municipio cuenta con un portal de e-gobierno, se tomaron los parámetros siguientes: primero, éste debía poseer un dominio.gob.mx, garantizando con ello, fuese un sitio oficial de gobierno. En caso de no contar con uno, el indicador colocado fue “No Tiene”. Cabe mencionar que, en aquellos casos que no se localizó un dominio del municipio en la búsqueda en la *web*, para reforzar la validez descriptiva de este dato, se realizó una segunda validación que fue vía telefónica. Corroborando su inexistencia con el personal de atención ciudadana de cada municipio revisado.

Segundo, para los casos que se determinó que si había un sitio *web* de e-gobierno en un municipio. Se realizó un concentrado, en donde, se ubicó a cada entidad federativa con sus respectivos municipios de acuerdo con los datos obtenidos a través de la plataforma digital del INEGI (2022), de donde, se tomaron nombres de entidades y sus municipios correspondientes. Se realizó la evaluación de las funciones que ofrecía cada sitio *web* utilizando la clasificación de la tabla 2, para determinar el estado de cada uno y poder determinar la categoría correspondiente de cada uno de ellos.

Estatus del sitio <i>web</i> de e-gobierno	Descripción
No tiene	No cuenta con una dirección de sitio web de e-gobierno (No se localizó el URL).
No Funciona	Cuenta con una dirección de sitio web de e-gobierno, pero no funciona al intentar acceder.
En mantenimiento	Cuenta con una dirección de sitio web de e-gobierno, pero indica que se encuentra en mantenimiento.

Estatus del sitio <i>web</i> de e-gobierno	Descripción
<i>Presencia</i>	Los sitios web de e-gobierno proporcionan únicamente información estática.
<i>Interacción</i>	El sitio web permite la comunicación entre el gobierno y el ciudadano de manera: Asincrónica, pudiendo contener aplicaciones como: correo electrónico, blogs, formularios de contacto. Sincrónica (en tiempo real): cuenta con aplicaciones de comunicación como: Chat, foros y servicios de mensajería no integrados (WhatsApp, Messenger Facebook, etc.).
<i>Transacción</i>	El usuario puede realizar transacciones completas en línea, esto incluye pagos de servicios, de impuestos, etc.

Fuente: Autor.

Tabla 2 – Categorías para determinar el estado de los sitios *web* de e-gobierno.

La revisión realizada a cada sitio *web* fue por medio de un estudio longitudinal, verificando el estado de los sitios *web* en diferentes momentos de tiempo. Tratando de garantizar con ello, la existencia o inexistencia de un portal, así como, los posibles cambios de estatus que pudiesen haber existido en el periodo de análisis que fue de febrero a octubre del año 2022. Para identificar los cambios se realizó una base de datos en la cual se anotaron observaciones con las características encontradas de cada sitio web en la primera revisión. Posteriormente se realizó una segunda verificación por otro colaborador para corroborar que el sitio web mantuviera las mismas características, de haber cambiado se actualizaba la base de datos. En caso de tener perspectivas diferentes, en cuanto a las características del sitio web, un tercer colaborador revisaba el portal para asegurar la veracidad de la información. Esta labor fue realizada por el investigador de este estudio, en colaboración con su equipo de trabajo (conformado por 3 integrantes con conocimientos en el área de sistemas de información), quienes, a partir de los parámetros presentados realizaron la revisión de cada sitio web. Los parámetros de la revisión consistieron en hallar a través de buscadores los sitios web conforme el procedimiento antes mencionado. Después, identificar el nivel correspondiente a la categoría correspondiente (ver Tabla2). Finalmente se procedía a examinar que la información del portal fuera actual, esto se determinó buscando que la información presentada fuera la correspondiente al último periodo gubernamental y/o las fechas de las publicaciones fuera actual. Por último, se anotaban las funciones y observaciones con respecto a la categoría que cada sitio web presentó.

4. Resultados

Se identificó que, solo 1494 de los 2470 municipios contaban con un URL que aloja oficialmente un sitio *web* de e-gobierno, lo cual, representa solo el 57.7% de ellos. Esto indica una disponibilidad limitada de e-gobierno en México a nivel municipal. Asimismo, de esos sitios hallados, se encontró sin funcionamiento y/o en mantenimiento a 68 casos. Por lo tanto, la base de análisis fueron únicamente 1426 sitios *web* de e-gobierno. Encontrando superioridad de los sitios en el nivel de madurez de presencia con 836, lo cual representa el 58.63% del total de la muestra de sitios existentes (ver Gráfico 2).

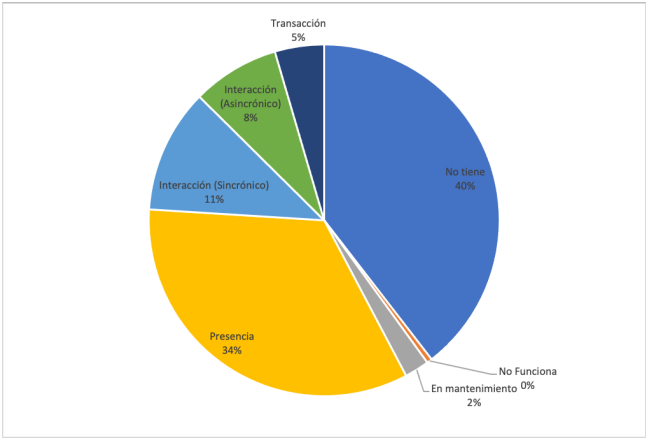


Gráfico 2 – Estatus de los sitios web de e-gobierno municipal. Fuente: Autor.

En ese sentido, podemos observar que más de la mitad de las plataformas se encuentran en el nivel más bajo de madurez. Situación que representa insuficiencia de los servicios de e-gobierno en México. Seguido de ello, hallamos que el nivel interactivo está presente en 481 municipios con un sitio web. Por último, el nivel transaccional está presente en 112 casos (ver Tabla 3).

Id	Entidad	Total Municipios	Estatus del Sitio web de e-gobierno						
			NT	NF	EM	P	I-S	I-A	T
1	Aguascalientes	11	1	0	0	4	2	2	2
2	Baja California	6	1	0	0	1	0	0	4
3	Baja California Sur	5	0	0	0	3	0	0	2
4	Campeche	12	2	0	0	5	0	4	1
5	Chiapas	125	52	6	0	47	18	0	2
6	Chihuahua	67	52	0	1	6	2	5	1
7	Coahuila	38	22	0	0	5	1	8	2
8	Colima	10	1	0	0	2	0	3	4
9	CDMX	16	0	0	0	10	0	6	0
10	Durango	39	11	1	0	15	1	10	1
11	Guanajuato	46	5	0	3	16	2	13	7
12	Guerrero	81	23	0	0	34	2	20	2
13	Hidalgo	84	4	0	4	43	29	2	2
14	Jalisco	125	4	0	7	68	38	2	6

Id	Entidad	Total Municipios	Estatus del Sitio web de e-gobierno						
			NT	NF	EM	P	I-S	I-A	T
15	Estado de México	125	6	3	1	65	28	2	20
16	Michoacán	113	8	0	4	60	38	2	1
17	Morelos	36	3	0	4	17	1	6	5
18	Nayarit	20	2	1	1	9	0	6	1
19	Nuevo León	51	21	0	0	9	1	12	8
20	Oaxaca	570	538	1	3	10	0	16	2
21	Puebla	217	35	0	3	132	44	2	1
22	Querétaro	18	0	0	1	5	1	5	6
23	Quintana Roo	11	1	0	0	5	0	2	3
24	San Luis Potosí	58	9	0	7	34	1	7	0
25	Sinaloa	18	0	0	0	3	0	8	7
26	Sonora	72	29	0	0	2	35	2	4
27	Tabasco	17	1	0	0	9	1	5	1
28	Tamaulipas	43	5	0	4	8	23	2	1
29	Tlaxcala	60	4	0	0	48	4	3	1
30	Veracruz	212	76	0	0	115	8	2	11
31	Yucatán	106	40	0	10	22	1	32	1
32	Zacatecas	58	20	0	3	21	1	10	3
Total		2470	976	12	56	833	282	199	112

NT= No tiene; NF= No funciona; EM= En mantenimiento; P= Presencia; I-S= Interacción Sincrónica; I-A= Interacción Asincrónica; T= Transaccional.

Tabla 3 – Clasificación de los sitios web de e-gobierno municipal según su estatus.

El nivel interactivo, es el que presenta mayores matices en este análisis, además de ser el nivel dominante en este caso de estudio. Por otra parte, como se identificó en la literatura, para este nivel hay diferentes tipos de comunicación (sincrónica y asincrónica). Motivo por el cual, el análisis de los resultados se centra en el nivel interactivo, ya que, además, representa el 33.7% del total de los municipios con un sitio *web*. Situación que indica que una tercera parte de todos los sitios *web* a nivel municipal existentes en todo el territorio mexicano se encuentran en el nivel de interacción. Entonces 481 sitios *web* de e-gobierno municipal, se encuentran en el nivel de madurez de interacción. Asimismo, según datos del índice de e-gobierno (UN, 2022), México cuenta con un alto nivel de interacción. De tal forma que, la revisión de cada uno de los sitios web interactivos permitió identificar lo siguiente:

En primer lugar, se observa en 199 de los sitios web identificados como interactivos que a través de ellos se realiza únicamente una comunicación a partir de formularios de contacto y/o correo electrónico, lo cual, de acuerdo con lo expuesto en la teoría es una comunicación asincrónica. Esto indica que, en estos municipios, los sitios *web* de e-gobierno no tienen la posibilidad de tener una comunicación en tiempo real entre el ciudadano y el servidor público, a pesar de ser definidos como interactivos. En segundo lugar, se observó en el resto de los sitios *web* interactivos, los cuales son 282, que adicionalmente a los elementos de comunicación asincrónica, tienen funciones de comunicación que se encuentran integrados en la plataforma, tales como chat y foros para una comunicación en tiempo real, así como, elementos de comunicación no integrados a ellas, como redes sociales, entre las que dominan Messenger Facebook y WhatsApp, mismos que, permiten una comunicación sincrónica y/o en tiempo real, el resto de casos cuentan con Twitter e Instagram, entre otros medios de comunicación masivos.

En los sitios web con una comunicación asincrónica, predominan los formularios de contacto que vienen predeterminados en los creadores de sitios *web* y en la mayoría de los casos son únicamente una función decorativa del sitio, porque al enviar una solicitud por ese medio, no hay ninguna respuesta por parte de las autoridades municipales, al menos no la hubo durante el tiempo de realización del estudio.

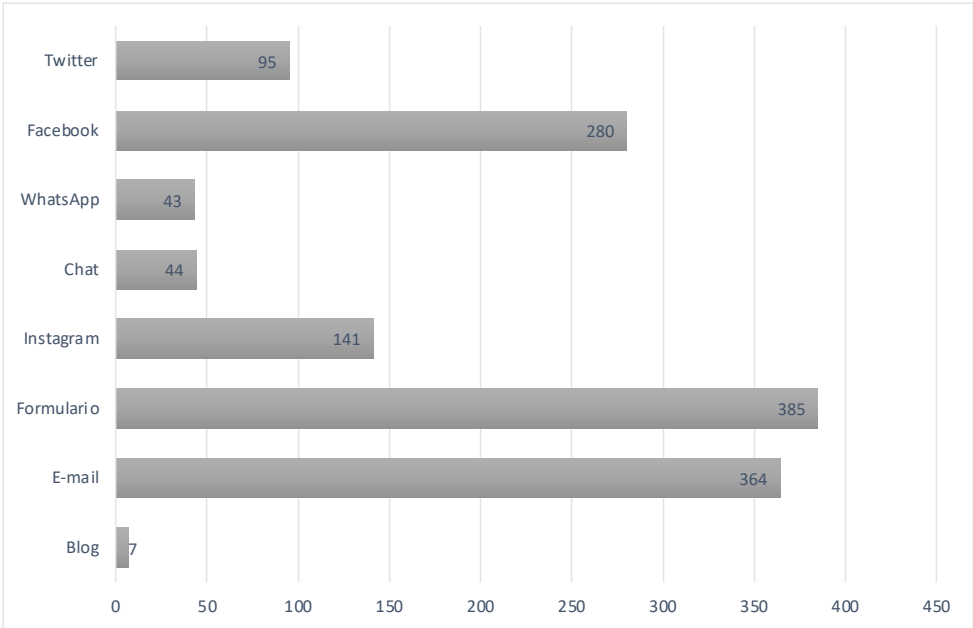


Gráfico 3 – Elementos de interacción en los sitios web de e-gobierno. Fuente: Autor.

En los sitios web con una comunicación sincrónica, se observó que, hay un predominio de sitios que cuentan con una comunicación no integrada en la plataforma a través

de redes sociales, a pesar de ser sitios caracterizados por encontrarse en un nivel interactivo del modelo de madurez. Destacando Facebook en el 99% de los sitios web con una comunicación sincrónica (ver Gráfico 3). Asimismo, en aquellos sitios web con una comunicación sincrónica se observaron mejores características visuales de la plataforma, comparadas con las que no tienen estas opciones. Además, en aquellas que cuentan con funciones de comunicación integradas en la misma plataforma como lo es: un chat, representan solo el 9% de todos los sitios web en el nivel interactivo, pero también representan un aspecto visual más evolucionado con respecto a aquellas que únicamente tienen una comunicación no integrada, como las que poseen solo redes sociales. Dentro de las opciones que destacan encontramos banners informativos, correcta distribución de la información, botones de acceso directo a redes sociales y fácil acceso a las opciones; a diferencia de los sitios web con una comunicación asincrónica y/o comunicación sincrónica no integrada en la plataforma.

Finalmente, podemos observar que, del total de los sitios *web* analizados, existe una clara heterogeneidad en el nivel de interacción del modelo de madurez, en donde, el 58.63% de ellos, tienen una interacción a través de herramientas de comunicación sincrónica. Asimismo, en el 41.63% su forma de interacción es asincrónica, pero también ha sido posible identificar diferencias estructurales en las plataformas, en cuanto a su apariencia y funcionamiento. Resultado que pone de manifiesto la divergencia que existe entre las funciones de un sitio *web* en un mismo nivel del modelo de madurez como lo es el interactivo.

5. Conclusiones

A partir de la revisión de todas las plataformas de e-gobierno a nivel municipal de este caso de estudio particular que fue el de México, se pudo comprobar la hipótesis con los hallazgos, mismos que demuestran una divergencia a partir de lo siguiente: los sitios web ubicados en el nivel transaccional no difieren demasiado sus funcionalidades entre una plataforma y otra de diversos municipios. En estos sitios web, el único servicio similar en todos los ubicados en este nivel transaccional son el pago de impuestos principalmente el denominado predial¹, por lo cual, no hay mucha variación en las funciones, ya que, prácticamente todos presentan las mismas funciones y/o carencias (mismas que serán motivo de otros estudios). En contra parte, en el nivel inicial que es el de presencia, los hallazgos son similares a los sitios web del nivel transaccional, solo que en este caso los matices se dan principalmente en la profundidad del diseño, como lo son colores, imágenes, banners, entre otros elementos que en definitivo no modifican la funcionalidad. Sin embargo, en el nivel interactivo es, en donde, se encontraron diferencias importantes de resaltar y describir con detenimiento.

En estos sitios web ubicados en el nivel interactivo se encontró, por un lado, plataformas interactivas con funciones básicas de comunicación de tipo asincrónica como lo son: el correo electrónico, formularios y/o blog, pero en contra parte, hay otros sitios web que a pesar de contar con estas mismas funciones de comunicación, adicionalmente también

¹ El predial es un impuesto que debe pagarse al municipio todos los años. Este pago debe ser cubierto por todos los propietarios de un inmueble, ya sea una casa, departamento, edificio o terreno.

cuentan con otras alternativas de comunicación de tipo sincrónica como: chat integrado en la plataforma, redes sociales como: Facebook y WhatsApp, entre otros elementos. Por lo tanto, al tipificarlas con base en los modelos de niveles madurez de e-gobierno expuestos en la literatura actual y que presentan entre 3 y 5 etapas (Almuftah, *et al*, 2016; Lee, 2010; Zahran, *et al*, 2015), es posible evidenciar una importante divergencia que pone de manifiesto lo expuesto por Patergiannaki y Pollalis (2022) quienes identificaron la necesidad de ampliar estos modelos de madurez de e-gobierno, los cuales explican son deficientes porque no profundizan con sub-niveles o sub-categorías que permitan definir con más detalle las funcionalidades de estas plataformas y ampliar con ello la visión del estado que tienen.

Por lo tanto, este nivel interactivo es el que pone de manifiesto la divergencia en un mismo nivel de madurez a partir de la revisión de los sitios *web* de e-gobierno a nivel municipal en México. Pudiéndose identificar tres sub-niveles para el nivel interactivo, de los cuales se puede observar, uno básico en el cual la comunicación es de tipo asincrónica; uno intermedio, en el cual la comunicación tiene elementos de tipo asincrónicos, pero también de tipo sincrónicos que no están integrados en la plataforma como lo son: las redes sociales. Por último, se observó un nivel interactivo avanzado con funciones de comunicación sincrónica e integradas en la plataforma, como se explica con detalle en la tabla 6.

Nivel básico de interacción	Sitios <i>web</i> de e-gobierno que permiten la interacción entre el gobierno y el ciudadano de manera asincrónica, pudiendo contener aplicaciones de comunicación como: correo electrónico, blogs, formularios de contacto, entre otros.
Nivel intermedio de interacción	Sitios <i>web</i> de e-gobierno que además de las funciones asincrónicas, permiten la interacción entre el gobierno y el ciudadano de manera sincrónica (en tiempo real), a través de funciones de comunicación no integradas en la plataforma, por medio de redes sociales como: WhatsApp, Messenger Facebook, etc.
Nivel avanzado de interacción	Sitios <i>web</i> de e-gobierno que permiten una interacción sincrónica (en tiempo real) entre el gobierno y el ciudadano con funciones de comunicación integradas en la plataforma como: Chat y/o servicios de mensajería instantánea. Además de poder contar con funciones asincrónicas y sincrónicas no integradas en la plataforma.

Fuente: Autor.

Tabla 6 – Niveles de madurez interactivos hallados.

Evidenciándose con lo anterior, la necesidad de profundizar en esas categorías definidas tradicionalmente en otros estudios (Almuftah, *et al*, 2016; Lee, 2010; Patergiannaki y Pollalis, 2022; Shareef, *et al*, 2010; Zahran, *et al*, 2015), porque ello permitirá a los tomadores de decisiones del gobierno tener una perspectiva más clara del estado de sus plataformas en la administración pública. Pudiendo con ello, incentivar el uso de ellas con la implementación de más funciones a partir del diseño de políticas públicas enfocadas a cada caso en específico como lo explican Almuftah, *et al*, (2016) y Patergiannaki y Pollalis (2022). En el caso particular de México, es importante precisar que el 41.3% de todos los sitios *web* de e-gobierno a nivel municipal, oficialmente se encuentran en el nivel interactivo, lo cual, bajo el esquema tradicional podría generar un espejismo de la eficiencia y estructura del e-gobierno que un país posee. Esto porque solo la mitad de

esos sitios web favorecen la comunicación en tiempo real entre el gobierno y el ciudadano y, de esa mitad solo un 8% posee funciones integradas en la plataforma. Basándose en la visión de Lee (2010) quién explica que existen tanto niveles altos como bajos en los niveles de madurez del e-gobierno, este enfoque presentado es factible.

Asimismo, es interesante el hallazgo porque muestra que los sitios *web* con un nivel básico de interacción, en su mayoría no cuentan con características visuales adecuadas. Además, tienen elementos completamente básicos, incluso en la mayoría de los casos poseen funciones de correo electrónico, blog y/o formularios de contacto como elementos interactivos, pero estos mecanismos son decorativos porque si un ciudadano realiza una consulta a través de estos medios, la administración pública no responde a las solicitudes por estos medios de comunicación. Por otra parte, este hallazgo es importante porque al estructurar estrategias para el desarrollo de e-gobierno, un sitio *web* en un nivel asincrónico, se encuentra en condiciones tecnológicas y de madurez totalmente diferentes a uno con funciones sincrónicas que además tiene elementos de comunicación integrados en el mismo sitio web. Desfavoreciendo la toma de decisiones de los responsables de la administración pública y limitando la eficiencia de los sitios *web* para la ciudadanía, lo cual limita la usabilidad de éstos. Si a simple vista observamos los datos, podríamos decir que es un buen panorama evolutivo contar con 481 sitios *web* en el nivel de interacción, pero la realidad es que solo la mitad de esos portales tienen la capacidad de dar una respuesta inmediata a la ciudadanía. En ese sentido, Müller y Skau (2015) explican que es necesario realizar la distinción entre factores de éxito de bajo y alto nivel para ayudar a los profesionales a modificar sus planes y tomar las decisiones correctas al digitalizar los servicios de la administración pública en los distintos niveles de madurez del e-gobierno.

Por lo tanto, con estos resultados se puede asumir que la clasificación utilizada tradicionalmente, abre la puerta para simular indicadores de desarrollo de e-gobierno. Pudiéndose decir que una gran parte de las plataformas se encuentra en el nivel de interacción (aunque en realidad se encuentren en situaciones divergentes, a pesar de realmente estar en el mismo nivel), porque al colocar en los sitios web funciones de correo electrónico y/o un formulario, ello ya los define como una plataforma de e-gobierno en el nivel interactivo, pero no necesariamente todas los sitios web ubicadas en este nivel (bajo los esquemas tradicionales de los niveles de madurez de e-gobierno) poseerán las mismas funciones o tendrán el mismo alcance. También con lo observado, se puede deducir que no hay una figura en la administración pública designada para atender esos medios, lo cual podría poner en duda el objetivo real del desarrollo de estas plataformas. Poniendo de manifiesto la necesidad de investigar si esta suposición es real, así como, la importancia de indagar sobre los recursos humanos asignados para la operación de estas plataformas de e-gobierno, las reglas de operación definidas para estas plataformas, etc.

Referencias

- Almuftah, H., Weerakkody, V., & Sivarajah, U. (2016). Comparing and contrasting e-government maturity models: a qualitative-meta synthesis. In: Brunel Business School Research Papers. IOS PRESS. <https://doi.org/10.3233/978-1-61499-670-5-69>.

- Almarabeh, T., & Abuali, A. (2010). A general framework for e-government: definition maturity challenges, opportunities, and success, *European Journal of Scientific Research*, (39), 1216-1450. Disponible en: https://elearn.uni-sofia.bg/pluginfile.php/24221/mod_resource/content/0/E-Government-DefinitionMaturitySuccess.pdf
- Carrera-Mora, O., Villafuerte, L., & Rodríguez, A. (2018). Funcionalidad y eficiencia del gobierno electrónico local en el Estado de Veracruz, México: ¿realidad o simulación? En Lagunes, A., Jiménez, Y., & Mancilla, M. (Coord). Estrategias organizacionales e innovación tecnológica. Ed. Centro de estudios e investigaciones para el desarrollo docente (pp. 52-67). México.
- Carrera-Mora, O., Ovando, M., Villafuerte, L., & Rodríguez, A. (2019). La relación de la perspectiva de eficiencia del ciudadano con su comportamiento de uso de los servicios de e-gobierno municipal. *Innovar*, 29(74), 133-146. <https://doi.org/10.15446/innovar.v29n74.82096>.
- Chohan, S. R., Hu, G., Si, W., & Pasha, A. T. (2020). Synthesizing e-government maturity model: a public value paradigm towards digital Pakistan. *Transforming Government: People, Process and Policy*, 14(3), 495–522. Disponible en: <https://doi.org/10.1108/tg-11-2019-0110>.
- Dener, C., Nii-Aponsah, H., Ghunney, L. E., Johns, K. D. (2021). Índice de Madurez del GovTech: El estado de la transformación digital del sector público. *International Development in Focus*; Washington, DC: World Bank. <https://openknowledge.worldbank.org/handle/10986/36233>.
- Gartner Group. (2017). 5 Levels of Digital Government Maturity. Disponible en: <https://www.gartner.com/smarterwithgartner/5-levels-of-digital-government-maturity>
- Joshi, P. R., & Shareeful, I. (2018). E-Government Maturity Model for Sustainable E-Government Services from the Perspective of Developing Countries. *Sustainability* 10 (6), 1882. <https://doi.org/10.3390/su10061882>.
- Lee, J. (2010). 10-year retrospect on stage models of e-Government: a qualitative meta-synthesis. *Government Information Quarterly*, 27 (3), 220–230. <https://doi.org/10.1016/j.giq.2009.12.009>.
- Morales, V., & Bayona, S. (2020). Modelo conceptual de factores y su influencia en los niveles de desarrollo del e-gobierno municipal. *RISTI - Revista Ibérica de Sistemas e Tecnologías de Informação*, (28), 28-41. Disponible en: <https://www.risti.xyz/issues/ristie28.pdf>.
- Müller, S. D., & Skau, S. A. (2015). Success factors influencing implementation of e-government at different stages of maturity: a literature review. *International Journal of Electronic Governance*, 7(2), 136. <https://doi.org/10.1504/ijeg.2015.069495>.
- Naser, A., & Concha, G. (2011). El gobierno electrónico y la gestión pública. CEPAL. Serie gestión pública, 73. Disponible en: https://www.cepal.org/sites/default/files/publication/files/7330/S1100145_es.pdf.

- OCDE. (2019). Índice de gobierno digital. Resultados y mensaje clave. Disponible en: <https://www.oecd.org/gov/digital-government/digital-government-index-2019-highlights-es.pdf>.
- Ore, S. B., & Lozada, V. M. (2017). Maturity Model for Local E-Government: A Case Study. In ICCMS' 17, 8th International Conference on Computer Modeling and Simulation (pp. 78-83).. <https://doi.org/10.1145/3036331.3050419>.
- Patergiannaki, Z., & Pollalis, Y. (2022a). E-Government maturity assessment: Evidence from Greek municipalities. *Policy & Internet*, 15(1), 6-35. <https://doi.org/10.1002/poi3.317>.
- Patergiannaki, Z., & Pollalis, Y. (2022b). A Review of the e-Government Maturity Models: Are They Still Relevant? *Journal of Politics and Ethics in New Technologies and AI*, 1(1), 1-16. <https://doi.org/10.12681/jpentai.30872>.
- Suzuki, T., & Romualdo-Suzuki, L. (2010). On the benefit of 3-tier SOA architecture promoting information sharing among TMS systems and Brazilian e- Government Web Services: A CT-e case study. Extended Paper of the Proceedings of the 1st Conference on Computational Interdisciplinary Sciences (CCIS). Disponible en: <https://arxiv.org/pdf/2005.13047.pdf>.
- United Nations. (2020). Gobierno digital en la década de acción para el desarrollo sostenible. Encuesta sobre E-Gobierno, 2020. Disponible en: [https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2020-Survey/2020%20UN%20E-Government%20Survey%20\(Spanish%20Edition\).pdf](https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2020-Survey/2020%20UN%20E-Government%20Survey%20(Spanish%20Edition).pdf).
- United Nations. (2022). UN E-Government Survey 2022. Disponible en: <https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2022>.
- Shareef, M., Kumar, V., Kumar, U., Dwivedi, Y. (2010). e-Government Adoption Model (gam): Differing Service Maturity Levels. *Government Information Quarterly*, 28(1), 17-35. <https://doi.org/10.1016/j.giq.2010.05.006>.
- Statista. (2022). Producto interno bruto en América Latina y el Caribe en 2021, por país. Disponible en: <https://es.statista.com/estadisticas/1065726/pib-por-paises-america-latina-y-caribe/>
- Vera-Revilla, C., Pazo-Romero, A., Torreblanca-Gómez, G., Lazo, L., & Rodríguez-Cáceres, P. C. (2023). Concepción y discurso del desarrollo desde los gobiernos locales: un estudio de caso en un municipio del sur del Perú. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*. (E58), 189-206. Disponible en: <https://www.risti.xyz/issues/ristie58.pdf>.
- Zahrán, D. I., Al-Nuaim, H. A., Rutter, M. J., & Benyon, D. (2015). A Critical Analysis of E-Government Evaluation Models at National and Local Municipal Levels. *Electronic Journal of e-Government*, 13(1), 28- 42. Disponible en: <https://academic-publishing.org/index.php/ejeg/article/view/612>.

El efecto de un programa de psicoterapia de telesalud en la salud mental de jóvenes en alto riesgo psicosocial

Manchego Carnero Brígida-Aurora¹, Sivincha Huamani Karen²,
Manchego Carnero Rocío-Edith³

bmanchego@unsa.edu.pe; ksivincha@unsa.edu.pe; rmanchegoc@unsa.edu.pe

¹ Universidad Nacional de San Agustín de Arequipa, 04001, Arequipa, Perú.

² Hospital III Regional Honorio Delgado, 04001, Arequipa, Perú.

³ Universidad Nacional de San Agustín de Arequipa, 04001, Arequipa, Perú.

DOI: 10.17013/risti.50.106–117

Resumen: Este estudio experimental tuvo como objetivo determinar el efecto de la psicoterapia de telesalud en la salud mental de jóvenes en riesgo. La salud mental se evaluó utilizando el Cuestionario de Salud General de Goldberg. La intervención consistió en talleres sincrónicos realizados por personal experto en enfermería, mientras que el grupo de control no recibió intervención alguna. Se utilizó la prueba de McNemar para evaluar las diferencias entre las pruebas previas y posteriores en cada grupo. La muestra comprendió a 60 individuos jóvenes. Antes de la intervención, se encontraron síntomas psicopatológicos en el 67% y 93% de los grupos de control y experimental, respectivamente. Después de la intervención, estas frecuencias disminuyeron al 73% y 27%, respectivamente. La diferencia no fue significativa en el grupo de control ($p = 0.316$), pero sí fue significativa en el grupo experimental ($p = 0.000025$). En conclusión, la psicoterapia de telesalud dirigida por profesionales de enfermería mejora significativamente la salud mental de jóvenes en riesgo.

Palabras-clave: Telesalud; psicoterapia; salud mental; autoestima; jóvenes.

The effect of a telehealth psychotherapy program on the mental health of youth at high psychosocial risk

Abstract: This experimental study aimed to determine the effect of telehealth psychotherapy on the mental health of at-risk youth. Mental health was assessed using the Goldberg General Health Questionnaire. The intervention consisted of synchronous workshops conducted by expert nursing staff, while the control group received no intervention. The McNemar test was used to evaluate the differences between pre- and post-tests within each group. The sample comprised 60 young individuals. Prior to the intervention, psychopathological symptoms were found in 67% and 93% of the control and experimental groups, respectively. After the intervention, these frequencies decreased to 73% and 27%, respectively.

The difference was not significant in the control group ($p = 0.316$), but it was significant in the experimental group ($p = 0.000025$). In conclusion, telehealth psychotherapy led by nursing professionals significantly improves the mental health of at-risk youth.

Keywords: Telehealth; psychotherapy; mental health; self-esteem; youth.

1. Introducción

La salud mental es parte importante de la salud integral, incluso llegando a ser una parte indispensable de ésta, según la Organización Mundial de la Salud (OMS) (OMS, 2018a). De acuerdo a las estadísticas proporcionadas por la OMS, cada año personas de 15 a 29 años se suicidan cada 40 segundos, siendo así la segunda causa de muerte. Por otra parte, se resalta incluso que las personas que han sido diagnosticadas con por lo menos una enfermedad mental grave llegan a una edad máxima entre 10 y 20 años menos que las personas mentalmente sanas. Y finalmente, adicionando más a esta problemática, el número de profesionales con alguna especialidad en salud mental en países del tercer mundo son 2 para 100000 personas (OMS, 2018b; OPS/OMS, 2019).

Ahora bien, se define a la salud mental como un estado de plenitud en el cual el individuo siente una realización personal y desarrollo de sus capacidades, lo cual le permite enfrentar a la vida, sus problemas y finalmente ser productivo en su vida y en la sociedad. Por todo esto se puede afirmar que la salud mental es la base de la plenitud y bienestar del individuo, así como de su integración eficaz en la comunidad (MINSA, 2016, 2017; OMS, 2018a).

La Telesalud se ha visto como una alternativa para hacer llegar los servicios de salud a las personas de riesgo que residen en zonas alejadas a los centros de salud, facilitando así la descentralización de la atención médica, psicológica, y demás servicios benefactores de la salud. (Congreso de la República del Perú, 2019; Sánchez-Zapata & Zapata, 2019). Asimismo, la telesalud puede impulsar la involucración de las personas en los temas relacionados a la salud mental, especialmente en los tiempos de pandemia que vive el mundo, dando la oportunidad a la población de participar en los aportes y nuevas investigaciones de salud que surgen día a día en la comunidad científica, incrementando así los niveles de educación en el Perú. La telesalud, gracias a la tecnología de comunicación existente, se ha convertido en parte del servicio de salud, la cual es aplicada y dirigida por un personal de salud capacitado con el principal objetivo de que este servicio pueda ser más oportuno y accesible a la población en general.

Sin embargo, la aplicación de estas tecnologías en la salud mental también ha generado cuestionamientos sobre su efectividad, marco ético-legal, prejuicios, entre otros (Barnett, 2019; Hilty et al., 2018; Stoll et al., 2019). Es así que se plantea la pregunta, ¿La telesalud mejora los niveles de salud mental en jóvenes de una iglesia cristiana de un pueblo joven de Miraflores? De esta foma, el presente estudio pretende contribuir con evidencia sobre la existencia de beneficios de una intervención por Telesalud en la salud mental en jóvenes que residen en una zona geográfica considerada como de alto riesgo psicosocial, así como permitirá poner en práctica la intervención de Enfermería a distancia.

El resto del artículo está organizado de la siguiente manera, en la sección 2 se presenta el método empleado en este estudio cuasi experimental, donde se describe el diseño del estudio, la muestra utilizada y las variables de interés, así como las mediciones aplicadas para evaluar la salud mental de los participantes. En la sección 3, se exponen los resultados obtenidos a través del análisis estadístico de los datos recopilados, revelando las impactantes diferencias entre el grupo de control y el grupo experimental. Estos resultados se discuten y contextualizan en la sección 4, donde se exploran las implicaciones y conclusiones extraídas de este estudio, subrayando la relevancia de la intervención de Telesalud en la mejora de la salud mental de los jóvenes en situaciones de vulnerabilidad.

2. Método

2.1. Diseño de estudio

La investigación es de tipo cuasi experimental, con un diseño pre-test y post-test con dos grupos, un grupo de control y un grupo experimental.

2.2. Descripción de la muestra

Se realizó un estudio experimental, aleatorizado, sin cegamiento, aplicando un pre y post test a cada grupo. La población de estudio estuvo conformada por 60 jóvenes de una iglesia cristiana en un pueblo joven de Miraflores, Arequipa, Perú, que fue dividida de forma randomizada la cual fue repartida de manera aleatoria donde los números pares son el grupo de control y los impares el grupo experimental, la población joven de estudio radica en un asentamiento humano en el distrito de Miraflores, se encuentra dentro de un grupo vulnerable debido a la carencia de algunos recursos, tanto económicos (bajos ingresos) como también sociales. Además de los índices de alcoholismo, violencia familiar elevados y bajos niveles de educación que predominan en dicha zona.

Esta zona geográfica es considerada como de alto riesgo según el “Plan Nacional del Fortalecimiento de Servicios de Salud Mental Comunitaria” (MINSA, 2018) de Perú, debido a la carencia de algunos recursos, tanto económicos (bajos ingresos) como sociales, así como por tener los índices de alcoholismo y violencia familiar más elevados, y los niveles de educación más bajos.

Las características sociodemográficas estudiadas fueron sexo (femenino y masculino), edad (18 a 20, 21 a 23, 24 a 26, 27 a 29 años), grado de instrucción (primaria, secundaria, superior técnico y superior universitario) y estado civil (soltero, casado, conviviente y divorciado).

Cabe precisar que durante el tiempo del COVID-19 las universidades estatales peruanas implementaron un programa llamado “Beneficio de conectividad” que consistía en la entrega de equipos móviles con internet, donde los beneficiarios eran estudiantes con matrícula vigente y en condición de pobreza o pobreza extrema para asegurar su acceso a las clases virtuales.

Además, uno de los criterios de exclusión en la investigación era el acceso a internet, por lo tanto, por dicho criterio una gran parte de la población que participó en la investigación

era población universitaria, la cual tenía acceso a internet por parte del Estado Peruano a través de las universidades públicas.

Por otra parte, para la población no universitaria, la iglesia que participó en la investigación, siendo una institución de ayuda social y al tener trabajando más de 10 años en la zona, antes del COVID-19, tenían un programa de informática para adultos llamado Programa de Informática al Adulto Mayor (PROINAM), implementado con un área de informática en las instalaciones de la iglesia, lo cual ayudó a brindar a la población no universitaria las facilidades de conexión con el investigador.

2.3. Variables

Salud Mental: Es la capacidad de la persona de afrontar el estrés y problemas normales de la vida, y ejercer plenamente todas sus capacidades. Además, el individuo puede trabajar productivamente y contribuir al desarrollo pleno de su comunidad. Por lo tanto, la salud mental es el ejercicio pleno individual y de comunidad del bienestar.

Telesalud: Es el uso de los avances tecnológicos de información y comunicación para la aplicación del servicio de salud a pacientes cuyas residencias están alejadas geográficamente de los establecimientos de salud. Este servicio incluye el uso de diversos recursos tecnológicos informáticos como el internet, transmisiones de información, archivos, etcétera. Además de equipos de transmisión como teléfonos inteligentes, tabletas y ordenadores, los cuales forman parte de una nueva manera de ejercer la atención y cuidado de la salud en la población.

2.4. Mediciones

Para la valoración de la salud mental, se utilizó el Cuestionario de Salud General de Goldberg (The 12-Item General Health Questionnaire (GHQ-12)) (Isabel et al., 2013). Esta técnica fue ideada como un cuestionario auto administrativo. La fuente de los ítems de esta escala se basa en estudios precedentes sobre enfermedades mentales y la vivencia clínica de los compañeros de trabajo de Goldberg. De acuerdo a su metodología este cuestionario presenta tres subdimensiones que evalúan y examinan el estado de salud mental de la persona en el último mes respecto a su estado de salud mental. Las subdimensiones que se miden son: la autoestima que corresponde a los ítems 6, 9, 10, 11 y 12, el estrés que corresponde a los ítems 2, 5 y 7 y el afrontamiento que corresponde a los ítems 1, 3, 4 y 8. El Cuestionario de Salud General de Goldberg fue validado para ser aplicado en todo tipo de personas. Dentro de las ventajas de su aplicación es un instrumento confiable, ágil y simple, útil para llevar a cabo una detección del riesgo de mala salud mental en la población. De acuerdo a la escala se asigna 0 a las dos primeras respuestas de cada enunciado y 1 a las dos últimas, de esta forma Goldberg adquiriría un puntaje total, que mostraba evidencia de la presencia de disturbios mentales más severos, mientras mayor fuese su valor. El baremo de calificación es ausencia de psicopatología (0 – 4), sospecha de psicopatología (5 – 6) y presencia de psicopatología (7 – 12).

La intervención de psicoterapia a través de Telesalud consistió en 10 sesiones virtuales sincrónicas tipo taller que duraban 60 minutos durante 6 semanas (octubre a noviembre de 2020) a cargo de personal de enfermería con experiencia en el tema. El programa se diseñó en forma específica para la población según las subdimensiones propuestas

por Goldberg (autoestima, estrés y afrontamiento). Cada sesión estaba estructurada de la siguiente forma: presentación, momento lúdico, tema principal y final de la sesión. Además las 10 sesiones se agruparon en actividades. Para autoestima las actividades desarrolladas fueron: “amo como soy” (1. Autovaloración; 2. Valores y principios), y “¿qué siento?” (3. Sentimientos: formas de expresarse’ 4. Control interno: expresión de emociones). Para afrontamiento las actividades fueron: “yo sí puedo” (5. Orientación y actitud frente a los problemas; 6. Generación de soluciones alternativas) y “motivándome” (7. Sueños; 8. Establecer metas). Para estrés la actividad fue: “disfruto la vida” (9. Desarrollar las percepciones sensoriales; 10. Sumergirse en el presente).

Se realizaron capacitaciones al equipo de investigación y colaboradores para aplicar los instrumentos en forma estandarizada. Previo a la ejecución, se realizó una prueba piloto en una población similar de otra iglesia para verificar la comprensión de la ficha de recolección de datos. Posteriormente, previa coordinación con el encargado de la iglesia se obtuvieron los permisos necesarios para obtener la información de contacto de la población de estudio, y posteriormente se les contactó por vía telefónica.

2.5. Sesgos

En el grupo control se tuvo un sesgo hacia la derecha con un 60% de jóvenes con presencia de alguna psicopatología, lo cual indica, según el cuestionario de salud general de Goldberg, que este grupo presenta niveles bajos de autoestima, baja capacidad de afrontamiento a los problemas y elevados índices de estrés.

En el grupo experimental en el pretest se observa un sesgo hacia la derecha, lo que indica la presencia de alguna psicopatología, sin embargo, en el postest el sesgo se dirige hacia la izquierda, lo cual indica que la mayoría de los individuos presentan ausencia de psicopatología.

2.6. Tamaño muestral

La población de estudio estuvo conformada por jóvenes de una iglesia cristiana en un pueblo joven de Miraflores. La muestra estuvo conformada por 60 jóvenes entre los 18 y 29 años, la cual fue repartida de manera aleatoria donde los números pares son el grupo control y los impares el grupo experimental.

2.7. Análisis estadístico

Los datos fueron analizados con el programa SPSS - IBM 24, donde se crearon las tablas de frecuencia y de contingencia, así como la realización de pruebas estadísticas como McNemar para la comparación de la variable dependiente (salud mental) del antes y después de cada intervención (telesalud vs no intervención). Se consideró un valor de $p < 0.01$ como estadísticamente significativo.

3. Resultados

En total se estudiaron a 60 personas, sus características sociodemográficas se presentan en la Tabla 1, donde se observa que el porcentaje de dichas características son similares

en cada grupo de intervención, predominando población masculina (60%), de 18 a 20 años (40%), de grado de instrucción superior (~ 50%) y estado civil soltero (80%).

Características	Grupo Experimental		Grupo Control	
	N	%	N	%
Sexo				
Femenino	12	40	13	43.3
Masculino	18	60	17	56.7
Edad				
De 18 a 20 años	12	40	12	40
De 21 a 23 años	4	13	3	10
De 24 a 26 años	2	7	7	23
De 27 a 29 años	12	40	8	27
Grado de Instrucción				
Primaria	0	0	0	0
Secundaria	14	47	5	17
Superior Técnico	2	7	7	23
Superior Universitario	14	47	18	60
Estado Civil				
Soltero (a)	24	80	24	80
Casado (a)	6	20	3	10
Conviviente	0	0	3	10
Divorciado (a)	0	0	0	0
Total	30	100	30	100

Tabla 1 – Características sociodemográficas de la población de estudio según grupo de intervención

Salud Mental (Cuestionario de Goldberg)	Grupo Experimental				Valor p
	Pretest		Posttest		
	N	%	N	%	
Ausencia de psicopatología	2	7	22	73	0.000025*
Sospecha de psicopatología	4	13	6	20	
Presencia de psicopatología	24	80	2	7	
Total	30	100	30	100	

*Prueba de McNemar

Tabla 2 – Comparación de salud mental en el grupo experimental antes y después de la intervención

En la Tabla 2 se observa que en el pretest 28 jóvenes (93%) presentaron sospecha o presencia de psicopatología, mientras que en el posttest sólo 8 jóvenes (27%) presentaron sospecha o presencia de psicopatología, resultando que el grupo que recibió la intervención de psicoterapia por telesalud tuvo una mejora significativa en su salud mental al comparar la salud mental del pretest y el posttest ($p = 0.000025$).

En la Tabla 3 se observa que en el pretest 20 jóvenes (67%) presentaron sospecha o presencia de psicopatología, mientras que en el posttest 22 jóvenes (73%) presentaron sospecha o presencia de psicopatología, resultando que el no realizar intervención alguna en el grupo no tuvo efecto significativo en su salud mental al comparar la salud mental del pretest y el posttest ($p = 0.316$).

Salud Mental (Cuestionario de Goldberg)	Grupo Control				Valor p
	Pretest		Posttest		
	N	%	N	%	
Ausencia de psicopatología	10	33	8	27	0.316*
Sospecha de psicopatología	2	7	6	20	
Presencia de psicopatología	18	60	16	53	
Total	30	100	30	100	

*Prueba de McNemar

Tabla 3 – Comparación de salud mental en el grupo control antes y después de la intervención

En la Tabla 4 se observa que en el grupo experimental, la intervención de psicoterapia por telesalud tuvo un efecto significativo en cada una de las dimensiones: autoestima ($p = 0.0008$), afrontamiento ($p = 0.000046$) y estrés ($p = 0.0000057$). Además, se observa que después de aplicada la intervención, existe una disminución del 83% a 27% en el número de jóvenes con presencia de problemas en la autoestima y el incremento de 3% a 46% en jóvenes que no presentan ningún problema en la autoestima. Posteriormente, se observa una disminución del 72% al 7% en el número de jóvenes con presencia de problemas en el afrontamiento y el incremento de 14% a 66% en jóvenes que no presentan ningún problema en el afrontamiento de problemas. Finalmente, se observa una disminución del 83% al 3% en el número de jóvenes con presencia de problemas en el estrés y el incremento de 1% al 90% en jóvenes que no presentan ningún problema en el estrés.

Salud Mental (Cuestionario de Goldberg)	Grupo Control				Valor p
	Pretest		Posttest		
	N	%	N	%	
Autoestima					
Ausencia de psicopatología	1	3	14	46	0.0008*
Sospecha de psicopatología	4	14	8	27	
Presencia de psicopatología	25	83	8	27	

Salud Mental (Cuestionario de Goldberg)	Grupo Control				Valor p
	Pretest		Postest		
	N	%	N	%	
Afrontamiento					
Ausencia de psicopatología	4	14	20	66	0.000046*
Sospecha de psicopatología	4	14	8	27	
Presencia de psicopatología	22	72	2	7	
Estrés					
Ausencia de psicopatología	1	3	27	90	0.0000057*
Sospecha de psicopatología	4	14	2	7	
Presencia de psicopatología	25	83	1	3	
Total	30	100	30	100	

*Prueba de McNemar

Tabla 4 – Comparación de las dimensiones de salud mental en el grupo experimental antes y después de la intervención

4. Discusión y conclusión

La salud mental ha sido y es una parte primordial para el desarrollo y bienestar integral de la persona. Sin embargo, el poco estudio y conocimiento de esta área ha traído como consecuencia que gran parte de la población joven en el Perú no logre este desarrollo a plenitud, acarreando consecuencias como: una deficiente educación, carencias económicas, alcoholismo, entre otros problemas tanto físicos, psicológicos y sociales. Estos problemas se encuentran concentrados en mayor proporción dentro de las zonas más pobres del Perú, más aún en época de pandemia (Piazza & Fiestas, 2014).

Respecto a la deficiente educación en el Perú, el Ministerio de Educación (MINEDU), creó una Política Nacional de Juventud (PNJ) que tuvo como objetivo incrementar el acceso a la educación superior de jóvenes, basándose en los datos brindados por la Organización Mundial de la Salud (OMS) que el 2.3% de la población mundial es superdotada y que esa condición no tiene un carácter social o económico, lo que quiere decir que hijos de modestos campesinos u obreros pueden tener la capacidad de poder estudiar estudios superiores. En este proceso de inclusión social y en el marco del cumplimiento orientado a desarrollar competencias en el proceso educativo de la población joven, se crea el Programa Nacional de Becas y Crédito Educativo (Pronabec).

Es así que, de acuerdo con el informe del Instituto Nacional de Estadística e informática (INEI), en la población de pobreza y pobreza extrema se aprecia que el 30% de peruanos tienen algún año de estudios superiores. En el Perú, las universidades estatales brindan una educación gratuita, por lo cual la población estudiantil en su mayoría está constituida por jóvenes categorizados en situación de pobreza y pobreza extrema según el SISFOH (sistema de información sobre las características socioeconómicas de los hogares peruanos).

Asimismo, la crisis vivida durante los meses de pandemia que ha vivido el mundo entero ha incrementado la problemática de salud mental a nivel global (Cortes, 2020). Por lo tanto, esta crisis ha aumentado considerablemente la necesidad de implementar la adquisición y desarrollo de tecnología de comunicación a distancia como lo es el internet, el cual, en las zonas rurales era muy escasa y tenía bastantes inconvenientes (Perales, 2020). Toda esta necesidad, fomentó la implementación de la tecnología y los recursos necesarios para la aplicación de estas intervenciones en telesalud que hasta antes de la pandemia se había visto muy limitada (Huarcaya-Victoria, 2020).

En base a estos datos se puede corroborar en los resultados que tanto el grupo control, como el grupo experimental antes, de aplicar el test, reportan resultados con un nivel estadísticamente alto de sospecha y presencia de alguna psicopatología (alrededor del 80%) en jóvenes entre 18 y 29 años, el cual, según datos del INEI en el 2015, es el grupo con mayor crecimiento en el Perú (Instituto Nacional de Estadística e Informática, 2016).

Posterior a la intervención de telesalud en el grupo experimental se puede observar la disminución significativa de casos de problemas de psicopatología en jóvenes. Estos resultados coinciden con lo evidenciado en estudios previos, donde la intervención por telesalud mostró ser similar a la realizada en forma presencial en cuanto a eficacia (Barrera-Valencia et al., 2017; Hungerbuehler et al., 2016; Martínez et al., 2014; Osma et al., 2016). La efectividad de la intervención realizada en el presente estudio puede explicarse por su carácter sincrónico, que a diferencia de su contraparte asincrónica, presenta mayores tasas de adherencia (Barrera-Valencia et al., 2017; Hungerbuehler et al., 2016; Sánchez-Zapata & Zapata, 2019). Además, se conoce que la población joven está más familiarizada con la tecnología, lo que facilita su uso y satisfacción en la misma (Cortes, 2020; Huarcaya-Victoria, 2020; Martínez et al., 2014; Osma et al., 2016), reflejada en la aceptación de estas intervenciones de telesalud observada durante la fase experimental.

Los indicadores que define Goldberg para determinar los problemas psicopatológicos presentes en la persona son la autoestima, el afrontamiento a los problemas y el estrés (Isabel et al., 2013). En el presente estudio se observa un efecto importante de la psicoterapia por telesalud en cada una de estas tres dimensiones y en forma global, resultados similares a otros estudios (Barbosa & Silva, 2017; Fidalgo et al., 2021; Morón & Aguayo, 2018).

Con el fin de fortalecer estas intervenciones, es importante además que puedan establecerse marcos normativos éticos y de competencias que permitan garantizar un adecuado uso de estas tecnologías, como la confidencialidad de datos e información, la seguridad y transmisión de los datos, entre otros. (Morón & Aguayo, 2018)

En síntesis, este trabajo evidencia que la telesalud promueve la salud mental en jóvenes de una iglesia cristiana de un pueblo joven de Miraflores, demostrándose en los resultados del grupo experimental, después de aplicarse la intervención de telesalud.

Agradecimientos

La presente investigación agradece a los participantes de la iglesia cristiana de Miraflores, Arequipa, Perú y al equipo de investigación y colaboradores, por su participación y aporte al desarrollo de la misma.

Autorización Ética

En cuanto a las consideraciones éticas, la participación en el estudio fue de forma voluntaria brindando un consentimiento informado, asegurando el anonimato de los resultados y además se cumplieron con las disposiciones de la declaración de Helsinki y conducta responsable en investigación.

Referencias

- Barbosa, I. D. A., & Silva, M. J. P. D. (2017). Nursing care by telehealth: what is the influence of distance on communication?. *Revista brasileira de enfermagem*, 70, 928-934. <https://doi.org/10.1590/0034-7167-2016-0142>
- Barnett, J. E. (2019). The ethical practice of psychotherapy: Clearly within our reach. *Psychotherapy*, 56(4), 431-440. <https://doi.org/10.1037/pst0000272>
- Barrera-Valencia, C., Benito-Devia, A. V., Vélez-Álvarez, C., Figueroa-Barrera, M., & Franco-Idárraga, S. M. (2017). Cost-effectiveness of synchronous vs. asynchronous telepsychiatry in prison inmates with depression. *Revista Colombiana de Psiquiatría (English ed.)*, 46(2), 65-73. <https://doi.org/10.1016/j.rpeng.2017.05.005>
- Campos, N. A., Rivera, R. J. C., Aldon, L. F. V., Quispe, G. B. G., & Caballero, J. E. A. P. (2022). Salud mental de los docentes que realizan trabajo virtual en tiempo de pandemia. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (E48), 320-331.
- Cortes, M. E. (2020). COVID-19 como amenaza a la salud mental de los adolescentes. *Revista Peruana de Investigación en Salud*, 4(4), 192-194. <https://doi.org/10.35839/repis.4.4.757>
- El Peruano. (2019). Decreto legislativo que fortalece los alcances de la telesalud. <http://busquedas.elperuano.pe/normaslegales/decreto-legislativo-que-fortalece-los-alcances-de-la-telesal-decreto-legislativo-n-1490-1866212-2/>
- Fidalgo, Á. P., Ruiz, C. D., Vega, M. A. C., & Moreno, A. D. (2021). Revisión del uso de la videoconferencia en atención de salud mental en el contexto hispanohablante. *Revista de ciencias sociales*, 27(3), 87-106.
- García-Salirrosas, E. E., Cieza-Villavicencio, J. S., Pastor, A. B., & Mariano, A. M. (2023). Revisión bibliométrica de la aceptación y uso de la tecnología de la información y comunicación con el uso de la teoría del UTAUT aplicado a la telemedicina. *RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação*, (E56), 516-527.

- Hilty, D. M., Maheu, M. M., Drude, K. P., & Hertlein, K. M. (2018). The need to implement and evaluate telehealth competency frameworks to ensure quality care across behavioral health professions. *Academic Psychiatry*, 42(6), 818-824. <https://doi.org/10.1007/s40596-018-0992-5>
- Huarcaya-Victoria, J. (2020). Consideraciones sobre la salud mental en la pandemia de COVID-19. *Revista Peruana de Medicina Experimental y Salud Pública*, 37(2), 327-334. <https://doi.org/10.17843/rpmesp.2020.372.5419>
- Hungerbuehler, I., Valiengo, L., Loch, A. A., Rössler, W., & Gattaz, W. F. (2016). Home-based psychiatric outpatient care through videoconferencing for depression: a randomized controlled follow-up trial. *JMIR Mental Health*, 3(3), e5675. <https://doi.org/10.2196/mental.5675>
- Instituto Nacional de Estadística e Informática. (2016). Encuesta Nacional sobre Relaciones Sociales ENARES 2013 y 2015 (Vol. 1). Lima: Gráfica Burgos.
- Martínez, O., Jometón, A., Pérez, M., Lázaro, E., Amayra, I., López-Paz, J. F., ... & Bárcena, J. E. (2014). Effectiveness of teleassistance at improving quality of life in people with neuromuscular diseases. *The Spanish Journal of Psychology*, 17, E86. <https://doi.org/10.1017/sjp.2014.89>
- MINSA. (2016). Salud mental comunitaria: Nuevo modelo de atención. <https://www.gob.pe/institucion/minsa/informes-publicaciones/284981-salud-mental-comunitaria-nuevo-modelo-de-atencion>
- MINSA. (2017). Situación de salud de los adolescentes y jóvenes en el Perú. Documento técnico. <https://www.gob.pe/institucion/minsa/informes-publicaciones/284802-situacion-de-salud-de-los-adolescentes-y-jovenes-en-el-peru-documento-tecnico>
- MINSA. (2018). Plan nacional de fortalecimiento de servicios de salud mental comunitaria 2018 - 2021. Documento técnico. <https://www.gob.pe/institucion/minsa/informes-publicaciones/279690-plan-nacional-de-fortalecimiento-de-servicios-de-salud-mental-comunitaria-2018-2021-documento-tecnico>
- Morón, J. J. M., & Aguayo, L. V. (2018). La psicoterapia on-line ante los retos y peligros de la intervención psicológica a distancia. *Apuntes de Psicología*, 36(1-2), 107-113. <https://doi.org/10.55414/ap.v36i1-2.717>
- OMS. (2018a). Salud mental: fortalecer nuestra respuesta. <https://www.who.int/es/news-room/fact-sheets/detail/mental-health-strengthening-our-response>
- OMS. (2018b, September 28). Día Mundial de la Salud Mental 2018. Pan American Health Organization / World Health Organization. https://www3.paho.org/hq/index.php?option=com_content&view=article&id=14710:world-mental-health-day-2018&Itemid=42091&lang=es
- OPS/OMS. (2019). Prevención del suicidio. <https://www.paho.org/es/temas/prevencion-suicidio>

- Osma, J., Crespo, E., & Castellano, C. (2016). Terapia Cognitivo-Conductual multicomponente para el tratamiento del Trastorno Explosivo Intermitente por videoconferencia: un estudio de caso. *Anales de Psicología*, 32(2), 424-432. <http://dx.doi.org/10.6018/analesps.32.2.211351>
- Perales, A. (2020). Ética, salud mental y COVID-19. *Acta Médica Peruana*, 37(4), 532-535.
- Piazza, M., & Fiestas, F. (2014). Prevalencia anual de trastornos y uso de servicios de salud mental en el Perú: resultados del estudio mundial de salud mental, 2005. *Revista peruana de medicina experimental y salud pública*, 31(1), 30-38.
- Sánchez-Zapata, P., & Zapata, J. F. (2019). Telesalud Y Telemedicina Para El Manejo De La Epilepsia. *Revista Ecuatoriana de Neurología*, 28(1), 63-69.
- Stoll, J., Müller, J. A., & Trachsel, M. (2020). Ethical issues in online psychotherapy: A narrative review. *Frontiers in psychiatry*, 10, 993. <https://doi.org/10.3389/fpsyt.2019.00993>
- Villa, I. C., Arboleda, C. Z., & Roldan, L. F. R. (2013). Psychometric properties of the general health Goldberg GHQ-12 questionnaire applied at a hospital facility in the city of Medellin. *Avances en Psicología Latinoamericana*, 31(3), 532-546.

Sistemas conversacionais: aspetos culturais influenciam a avaliação de UX de um protótipo conceptual de assistente de voz

Juliana Camargo¹, José Nunes¹, Maria João Antunes¹, Óscar Mealha¹,
Carolina Abrantes² e Luís Nóbrega²

julianacamargo@ua.pt; jnunes@ua.pt; mariajoao@ua.pt; oem@ua.pt;
cabrantes@broadvoice.com; lnobrega@broadvoice.com

¹ DigiMedia, Universidade de Aveiro, Aveiro, Portugal

² GoTelecom, S.A, Aveiro, Portugal

DOI: 10.17013/risti.50.118–135

Resumo: Este artigo analisa a experiência de 100 indivíduos (50 brasileiros e 50 portugueses) no contexto de um sistema conversacional com assistente de voz automatizado. A avaliação da UX foi realizada em diferentes etapas: cálculo da estatística descritiva, análise das frases ditas pelos utilizadores e aplicação da escala SUS. Esta abordagem metodológica permitiu verificar que o protótipo conceptual simula um produto com grande potencial para facilitar o dia a dia dos profissionais que atuam em salões de estética e beleza. No entanto, há funcionalidades que devem ser redesenhadas, como a escolha dos nomes dos botões e a acessibilidade da ferramenta. Outra contribuição do estudo é a confirmação de que há opiniões divergentes entre as populações da amostra, brasileira e portuguesa, mostrando que a solução teve melhor aceitação entre os portugueses. Esta última questão é um dos aspetos mais relevantes dos achados deste trabalho, pois pode comprometer o sucesso de tal assistente.

Palavras-chave: sistemas conversacionais, chatbots, assistentes virtuais, inteligência artificial, acessibilidade, experiência do utilizador (UX).

Conversational systems: cultural aspects influence the UX evaluation of a conceptual voice assistant prototype

Abstract: This paper analyzes the experience of 100 individuals (50 Brazilians and 50 Portuguese) in the context of a conversational system with an automated voice assistant. The UX evaluation consisted of calculating descriptive statistics, analyzing the phrases spoken by users during the tests and applying the SUS scale. This methodological approach made it possible to verify that the conceptual prototype simulates a product with great potential for facilitating the daily lives of professionals working in beauty salons. However, there are features that need to be redesigned, such as the choice of button names and the accessibility of the tool. Another contribution is the confirmation that there are divergent opinions between

the Brazilian and Portuguese sample populations, showing that the solution was better accepted by the Portuguese. This last issue is one of the most relevant aspects of this study's findings, as it could jeopardize the success of such an assistant.

Keywords: conversational systems, chatbots, virtual assistants, artificial intelligence, accessibility, User Experience (UX).

1. Introdução

A avaliação da Experiência do Utilizador (UX) tem ganhado cada vez mais relevância na investigação académica (Lallemant & Koenig, 2020). Tal disciplina consiste em verificar o comportamento de um grupo de pessoas durante a utilização inicial de um produto, sistema ou serviço (Pettersson et al., 2018). É importante ressaltar que esse tipo de estudo vai além da observação das respostas dadas pelos participantes durante os testes: com uma metodologia adequada, é possível detetar características hedônicas e emocionais, como estética, estímulo e identificação, que são extremamente relevantes em qualquer tipo de interação entre humanos e sistemas (Bernhaupt & Pirker, 2014).

Uma avaliação bem estruturada, portanto, não se limita à verificação da usabilidade do produto ou da solução a ser testada. Ela deve ser construída a partir de uma combinação de técnicas capazes de analisar todos os aspetos envolvidos na jornada do utilizador (Ludwig et al., 2021). Trata-se de uma atividade multidisciplinar, uma vez que envolve várias áreas do conhecimento, como design, engenharia e psicologia (Pettersson et al., 2018).

Quando nos referimos a um sistema conversacional, tema central deste estudo, a análise de UX é importante para validar recursos e propor melhorias (Guerino & Valentim, 2020). Este tipo de sistema consiste no uso de tecnologias capazes de compreender a linguagem humana e conduzir uma conversa entre pessoas e dispositivos, de forma natural, que pode acontecer por meio da escrita ou da fala (Trippas et al., 2020). Alguns exemplos mais recorrentes são os assistentes pessoais, como Siri e Alexa, e os chatbots, muito utilizados no atendimento ao cliente.

A utilização mais ampla desta tecnologia tem sido amplamente suportada por avanços significativos nas áreas de Processamento de Linguagem Natural (PLN), Inteligência Artificial (IA) e Aprendizado de Máquina. Estas abordagens permitem uma interação mais fluida e consistente entre seres humanos e computadores, justificando seu uso em situações cotidianas (Kiseleva et al., 2016). Considerando que estas abordagens são sustentadas pelas necessidades e preferências dos utilizadores, o software depende de uma avaliação de UX consistente para identificar se a solução proposta é adequada. Isto geralmente é feito por meio de técnicas específicas, como entrevistas, observação, narrativas e diários.

O objetivo deste artigo é, portanto, relatar e discutir os resultados da avaliação de UX de um protótipo conceptual de um sistema de conversação. Este estudo empírico contou com uma amostra de 100 participantes, sendo metade da população brasileira e metade composta por cidadãos portugueses. O paper foi organizado da seguinte forma: o primeiro passo foi a revisão da literatura. Em seguida, na segunda secção, há uma breve descrição do protótipo conceptual e suas diferentes iterações de design. A terceira secção

relata o teste de experiência do utilizador, ou seja, quais técnicas foram empregadas para a coleta de dados e sua análise. Por fim, a quarta secção apresenta os resultados e sua discussão, com destaque para os principais achados e contribuições deste trabalho.

2. Estudos relacionados

Antes de entrar na análise dos métodos utilizados para avaliar os sistemas de conversação, é importante esclarecer a diferença entre Experiência do Utilizador (UX) e usabilidade. Esta última, segundo a norma ISO 9241: Parte 11, mede a eficácia, a eficiência e a satisfação de um produto num contexto de utilização específico, com base num conjunto pré-determinado de objetivos e tarefas (International Organization for Standardization, 2010). Relativamente à experiência do utilizador, de acordo com a ISO 9241-110:2010 (cláusula 2.15), trata-se da análise das respostas geradas a partir da utilização inicial de um produto, sistema ou serviço. Por conseguinte, são conceitos distintos, precisamente porque os dados da avaliação de UX dependem essencialmente de emoções, sentimentos e perceções de cada participante durante a utilização de um produto.

A recolha de informações sobre a experiência do utilizador pode ocorrer de várias formas: por meio de entrevistas, observações, questionários validados ou uma combinação de técnicas. Um exemplo de abordagem metodológica mista é o estudo realizado por Abreu et al. (2021) para avaliar a UX de um sistema de interação por voz no contexto da televisão interativa (iTV). Os investigadores optaram por uma triangulação de instrumentos: AttrakDiff, System Usability Scale (SUS) e Self-Assessment Manikin (SAM), complementados por entrevistas semiestruturadas.

Iniesto et al. (2021) também apostaram em uma combinação de técnicas para avaliar a UX de um assistente virtual destinado a estudantes com deficiência física. Os investigadores utilizaram a observação direta, questionários (aplicados antes e depois das atividades) e entrevistas semiestruturadas. Além disso, pediram aos participantes que preenchessem dois questionários quantitativos: SUS e SUISSQ-R, uma versão simplificada do questionário Speech User Interface Service Quality (SUISQ).

Outro exemplo foi conduzido por Brüggemeier et al. (2020), que adotaram escalas validadas para avaliar a UX referente à Alexa, a assistente pessoal da Amazon, especificamente em termos de seleção de músicas. Este estudo utilizou os questionários Subjective Assessment of Speech System Interfaces (SASSI), SUISSQ-R, SUS e AttrakDiff. Ao longo do processo de avaliação, os participantes foram questionados sobre quais instrumentos consideravam mais eficazes para medir ou classificar as suas perceções, tendo o questionário SASSI sido identificado como a escolha preferida. O estudo começou com um teste inicial para identificar os dados pessoais e o estado emocional dos participantes. Posteriormente, utilizaram o Computer System Usability Questionnaire (CSUQ), que se baseia na usabilidade.

Procedimentos semelhantes foram realizados por Ludwig et al. (2021) e Le Pailleur et al. (2020), que objetivaram identificar aspetos emocionais relacionados a um sistema inteligente empregando uma combinação de instrumentos, como o questionário SAM e entrevistas semiestruturadas.

Ainda sobre a melhoria da experiência do utilizador, o estudo conduzido por Watterachchi et al. (2022) teve como objetivo desenvolver um modelo de dispositivo móvel que possa tomar decisões acertadas – e em tempo real – sobre funcionalidades com base nas emoções e no contexto do utilizador. Os investigadores aplicaram diversos inquéritos e, a partir das respostas dos utilizadores, criaram protótipos utilizando a técnica de prototipagem evolutiva. O modelo foi considerado aceitável como um sistema adaptativo com controlo do utilizador, capaz de sugerir funções com base nas emoções e no contexto.

Importa ressaltar que a análise destes trabalhos aqui apresentados foi crucial para apoiar as escolhas dos instrumentos utilizados na avaliação da UX relativa ao protótipo desenvolvido no âmbito deste estudo, cuja descrição pode ser encontrada na secção seguinte.

3. Protótipo

O protótipo conceptual apresentado neste artigo é derivado do projeto CallBob - Voice Bots Builder, que funciona como um serviço de mediação de chamadas telefónicas. O objetivo foi o desenvolvimento de um software adaptado a pequenas e médias empresas, permitindo-lhes receber e gerir marcações (por voz) automaticamente, via chamadas telefónicas. A partir deste objetivo, foi criado um protótipo de assistente automatizado utilizando o software InVision¹.

O produto, portanto, simula a capacidade de responder a ligações automaticamente e realizar tarefas específicas, como agendar horários e fornecer informações sobre os serviços oferecidos. A gestão de conteúdo é realizada pelos próprios empreendedores em uma plataforma que pode ser acedida via internet. Cada gerente é responsável por administrar os elementos de fala do robô e definir o conjunto de respostas a serem dadas aos clientes durante suas interações com o sistema.

O protótipo foi projetado a partir de um processo de investigação participativa, incorporando opiniões de profissionais de várias áreas. Estas informações foram coletadas por meio de grupos focais e entrevistas individuais com clientes e gerentes de pequenas empresas. Com base nos dados empíricos obtidos, foram desenvolvidas narrativas de comunicação que funcionaram como suporte para que os diálogos entre humanos e o sistema fossem eficazes e naturais. O resultado é um modelo de interação intuitivo para a configuração do assistente virtual, voltado a utilizadores com níveis de literacia digital de baixa a média.

4. Protocolo e amostra

A amostra do estudo é composta por proprietários e/ou profissionais que atuam em salões de beleza. Para simular a utilização da ferramenta, a equipa de investigação propôs seis cenários com base nas solicitações feitas com maior frequência pelos clientes (identificadas pelos empreendedores que participaram dos grupos focais iniciais). Tais

¹ Disponível em: <https://www.invisionapp.com/>.

cenários e tarefas estavam relacionados com as principais funcionalidades do software, incluindo: 1) gestão do fluxo inicial de serviços, 2) configuração da área de agendamento, 3) personalização dos serviços oferecidos, 4) gestão de pedidos, 5) definição do horário de funcionamento e localização e 6) configuração das informações dos clientes. O método utilizado para analisar a experiência do utilizador em relação aos cenários envolveu uma combinação de técnicas capazes de identificar múltiplos aspetos, como estimulação, satisfação, controlo e aceitação:

- Questionário inicial: os dados pessoais e as informações profissionais foram recolhidos por meio de um questionário;
- Observação direta: os participantes utilizaram o protótipo enquanto narravam os seus pensamentos e forneciam as suas perceções;
- Avaliação dos cenários: no final da apresentação de cada cenário, os participantes classificaram a sua experiência numa escala nominal de 1 a 6, em que 1 indicava uma experiência “altamente negativa” e 6 indicava uma experiência “altamente positiva”;
- Escala de usabilidade do sistema (SUS): todos os participantes preencheram um questionário SUS, que mede a usabilidade do protótipo;
- Avaliação qualitativa do discurso dos participantes: os comentários, sugestões, elogios e observações relevantes feitos pelos participantes foram categorizados com base nos seguintes critérios: “altamente negativo” (-2), “negativo” (-1), “neutro” (0), “positivo” (1) e “altamente positivo” (2).

Os testes foram realizados online, por meio de um software de videoconferência, de fevereiro a maio de 2020, devido ao contexto pandémico. Um total de 100 profissionais foram convidados a participar do estudo, sendo 50 de Portugal e 50 do Brasil. A escolha da amostra se deu por conveniência. Dos 100 participantes selecionados, 63% são mulheres e 37% homens, com uma média de idade de 37 anos. Todos os indivíduos partilhavam a característica comum de trabalharem em salões de beleza.

5. Avaliação do protótipo e descobertas

Os resultados desta secção apresentam as classificações de experiência dos participantes (numa escala nominal de 6 pontos) para cada um dos 6 cenários que lhes foram apresentados. As estatísticas descritivas das classificações de experiência para cada cenário estão representadas na **Figura 1** como um diagrama de radar. O diagrama de radar inclui os valores do primeiro quartil, do terceiro quartil e da mediana da amostra (n=100) para cada um dos 6 cenários que constituem o protótipo conceptual.

O diagrama de radar na **Figura 1** mostra que as perceções gerais dos participantes estavam, na sua maioria, alinhadas a uma experiência “altamente positiva”. A classificação mediana para todos os cenários apresentados foi de 5. No entanto, os participantes expressaram uma opinião mais negativa especificamente em relação ao cenário 6, que diz respeito à secção de informações sobre o cliente. Este aspeto apresentou a menor amplitude entre o primeiro quartil (Q1), com 4 (limite exterior do branco interior), e o terceiro quartil (Q3), com 5 (limite exterior do sector azul). Em geral, os participantes reconheceram a utilidade desta funcionalidade para os salões.

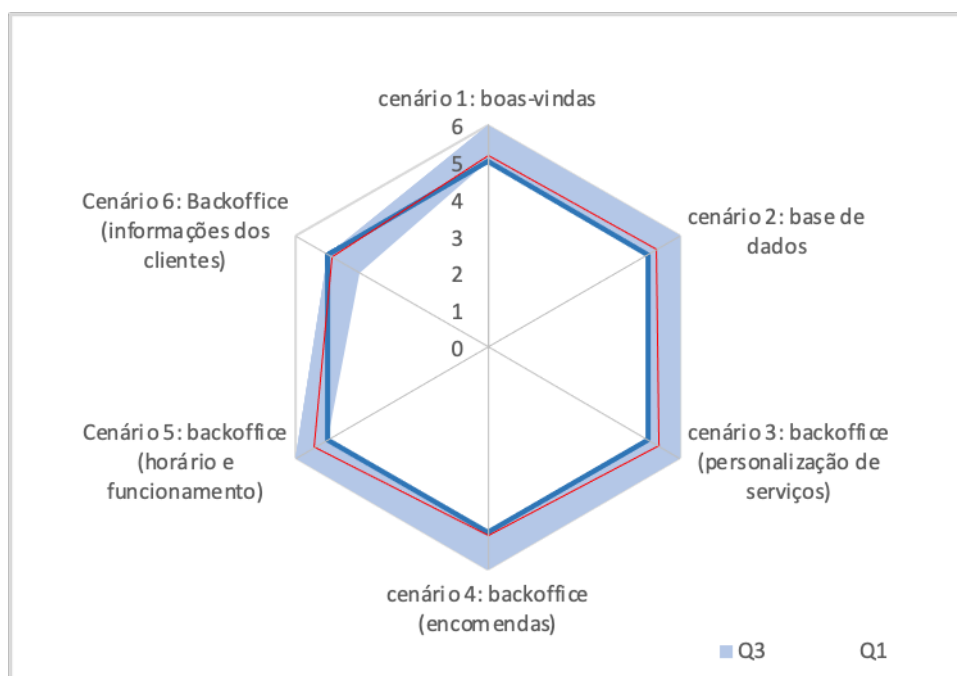


Figura 1 – Mediana, primeiro quartil (Q1) e terceiro quartil (Q3) das pontuações atribuídas pelos participantes a cada cenário proposto.

No entanto, 44 participantes sugeriram que o recurso poderia recolher dados mais pormenorizados sobre os clientes. Os seus comentários incluíam afirmações como: “Faltam informações relevantes, como o tipo de corte de cabelo feito e os produtos utilizados” e “Gosto muito da área de informações, mas acho que poderia ser mais detalhada, com a data de nascimento, por exemplo, e o tipo de corte de cabelo”. Destas 44 sugestões, 41 foram feitas por participantes brasileiros, indicando uma disparidade de opiniões entre os participantes das duas nacionalidades. Isto pode ser visto na **Figura 2** e na **Figura 3**, com os participantes brasileiros a exibirem uma reação menos positiva do que os participantes portugueses em todos os cenários, especialmente no 6. Uma possível razão para esta diferença é que 46 participantes brasileiros mencionaram que as marcações por telefone já não são o método mais comum utilizado nos salões do país. Afirmaram ainda que dependem atualmente de marcações feitas pela aplicação WhatsApp, o que pode ter contribuído para uma certa resistência em relação a algumas das funcionalidades apresentadas. Um participante comentou: “Seria mais interessante ter uma integração deste sistema com o WhatsApp, permitindo-nos enviar notificações pela aplicação.” Este contraste não foi observado entre os participantes portugueses, que indicaram utilizar prioritariamente o telefone para marcação de consultas.

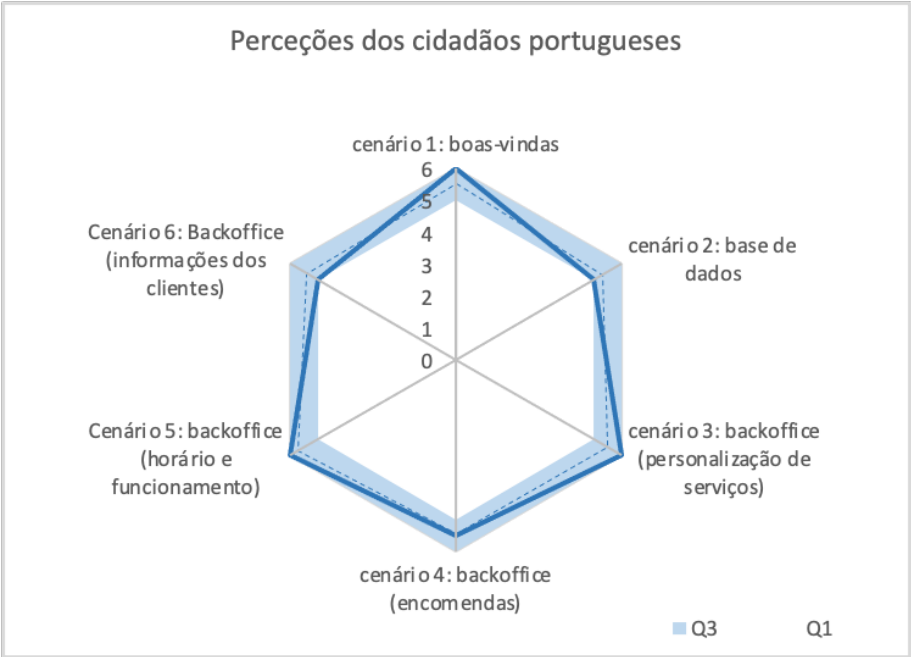


Figura 2 – Percepção dos cidadãos portugueses sobre os cenários apresentados.

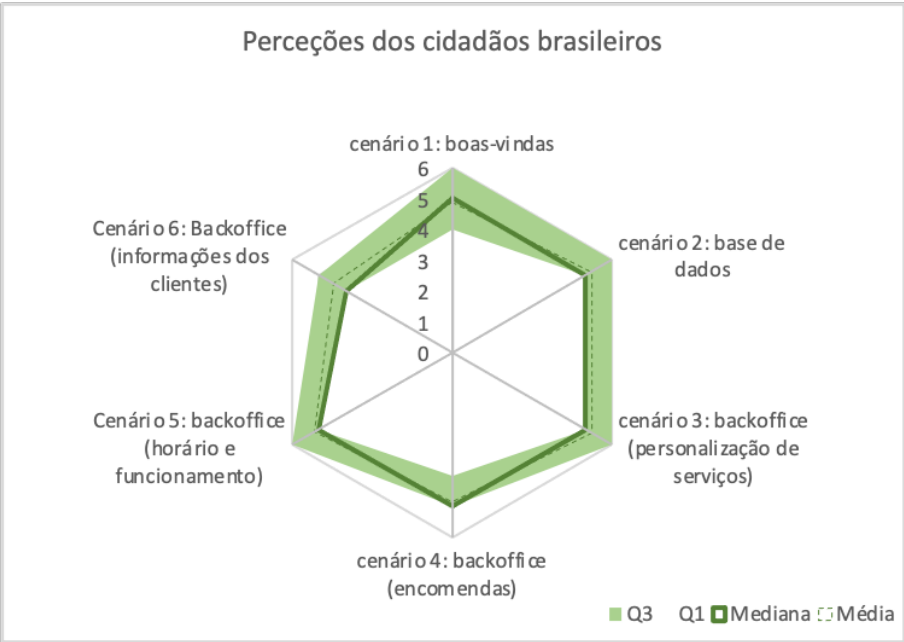


Figura 3 – Percepções dos cidadãos brasileiros sobre os cenários apresentados.

Relativamente ao questionário SUS, que foi utilizado para avaliar a usabilidade do software, a pontuação obtida foi de 80,83, indicando uma classificação entre “bom” e “excelente”. Isso significa que os participantes consideraram a ferramenta fácil de usar, mas também identificaram áreas de melhoria. Esta conclusão é apoiada pela análise individual de cada teste, em que os participantes forneceram críticas e sugestões para o sistema.

A secção seguinte apresenta os resultados destas análises, com base na recolha e organização das opiniões dos participantes. Foram analisadas 1.128 frases, classificadas nas seguintes categorias “altamente negativas” (em casos de fortes críticas ou dificuldades na realização das tarefas), “negativas” (críticas moderadas e pequenas dificuldades na realização das tarefas), “neutras” (aspectos não positivos ou negativos mencionados), “positivas” (tarefas realizadas sem ajuda) e “altamente positivas” (tarefas facilmente realizadas com elogios).

5.1. Gestão do fluxo de serviço inicial (cenário 1)

A tarefa de gestão do fluxo inicial do serviço consistia em selecionar as mensagens que seriam comunicadas pelo bot de voz aos clientes. Durante os testes, dois aspetos foram mencionados com frequência pelos participantes: “UX Writing” (mencionada 100 vezes) e “Design gráfico” (mencionada 62 vezes). A maioria (84%) deu um feedback positivo relativamente à “UX Writing”. Isto indica que boa parte dos entrevistados considerou que as mensagens apresentadas nesta etapa eram claras e concisas, melhorando a interação com o sistema (**Figura 4**).

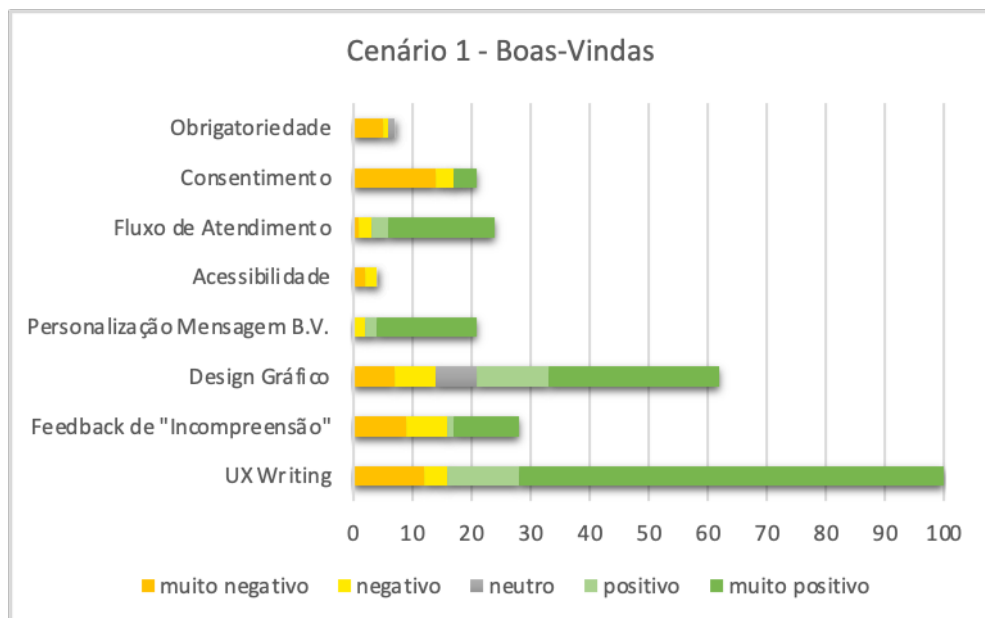


Figura 4 – Percepções sobre o cenário 1, que consistia em configurar o fluxo de boas-vindas.

Do total de 100 menções, apenas 15% expressaram aspetos negativos relativamente à gestão do fluxo inicial do serviço. Por exemplo, os participantes referiram ter dúvidas ou dificuldades em compreender certas partes do processo, como a inclusão de uma mensagem de boas-vindas ou a forma de começar a utilizar o sistema. Foram feitas sugestões no sentido de fornecer instruções mais claras ou um tutorial para facilitar a utilização.

Relativamente ao aspeto “design gráfico”, 46,7% (29) das menções foram “muito positivas” e 20% (7) foram “positivas”. Os participantes apreciaram a representação visual do fluxo do serviço, afirmando que esta proporcionava uma compreensão clara do funcionamento do sistema e das informações a preencher. No entanto, 11% (7) das menções foram classificadas como “negativas” e 11% (7) como “muito negativas”, principalmente devido a uma perceção de falta de atratividade do design. Adicionalmente, 11% (7) das menções foram consideradas neutras.

No cenário 1, três aspetos se destacaram com maior incidência de feedback negativo: “feedback de incompreensão”, “acessibilidade” e “consentimento”. Em relação ao “feedback de incompreensão”, que se refere à mensagem que o bot emitiria quando não entendesse a entrada do cliente, houve 28 menções. Destas, 57,1% (16) foram negativas e 42,8% (12) foram positivas. Os participantes que deram feedback negativo expressaram confusão quanto ao número de mensagens a registar no sistema, uma vez que havia espaço para três.

Todas as menções (4) sobre a acessibilidade da plataforma foram negativas, sendo 50% (2) classificadas como “negativas” e 50% (2) como “muito negativas”. De um modo geral, os participantes consideraram o tamanho do texto demasiado pequeno, o que compromete a legibilidade.

Relativamente ao aspeto “consentimento”, que consistia em criar uma mensagem para notificar os clientes sobre a utilização dos seus dados, houve 21 menções. Apenas 19% (4) dessas menções foram classificadas como “positivas”. Os participantes manifestaram dificuldades em compreender o conceito de consentimento e em identificar a mensagem de consentimento, o que levou a sugestões de explicações mais claras.

5.2. Configuração da área de marcação de consultas (cenário 2)

Nesta área, os utilizadores têm de registar as informações obrigatórias que o bot deve pedir aos clientes antes de marcar uma consulta. Para além disto, existe um campo dedicado à inclusão de frases para formação do sistema. A categoria “pedidos de informação dos clientes” foi a mais mencionada pelos participantes, com um total de 90 menções (**Figura 5**). Dos que comentaram, 95% (86) acharam fácil incluir as informações obrigatórias, como nome, data, serviço proposto e horário. Os 5% restantes (4) tiveram uma dúvida comum sobre a diferença entre as opções “obrigar” e “ativar” disponíveis na interface.

O campo destinado à inclusão de frases para auxiliar no treinamento do sistema foi bem recebido pela maioria dos participantes. Das 39 menções, 74% (29) foram classificadas como “positivas”. Os participantes reconheceram a importância da inclusão de novas

frases, especialmente nos casos em que interagem com clientes que falam de forma diferente devido a variações regionais. Por exemplo, um participante mencionou ter muitos clientes brasileiros em Portugal e apreciou a capacidade de acomodar diferentes padrões de fala. No entanto, 26% (10) das menções foram negativas, principalmente devido à falta de compreensão da funcionalidade. Tal aspeto influenciou outro, classificado como “terminologia do botão”. Este tema foi mencionado 38 vezes, sendo que apenas 1 opinião foi positiva, indicando que a palavra “treino” não foi a melhor escolha para o botão ou que deveria haver alguma explicação adicional junto ao ícone. Do total de menções negativas, 73% (27) foram feitas por entrevistados brasileiros, mostrando que a variação linguística entre Brasil e Portugal pode impactar a experiência do utilizador.

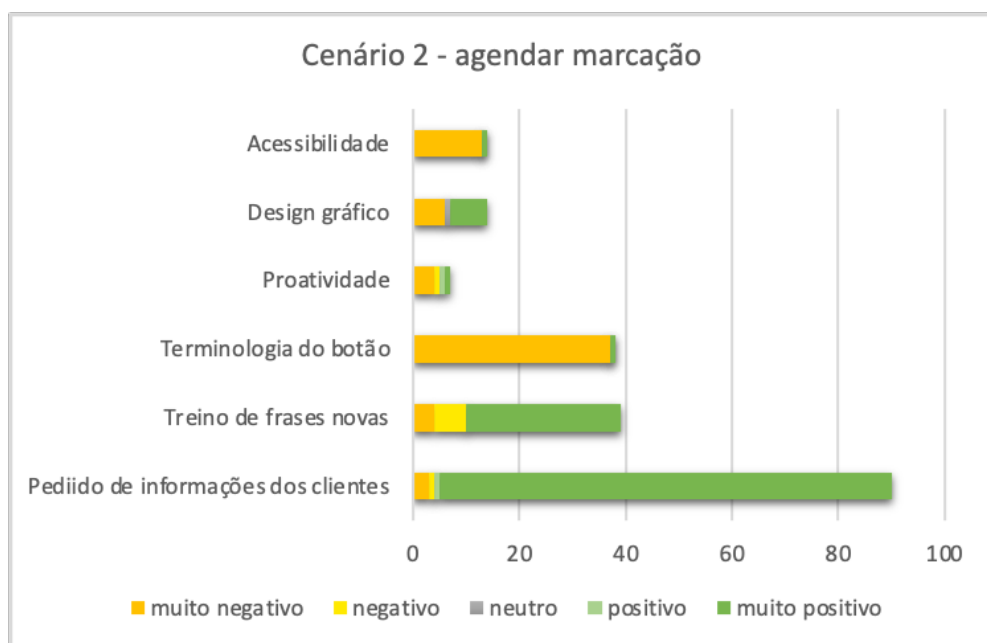


Figura 5 – Percepções em relação ao cenário 2, que consistia em configurar a área de marcações.

Outro aspeto mencionado 7 vezes foi a “proatividade”. Os participantes que o abordaram sentiram falta de uma componente proativa na plataforma. O “design gráfico”, por sua vez, foi citado 14 vezes, sendo 7 vezes “muito positivo”, 1 “neutro” e 6 “muito negativo”, mostrando que há espaço para melhorias. Alguns exemplos de sugestões dadas pelos participantes: “deveria ter uma cor diferente para cada opção do fluxo” e “o fluxo de atendimento deveria ser numerado, para que pudessemos seguir os passos”.

Por fim, a categoria “acessibilidade” foi mencionada 14 vezes, sendo que em 92% (13) das incidências o tamanho da fonte de letra foi questionado. “Deveria ser maior para facilitar a leitura”, disse um dos participantes.

5.3. Personalização dos serviços oferecidos (cenário 3)

Esta tarefa envolveu a adição de um novo serviço oferecido ao sistema. Os participantes consideraram este procedimento fácil de executar, tendo apenas uma pessoa tido dificuldades em encontrar a localização correta (**Figura 6**).

Outro aspeto que foi frequentemente mencionado nos testes foi a relevância da funcionalidade. No total, foram 37 menções, todas elas avaliadas como “muito positivas”. O aspeto da “terminologia dos botões” apareceu 14 vezes, com 7 menções “muito positivas” e 7 menções “muito negativas”. Os participantes consideraram que a escolha de “informações comerciais” como rótulo do botão não representava corretamente a funcionalidade.

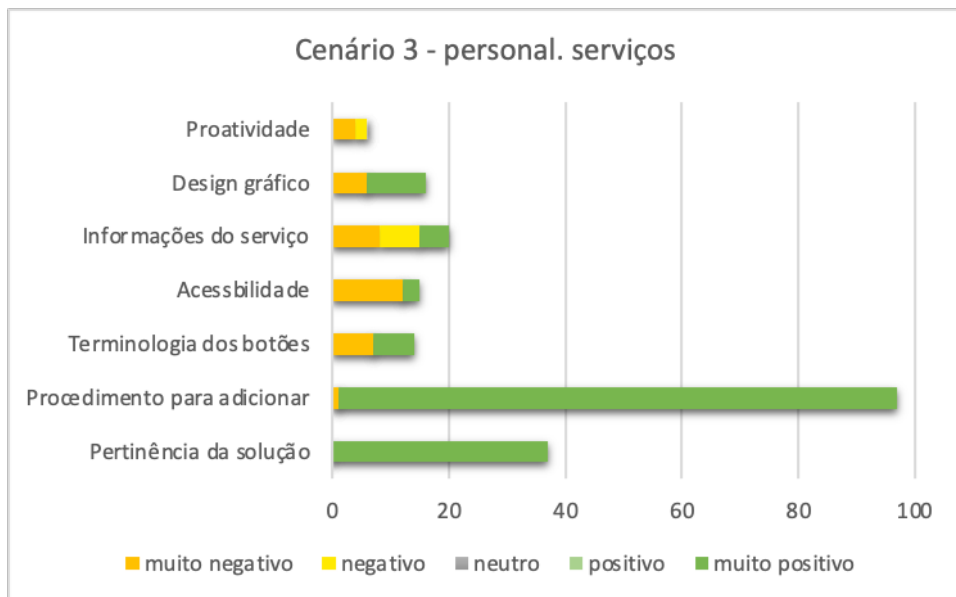


Figura 6 – Opiniões sobre o cenário 3, que consistia em incluir um novo serviço no sistema.

Assim como nos demais cenários, o aspeto “acessibilidade” foi mencionado pelos participantes. Das 15 citações, 80% (12) são “muito negativas”, referindo-se novamente ao tamanho escolhido para a fonte. “Deveria ter uma opção para aumentar as letras quando fosse mais conveniente”, disse um dos participantes.

No que diz respeito ao preenchimento das informações sobre o serviço a ser incluído, 15 entrevistados disseram que poderia haver mais campos para detalhá-lo. Por isso, consideraram este aspeto como “negativo” (7 pessoas) e “muito negativo” (8 pessoas).

5.4. Gestão das encomendas (cenário 4)

Neste cenário, os participantes foram encarregados de gerir uma encomenda efetuada por um cliente. O procedimento de marcação foi bem-sucedido, como indicado por

73 participantes que forneceram comentários classificados como “muito positivos”. Relativamente à “pertinência da solução”, apesar de ter sido elogiada por 25% (25/100) dos participantes, 12% (12/100) manifestaram preocupações quanto ao facto de poder não funcionar eficazmente (**Figura 7**). Apontaram, por exemplo, o risco de os clientes não se lembrarem do nome correto do produto quando fazem uma encomenda, uma vez que o sistema não sugere opções de forma proativa. Essa opinião foi mais prevalente entre os participantes brasileiros, representando 82% (9/12) das avaliações.

Entre os 32% (32/100) que questionaram o botão que leva à lista de pedidos, 81% (26/32) criticaram a escolha da terminologia. Entre estes, 88% (22/26) eram brasileiros, destacando mais uma vez os potenciais problemas relacionados com a variação linguística. Apenas 4% dos entrevistados sugeriram a inclusão de uma nova secção com o histórico de encomendas dos clientes.

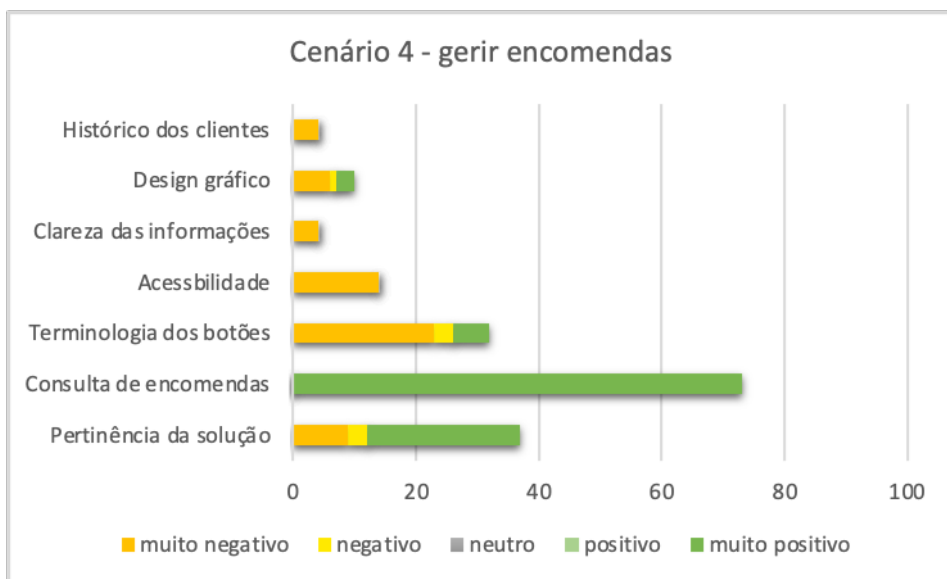


Figura 7 – Percepções sobre o cenário 4, que consistia na consulta de encomendas feitas pelos clientes.

5.5. Informações sobre horários e dias de funcionamento (cenário 5)

A tarefa atribuída para este cenário consistia em adicionar dados importantes ao sistema, como o horário de funcionamento e a localização dos salões. Entre os 100 entrevistados, 67% (67/100) mencionaram espontaneamente que conseguiram completar a tarefa de incluir a localização rapidamente, e elogiaram o fluxo escolhido. Apenas 1,4% (1) reconheceu que conseguiu completar a proposta, mas não fez nenhum elogio. No entanto, 2,98% (2) sentiram dificuldades na realização da tarefa, e seus comentários foram classificados como “negativos” (**Figura 8**).

As dificuldades foram mais proeminentes na tarefa de incluir o horário comercial. Entre os participantes, 84% (84/100) relataram uma experiência “muito positiva”, 1% (1/100) relatou uma experiência “positiva”, 4% (4/100) tiveram uma experiência “negativa” e 11% (11/100) tiveram uma experiência “muito negativa”. Entre os que tiveram dificuldades (12), 75% (9/12) eram brasileiros e 25% (3/12) eram portugueses. Alguns comentários dos participantes incluíram: “Não ficou claro se podíamos editar a hora introduzindo números ou se havia opções de hora pré-estabelecidas” e “Acho que devia haver um botão ou uma seta para indicar que é possível personalizar a hora”. Por outro lado, os participantes portugueses referiram a falta de clareza sobre qual o botão que deve ser ativado para personalizar a hora.

Os aspetos de “acessibilidade” e “design gráfico” voltaram a surgir, com sugestões de melhoria relacionadas com o tamanho da letra e a atratividade, respetivamente.

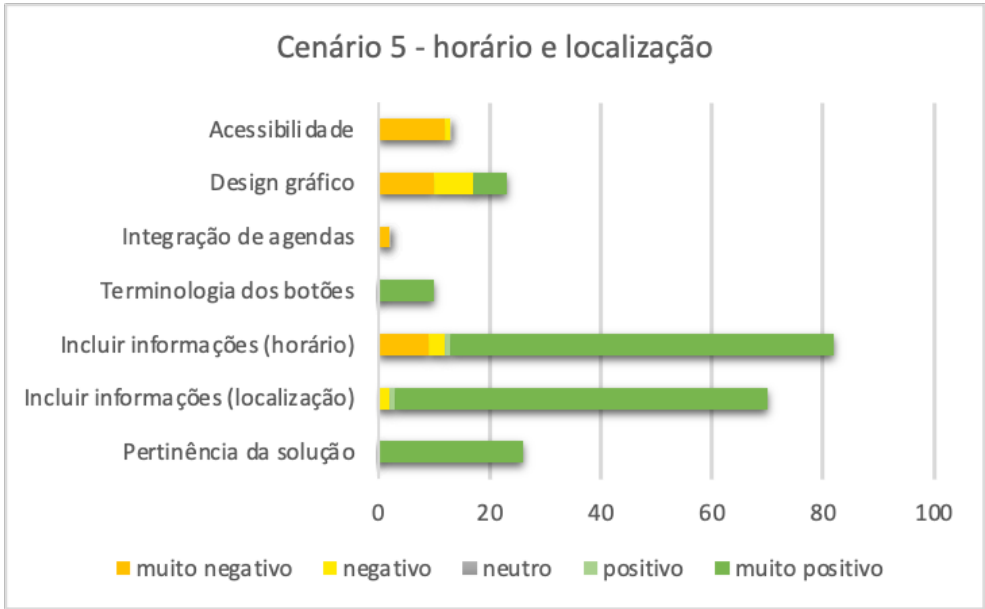


Figura 8 - Percepções sobre a inclusão do horário e da localização dos salões.

5.6. Informações sobre o horário e os dias de funcionamento (cenário 6)

Nesta tarefa, os participantes foram convidados a preencher informações sobre os seus clientes numa área especificamente destinada a armazenar estes dados. Do total de entrevistados, 63% (63/100) consideraram que se tratava de uma excelente opção para localizar mais facilmente os dados (**Figura 9**).

Foram feitos 80 comentários sobre a área de informação ao cliente, sendo 41% (33/80) classificados como “positivos”, 7,5% (6/80) como “negativos” e 51% (41) como “muito negativos”. Do total de opiniões negativas (47), 89% (42/47) foram expressas por

cidadãos brasileiros. A principal crítica foi a falta de pormenor na informação aos clientes. Alguns exemplos de comentários recolhidos sobre este tema são: “Acho que a secção de informações está incompleta. Deveria ser mais detalhada”, “ainda faltam informações importantes, como estilo de corte, cor e cabeleireiro preferido”, “falta uma quantidade significativa de informações relevantes, como data de aniversário e cortes anteriores” e “faltam tantas informações que acho que os dados disponíveis poderiam ser mais detalhados”. Conforme observado na **Figura 1**, essa área teve a menor percepção entre os entrevistados, principalmente entre os participantes brasileiros.

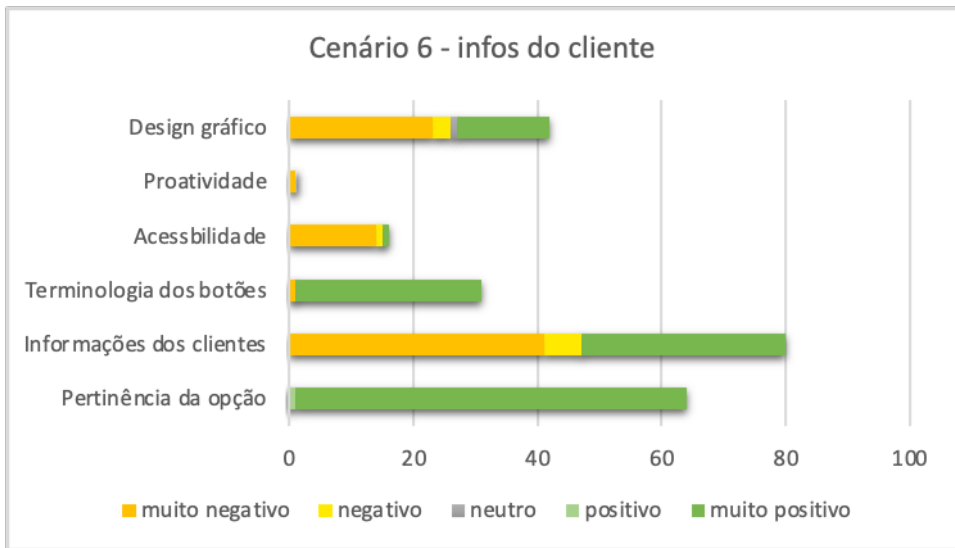


Figura 9 – Percepções sobre a área com os dados dos clientes.

Em relação aos aspetos “acessibilidade” e “design gráfico”, houve críticas quanto ao pouco espaço disponibilizado para os dados. Quarenta por cento dos participantes mencionaram que a área era muito pequena para a quantidade de informações que precisavam ser inseridas, resultando em consequências significativas para ambas as categorias.

6. Conclusões

A análise de UX descrita neste artigo empregou uma combinação de técnicas para avaliar tanto as qualidades hedônicas quanto pragmáticas do produto. Essas técnicas incluíram a avaliação de identificação, interesse, incentivo e atratividade. Todo o processo, incluindo seus elementos e etapas, foi detalhado no artigo. Um dos principais componentes da avaliação envolveu os participantes classificando cada cenário proposto em uma escala ordinal de 1 a 6. Além disso, uma análise abrangente de conteúdo dos comentários espontâneos dos participantes durante os testes de experiência de UX foi realizada e discutida.

Estatísticas descritivas foram utilizadas pelos pesquisadores para processar as avaliações de experiência dos participantes, que representaram dados qualitativos. A mediana, primeiro quartil e terceiro quartil, que são adequados para escalas ordinais, foram usados para fornecer uma opinião holística da amostra do estudo para cada cenário. Gráficos de radar foram utilizados para aumentar a perspetiva geral e resumir as opiniões dos entrevistados, destacando as divergências entre as populações portuguesa e brasileira.

A análise dos comentários dos participantes foi essencial para entender as avaliações atribuídas a cada cenário. Isso permitiu identificar áreas que não foram bem aceitas ou que poderiam ser melhoradas. A sistematização dos comentários levou à identificação de 28 sugestões de melhoria, incluindo renomear botões, abordar questões de acessibilidade (12 sugestões), integração com plataformas (6), confirmações via SMS (4) e informações detalhadas sobre os clientes (6).

Além disso, para medir a usabilidade geral do protótipo conceitual, todos os utilizadores foram convidados a completar um questionário SUS. Os resultados indicaram que a solução proposta foi útil e relativamente fácil de usar, embora algumas inconsistências tenham sido identificadas.

No geral, o método de P&D apresentado neste artigo mostrou-se eficiente na análise da experiência do utilizador, resultando em melhorias diretas na interface do sistema de conversação e na configuração de dados. Além disso, as diversas técnicas de coleta de dados empregadas ajudaram a identificar diferenças percebidas entre os dois grupos: cidadãos brasileiros e portugueses, em termos da configuração da interface do sistema de conversação.

Quanto à comparação entre os grupos, foi possível identificar um conjunto de diferenças na forma como avaliaram a plataforma. Isto é uma indicação importante de que diferenças culturais podem interferir na experiência do utilizador de um sistema. Por exemplo, uma parte significativa dos brasileiros não considerou a terminologia dos botões suficientemente clara. Em geral, eles acharam que os nomes escolhidos dificultam a localização de informações, o que não foi o caso dos cidadãos portugueses. Este comportamento se deve, em grande parte, às diferenças entre as palavras utilizadas no português do Brasil e no português de Portugal. No geral, os cidadãos brasileiros reagiram menos positivamente a todos os cenários, especialmente ao cenário 6. Uma possível razão para essa diferença, de acordo com 92% (46/100) dos entrevistados brasileiros, é que agendar compromissos por telefone deixou de ser o método mais comum usado em salões de beleza. A ferramenta mais frequentemente adotada é o WhatsApp. Portanto, este é outro exemplo importante de como diferenças culturais podem impactar a percepção dos utilizadores sobre sistemas e plataformas. Daí a necessidade de uma amostra diversificada para considerar a diversidade de opiniões em relação a novos produtos e soluções.

Agradecimentos. Este trabalho é apoiado pelo Fundo Europeu de Desenvolvimento Regional (FEDER), por meio do Programa Operacional de Competitividade e Internacionalização (COMPETE 2020) do quadro Portugal 2020 [Projeto CallBob com Número 038501 (POCI-01-0247-FEDER-038501)] e também é apoiado financeiramente por fundos nacionais por meio da FCT - Fundação para a Ciência e Tecnologia, I.P., no âmbito do projeto UIDB/05460/2020.

Referências

- Abreu, J., Camargo, J., Santos, R., Almeida, P., Beça, P., & Silva, T. (2021). UX Evaluation Methodology for iTV: Assessing a Natural Language Interaction System. In *Communications in Computer and Information Science* (Vol. 1433). https://doi.org/10.1007/978-3-030-81996-5_11
- Bernhaupt, R., & Pirker, M. M. (2014). User interface guidelines for the control of interactive television systems via smart phone applications. *Behaviour and Information Technology*. <https://doi.org/10.1080/0144929X.2013.810782>
- Brüggemeier, B., Breiter, M., Kurz, M., & Schiwy, J. (2020). User Experience of Alexa when controlling music: Comparison of face and construct validity of four questionnaires. *ACM International Conference Proceeding Series*, (pp. 1-9). <https://doi.org/10.1145/3405755.3406122>
- Guerino, G. C., & Valentim, N. M. C. (2020). Usability and User eXperience Evaluation of Conversational Systems: A Systematic Mapping Study. *ACM International Conference Proceeding Series*, (pp. 427–436). <https://doi.org/10.1145/3422392.3422421>
- Iniesto, F., Coughlan, T., & Lister, K. (2021). Implementing an accessible conversational user interface: Applying feedback from university students and disability support advisors. *Proceedings of the 18th International Web for All Conference, W4A 2021*, (pp. 1-5). <https://doi.org/10.1145/3430263.3452431>
- International Organization for Standardization. (2010). ISO 9241-210: Ergonomics of human–system interaction - Human-centred design for interactive systems. In *International Organization for Standardization*.
- Kiseleva, J., Williams, K., Jiang, J., Awadallah, A. H., Crook, A. C., Zitouni, I., & Anastasakos, T. (2016). Understanding user satisfaction with intelligent assistants. *CHIIR 2016 - Proceedings of the 2016 ACM Conference on Human Information Interaction and Retrieval*, (pp.121-130). <https://doi.org/10.1145/2854946.2854961>
- Lallemand, C., & Koenig, V. (2020). Measuring the Contextual Dimension of User Experience: Development of the User Experience Context Scale (UXCS). *ACM International Conference Proceeding Series*, (pp. 1-13). <https://doi.org/10.1145/3419249.3420156>
- Le Pailleur, F., Huang, B., Léger, P. M., & Sénécal, S. (2020). A New Approach to Measure User Experience with Voice-Controlled Intelligent Assistants: A Pilot Study. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-030-49062-1_13
- Ludwig, R., Bachmann, A., Buchholz, S., Ganser, K., Glänzer, D., & Matarage, A. (2021). How to measure ux and usability in today's connected vehicles. *Advances in Intelligent Systems and Computing*. https://doi.org/10.1007/978-3-030-55307-4_3

- Pettersson, I., Lachner, F., Frison, A. K., Riener, A., & Butz, A. (2018). A bermuda triangle? - A review of method application and triangulation in user experience evaluation. *Conference on Human Factors in Computing Systems – Proceedings*, (pp. 1-16). <https://doi.org/10.1145/3173574.3174035>
- Trippas, J. R., Spina, D., Thomas, P., Sanderson, M., Joho, H., & Cavedon, L. (2020). Towards a model for spoken conversational search. *Information Processing and Management*, 57(2), 102162. <https://doi.org/10.1016/j.ipm.2019.102162>
- Wattearachchi, W. D., Hewagamage, K. P., & Hettiarachchi, E. (2022). Emotional Keyboard: To Provide Adaptive Functionalities Based on the Current User Emotion and the Context. *Interaction Design and Architecture(s) Journal*, (54), 147 – 174. <https://doi.org/10.55612/s-5002-054-007>

Critérios Editoriais

A RISTI (Revista Ibérica de Sistemas e Tecnologias de Informação) é um periódico científico, que foca a investigação e a aplicação prática inovadora no domínio dos sistemas e tecnologias de informação.

O Conselho Editorial da RISTI incentiva potenciais autores a submeterem artigos originais e inovadores para avaliação pelo Conselho Científico.

A submissão de artigos para publicação na RISTI deve realizar-se de acordo com as chamadas de artigos e as instruções e normas disponibilizadas no sítio Web da revista (<http://www.risti.xyz/>).

Todos os artigos submetidos são avaliados por um conjunto de membros do Conselho Científico, não inferior a três elementos.

Em cada número da revista são publicados entre cinco a oito dos melhores artigos submetidos.

Criterios Editoriales

La RISTI (Revista Ibérica de Sistemas y Tecnologías de la Información) es un periódico científico, centrado en la investigación y en la aplicación práctica innovadora en el dominio de los sistemas y tecnologías de la información.

El Consejo Editorial de la RISTI incentiva autores potenciales a enviar sus artículos originales e innovadores para evaluación por el Consejo Científico.

El envío de artículos para publicación en la RISTI debe hacerse de conformidad con las llamadas de los artículos y las instrucciones y normas establecidas en el sitio Web de la revista (<http://www.risti.xyz/>).

Todos los trabajos enviados son evaluados por un número de miembros del Consejo Científico de no menos de tres elementos.

En cada número de la revista se publican cinco a ocho de los mejores artículos enviados.



Revista Ibérica de Sistemas e Tecnologias de Informação
Revista Ibérica de Sistemas y Tecnologías de Información

©RISTI 2023 <http://www.risti.xyz>

