

## Procesamiento del Lenguaje Natural y Inteligencia Artificial para Analizar Tendencias de Búsqueda de Medicamentos Potencialmente Peligrosos

Cembreros-Castaño D.<sup>1</sup>, Hervás-Pérez J.P.<sup>2\*</sup>, Ponte Cosentini M.E.<sup>3</sup>

ORIGINAL ARTICLE

### RESUMEN

**Introducción:** Este estudio presenta el desarrollo de la aplicación “Searching Help<sup>®</sup>”, diseñada para monitorear y prevenir comportamientos autolíticos en individuos a partir del análisis de su actividad en el móvil durante la búsqueda de medicamentos que pueden implicar un riesgo para la vida del usuario<sup>1</sup>.

Este estudio aplica técnicas de Procesamiento de Lenguaje Natural (PNL) e Inteligencia Artificial (IA) para predecir búsquedas potencialmente peligrosas de medicamentos mediante el análisis de las tendencias de búsqueda en Google Trends<sup>®2-3</sup>. La identificación de estos patrones de búsqueda es crucial para mejorar la capacidad de detección de comportamientos de riesgo de la aplicación “Searching Help<sup>®</sup>”, permitiendo una intervención temprana para prevenir posibles actos autolíticos<sup>4</sup>. Así, la implementación de IA y PNL en este contexto supone una innovadora estrategia de prevención en el campo de la salud mental.

**Material y Métodos:** Selección de medicamentos: Se determinó un conjunto de medicamentos utilizando como referencia ATC. Datos: las búsquedas se hicieron sobre 9 categorías de medicamentos compuestas en total por 66 principios activos y 137 nombres comerciales. Recopilación de Datos: se descargaron datos de las búsquedas en Google Trends<sup>®</sup> relacionadas con los medicamentos seleccionados, enfocándonos en las tendencias en España durante los últimos cinco años. Los datos obtenidos fueron preprocesados, lo que incluye la eliminación de búsquedas duplicadas y la normalización de variantes de palabras. Aplicación de PLN e IA: Se procesaron los datos con Sketch Engine<sup>®</sup>, una herramienta especializada en PLN, para descomponer las búsquedas en componentes semánticos y sintácticos y determinar las búsquedas más comunes. Posteriormente, se utilizó ChatGPT<sup>®</sup>, un modelo de lenguaje basado en IA, para analizar estos componentes, determinar las relaciones entre diferentes términos de búsqueda y categorizar las búsquedas. La IA también se utilizó para predecir nuevas búsquedas potencialmente peligrosas basándose en patrones de lenguaje identificados.

**Resultados:** Se obtuvieron 1391 cadenas de búsquedas frecuentes de los principios activos y nombres comerciales de los medicamentos, de las que, 1040 son búsquedas frecuentes en los últimos 5 años y 351 son búsquedas con un aumento inusual reciente. Se procedió al preprocesado de la información para utilizar formas lingüísticas estándar, y a la agrupación de las cadenas de búsqueda por campos semánticos. Posteriormente,

<sup>1</sup> Facultad de Educación, Universidad Camilo José Cela, Urb. Villafranca del Castillo, Calle Castillo de Alarcón, 49, 28692 Villanueva de la Cañada, Madrid, España.

<sup>2</sup> Faculty of Health, Camilo José Cela University, Urb. Villafranca del Castillo, Calle Castillo de Alarcón, 49, 28692 Villanueva de la Cañada, Madrid, España.

<sup>3</sup> SearchingHelp. C/ José Echegaray, 14. 28232 Las Rozas – Madrid.

\*Autor correspondiente: jphervas@ucjc.edu.

se procedió a la categorización y etiquetado de dichas búsquedas con el objetivo de discernir entre aquellas que presentan menor riesgo para la salud en contraposición a aquellas que pueden ser consideradas más peligrosas. Tras la completa clasificación de las búsquedas en las categorías mencionadas, se procedió a emplear la técnica de modelado de lenguaje ChatGPT® para desarrollar modelos predictivos destinados a la identificación de nuevas búsquedas potencialmente peligrosas relacionadas con los medicamentos previamente analizados. Tras la categorización y aplicar el modelado de IA, se obtuvieron un total de 1391 búsquedas, que se muestran en la Figura 1.



Figura 1. Categorización de las búsquedas tras modelado por IA.

Estos resultados indican que solamente el 15% de las búsquedas que se hacen en Internet sobre medicamentos delicados son potencialmente peligrosas. A su vez, de las 209 búsquedas peligrosas, se obtienen 6 categorías. La IA ha generado con estos datos 145 modelos de formatos de cadenas de búsqueda potencialmente peligrosas (Ej.: tomar un determinado medicamento y alcohol). Al combinarse estas cadenas con la base de datos de medicamentos, da la capacidad de predecir un amplio espectro de consultas potencialmente peligrosas relacionadas con medicamentos. Este enfoque predictivo ha permitido contribuir de manera valiosa a la herramienta para la prevención y alerta temprana ante búsquedas peligrosas en el ámbito de la salud y la farmacología en línea.

**Conclusiones:** Este estudio aprovecha las capacidades avanzadas de PLN e IA para proporcionar una visión detallada de las tendencias de búsqueda de medicamentos potencialmente peligrosos. Sólo el 15% de las búsquedas en internet pudieran ser objeto de seguimiento por parte del personal sanitario. La IA detecta 5 categorías denominadas menos peligrosas, mientras que son 6 las denominadas categorías peligrosas detectadas. La IA es capaz de generar 145 modelos de cadenas de búsqueda genéricos.

### Referencias Bibliográficas:

1. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the Methods, Tools, and Statistical Approaches in Google Trends Research: Systematic Review. *J Med Internet Res*. [Internet]. 2018;20(11):e270. Published 2018 Nov 6. doi:10.2196/jmir.9366.
2. Ning Wang, L.F. Yuvraj S. Varsha B. Koduvayur S. Rajarathnam C. Ellen L. Learning Models for Suicide Prediction from Social Media Posts. In *Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology*: [Internet]. 2021;87–92, Online. Association for Computational Linguistics. doi:10.18653/v1/2021.clpsych-1.9.
3. Zafra Cremades S, Gómez Sorian J. M, , Navarro-Colorado B. Diseño, compilación y anotación de un corpus para la detección de mensajes suicidas en redes sociales. *Procesamiento del Lenguaje Natural* [Internet]. 2017; (59):65-72. Recuperado de: <https://www.redalyc.org/articulo.oa?id=515754427007>.
4. Kid Valeriano, Alexia Condori-Larico and José Sulla-Torres, “Detection of Suicidal Intent in Spanish Language Social Networks using Machine Learning” *International Journal of Advanced Computer Science and Applications(IJACSA)*, 11(4), 2020.
5. <http://dx.doi.org/10.14569/IJACSA.2020.0110489>.

## Implementation of the Case Study Method in the Chemical-Biological Area with the Help of Artificial Intelligence

Weiss-Steider B.<sup>1</sup>, Romero-Altamirano I.S.<sup>1\*</sup>, Vázquez R.S.<sup>1</sup>, Vargas-Angeles C.A.<sup>1</sup>, Rangel-Corona R.<sup>1</sup>, Ortega M.T.C.<sup>1</sup>

ORIGINAL ARTICLE

### ABSTRACT

**Introduction:** The case study method is a teaching methodology that presents students with real-world situations, for them to analyze in groups and discuss solutions. It has been used in law, business, and medical schools since the 19th century, and has been shown to be effective for learning. The case study method has been used in the chemical-biological area to a limited extent, due to the need for high-quality cases with sufficient context. Artificial intelligence (AI) has emerged as a tool for generating accurate literary texts, and has been proposed as a solution to facilitate the use of the case study method in this area<sup>1-3</sup>. The objective of this study was to evaluate the utility of AI to implement the case study method in the chemical-biological area. To do this, AI prompts were designed to generate case studies.

**Methods:** AI prompts were created for case studies in the Bachelor of Pharmacy and Biology programs. They were based on case structures from the Case Center and improved through 300 interactions with the ChatGPT AI. 50 test cases were evaluated with 10 experts in the areas of educational psychology, higher education, genetics, pharmacology, pharmacy, chemistry, molecular biology, and cell biology, to determine their congruence with the professional field.

**Results:** The designed AI prompts allowed teachers to incorporate relevant elements from the student's environment, such as places, characters, or objects, to generate greater familiarity. They also provided the possibility of adjusting variables such as the length of the text and the area of knowledge, and defining the problem according to the established learning objectives. Although the evaluated cases showed consistency, a limitation was observed in the generation of real data and variables. While AI supported the creative and writing process, it did not replace the teacher's participation in defining learning.

**Conclusions:** The designed AI prompts were consistent and useful for writing cases. However, AI has limitations due to it does not have the specific knowledge about various fields. It is recommended to evaluate and enrich the products obtained with the professional knowledge of teachers.

**Keywords:** AI, case study method, teaching in pharmacy and chemistry.

<sup>1</sup> Facultad de Estudios Superiores Zaragoza. Universidad Nacional Autónoma de México (UNAM). Av. Guelatao no. 66, Colonia Ejército de Oriente, Iztapalapa C.P. 09230, Ciudad de México, México.

\*Corresponding author: tcvaldes@unam.mx.

**Acknowledgments:** The authors appreciate the support of DGAPA, UNAM, PE 211323 grant.

**Bibliographic References:**

1. Akgun, S., & Greenhow, C. (2021). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, 1-10. doi:10.1007/s43681-021-00067-8.
2. Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Available at SSRN: <https://ssrn.com/abstract=4337484>.
3. Dimitriadou, E., & Lanitis, A. (2023). A critical evaluation, challenges, and future perspectives of using artificial intelligence and emerging technologies in smart classrooms. *Smart Learning Environments*, 10(1), 1-26. doi:10.1186/s40561-022-00159-5.

## Leveraging Machine Learning for Protein Target Discovery in Natural Products

Dias A.L.<sup>1\*</sup>, Afonso M.M.<sup>2</sup>, Bernardes G.J.L.<sup>3,4</sup>, Rodrigues T.<sup>1</sup>

ORIGINAL ARTICLE

### ABSTRACT

**Introduction:** The discovery of new bioactive small molecules requires the exploration of uncharted and biologically relevant chemical spaces<sup>1</sup>. Natural products (NPs) are a privileged source of biologically active molecular scaffolds not yet probed by conventional synthetic small molecules<sup>2</sup>. This makes NPs a rich source for chemical space exploration and great starting points for hit/lead discovery. However, understanding the mode of action and target identification of NPs pose some challenges. In this work, we employed machine learning algorithms to explore a novel alkaloid NP, aiming to identify potential protein targets while circumventing the costs of chemoproteomics tools.

**Methodology:** Our approach utilized different machine learning algorithms, both supervised and unsupervised models, trained on a curated ChEMBL database (version 31). The database contained activity data of small molecules against up to 1167 and 2036 protein targets for supervised and unsupervised learning, respectively. The algorithms were designed to predict normalized activity values or cluster similar training molecules using CATS2 pharmacophore descriptors, and optimized by hyperparameter tuning. Model performance was evaluated using regression metrics, such as Coefficient of determination ( $r^2$ ), Mean Absolute Error (MAE), and Mean Squared Error (MSE) for supervised, and the percentage of true-positives/false-positives for unsupervised learning. All metrics were obtained on a 10-fold cross-validation.

**Results:** By integrating the predictions from both supervised and unsupervised models, we identified a GPCR target hit for the NP alkaloid, exhibiting 77.5% inhibition at 50  $\mu$ M.

**Discussion/Conclusion:** In conclusion, the application of machine learning led to the successful identification of a protein target for a novel alkaloid NP. This discovery enables further exploration of the NP scaffold, providing a platform for the development of novel synthetic high-quality molecular tools and understanding its biological mode of action. This study highlights the potential of machine intelligence in addressing challenges in medicinal chemistry and generating new knowledge from pre-existing chemical data.

**Keywords:** machine learning, natural products, alkaloids, activity prediction, ChEMBL Database.

<sup>1</sup> Research Institute for Medicines (iMed), Faculty of Pharmacy, University of Lisbon, Lisbon, Portugal.

<sup>2</sup> Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR/CIMAR), Porto, Portugal.

<sup>3</sup> Instituto de Medicina Molecular João Lobo Antunes, Faculty of Medicine, University of Lisbon, Lisbon, Portugal.

<sup>4</sup> Department of Chemistry, University of Cambridge, Cambridge, UK.

\*Corresponding author: ana.laura.dias@edu.ulisboa.pt.

**Acknowledgements:** This work was supported by the Fragment-Screen (101094131) project. Ana Laura Dias acknowledges the Fundação para a Ciência e Tecnologia (FCT) for the financial support from the PhD studentship 2022.09827.BD.

**Bibliographic References:**

1. Nelson A, Karageorgis G. Natural product-informed exploration of chemical space to enable bioactive molecular discovery. *RSC Medicinal Chemistry*. 2021;12(3):353-62.
2. Atanasov AG, Zotchev SB, Dirsch VM, the International Natural Product Sciences T, Supuran CT. Natural products in drug discovery: advances and opportunities. *Nature Reviews Drug Discovery*. 2021;20(3):200-16.