

## Resumen

La era digital ha marcado un antes y un después en la comunicación global, siendo una de las más relevantes la facilidad de compartir información, cada día se publican artículos científicos que son aportes muy importantes para la comunidad y estar al tanto de dichos avances es esencial, sin embargo, al existir tantos documentos digitales, el obtener información puntual suele ser una tarea que demanda mucho esfuerzo y tiempo. Para el caso de estudio se tomó en consideración los problemas que presentan los investigadores del área de Biotecnología de la Universidad de las Fuerzas Armadas – ESPE, donde a través de un análisis exploratorio, se determinó que existe una inadecuada gestión manual de la información en artículos científicos sobre nanopartículas. Por esta razón, el proyecto tiene como propósito desarrollar un modelo machine learning para la gestión de artículos científicos a través del cual se puedan establecer patrones de información que aporten al conocimiento científico.

Este proyecto se elaboró basándose en la metodología Design Science Research (DSR) y por medio de este, se definió e implementó el modelo machine learning para la generación de patrones de comportamiento en las investigaciones relacionadas con nanopartículas. Finalmente, se realizó la validación del modelo utilizando métricas de evaluación de rendimiento de clasificadores en tareas de extracción de datos y, además, al poner en marcha el modelo se obtuvieron resultados positivos donde se contrastaron los datos arrojados por el modelo y las anotaciones extraídas manualmente, determinándose que el modelo puede ser utilizado como una herramienta de apoyo para las investigaciones en el campo de las nanopartículas.

*Palabras clave:* machine learning, procesamiento del lenguaje natural, minería de textos, extracción automática de datos, nanopartículas.

## **Abstract**

The digital era has marked a before and after in global communication, being one of the most relevant the ease of sharing information, every day scientific articles are published that are very important contributions to the community and being aware of these advances is essential, however, with so many digital documents, obtaining timely information is usually a task that demands a lot of effort and time. For the case study, we took into consideration the problems presented by researchers in the area of Biotechnology at the Universidad de las Fuerzas Armadas – ESPE, where through an exploratory analysis, it was determined that there is an inadequate manual management of information in scientific articles on nanoparticles. For this reason, the project aims to develop a machine learning model for the management of scientific articles through which information patterns that contribute to scientific knowledge can be established.

This project was developed based on the Design Science Research (DSR) methodology and through this, the machine learning model was defined and implemented for the generation of behavioral patterns in research related to nanoparticles. Finally, the validation of the model was performed using classifier performance evaluation metrics in data extraction tasks and, in addition, when the model was implemented, positive results were obtained where the data yielded by the model and the manually extracted annotations were contrasted, determining that the model can be used as a support tool for research in the field of nanoparticles.

*Keywords:* machine learning, natural language processing, text mining, automatic data mining, nanoparticles.