



Análisis multiparamétrico de un sitio web especializado en medicina y tecnología utilizando inteligencia artificial mediante procesamiento de lenguaje natural (PLN)

»Luis Pino^{1,2,3}
»Eduardo Rico¹
»Iván Triana^{1,3}
»Liliana Mosquera¹
»Marco Anaya¹
»Rafael Brango¹
»Andrés Cardona⁴

¹ Centro de Pensamiento Aipocrates

² OxLER SAS

³ Fundación Santa Fe de Bogotá

⁴ Fundación Centro de Investigación y Tratamiento para el Cáncer "Carlos Sarmiento Angulo" CTIC

Introducción: Alpocrates es un centro de pensamiento para la inteligencia artificial en salud con profundización en medicina de alta complejidad incluyendo el cáncer, el cual fue fundado en junio de 2022 y cuenta con 20 miembros, entre ellos se incluyen cuatro hematólogos y oncólogos. El centro de pensamiento tiene una página web activa desde finales de julio de 2022 alojada en la dirección www.aipocrates.org, de libre acceso.

Metodología/Objetivos: se obtuvieron los datos crudos de utilización para el sitio web alojado en la plataforma WordPress para el periodo comprendido entre agosto de 2021 a agosto de 2022, con el fin de realizar:

1. Un análisis exploratorio de datos univariado, bivariado y multivariado.

2. Establecer coeficientes de correlación entre las variables definidas respecto a sus métricas de utilización.

3. Realizar un análisis mediante inteligencia artificial usando procesamiento de lenguaje natural (PLN) a los comentarios de los usuarios registrados en el sitio durante el periodo definido.

4. Definir estrategias de optimización derivadas del análisis multiparamétrico para sitios especializados en medicina de alta especialización y tecnologías.

El análisis se realizó sobre los datos crudos obtenidos desde la plataforma WordPress, el análisis exploratorio y el procesamiento de lenguaje natural utilizando el lenguaje Python versión 3.3.

Resultados: se alojaron 75 publicaciones en

Autor para correspondencia: Luis Pino

Correo electrónico: docpino2@gmail.com

Sociedad Colombiana de Hematología y Oncología. Este es un artículo Open Access bajo la licencia CC BY-NC-ND. (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

el período, consistentes en columnas escritas, infografías y comics relacionados con los temas de inteligencia artificial, genómica, gestión del riesgo, ciencia de datos, educación médica digital y tecnologías exponenciales. Se realizaron 164114 visitas realizadas por 16774 visitantes y un total de 62 comentarios. La Figura 1 muestra la tendencia de visitas.

Como se observa en la Figura 2, hay una correlación positiva entre el número de visitas y las vistas por persona, es decir que a mayor número de visitas (que no equivale al número de visitantes) hay mayor número de vistas por persona, pero se encuentra una correlación negativa entre el número de visitantes y el número de vistas por persona, lo que genera la hipótesis que en la medida en que el sitio web evoluciona y tiene mayor número de visitantes estos dejan de explorar al mismo y se enfocan en vistas de recursos específicos del mismo (ver la tendencia de agosto de 2022), es decir parecen “especializarse”.

De igual forma se realizó un gráfico de scatterplot con regresión (Figura 3) para evaluar la tendencia a la disminución en el número de vistas por persona en el tiempo. Como se deduce del gráfico hay una clara tendencia a la disminución progresiva en el número de vistas por persona que no se correlaciona con el número de visitantes y por tanto se confirma nuestra hipótesis de “especialización” progresiva del visitante.

Con relación al análisis de comentarios mediante procesamiento de lenguaje natural, en la Figura 4 se observa que las tres columnas que generaron mayor número de comentarios estuvieron relacionadas con el análisis de las aplicaciones digitales en salud, los tipos de modelos de ciencia de datos en salud y la educación médica 5.0., que corresponden en general a las tres líneas de énfasis del centro de pensamiento.

El análisis de sentimientos a los 42 comentarios se realizó utilizando una librería disponible en el software Python llamada Vader Sentiment (Figura 5). Un 56 % de los comentarios se clasificaron como neutros/negativos, por esto, se decidió hacer un análisis más profundo utilizando métricas de CountVectorizer y TF/IDF para determinar el peso de las palabras (Figura 6).

Conclusión: el análisis multiparamétrico de los datos de utilización en un sitio web médico especializado es de crucial importancia para diseñar estrategias de atractividad, adherencia y optimización, ya que define perfiles de usuarios, tendencias de uso y en este caso análisis de sentimientos que permiten reorientar los mensajes y el modelo de comunicación.

Palabras clave:

Inteligencia artificial; lenguaje natural; multiparamétrico; cáncer.

Figuras y Tablas:

Figura 1.

La tendencia de visitas

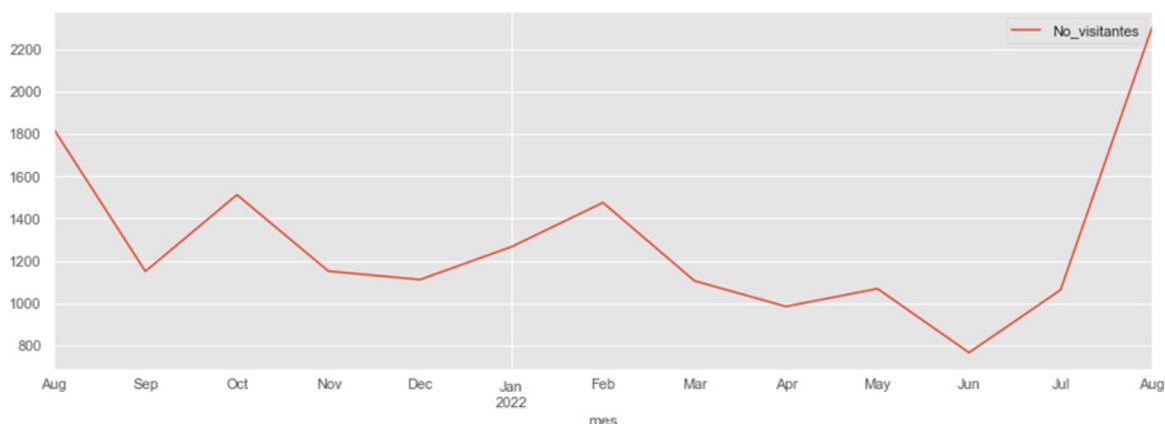




Figura 2.

Coefficientes de correlación



Figura 3.

Gráfico de scatterplot con regresión

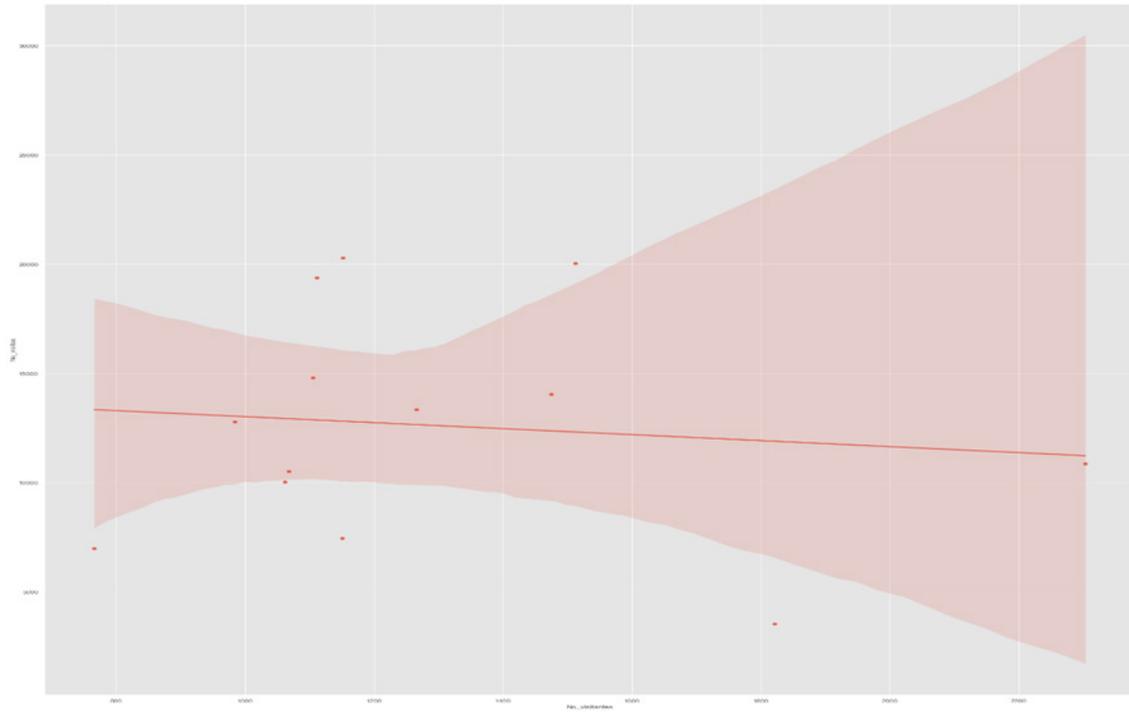


Figura 4.

Análisis de comentarios mediante procesamiento de lenguaje natural

