Universidad de los Andes
Ingeniería de Sistemas y Computación
ISIS 3301 Inteligencia de Negocios

Proyecto 1 – Etapa 2

Machine Learning: Analítica de Textos

Integrantes:

- Juan David Becerra – 201911588
- Juan Andrés Santiago – 201821950
- Nicolás Chalee Guerrero – 201912737

1 Proceso de automatización del proceso de preparación de datos

Descripción del proceso e implementación realizado por el ingeniero de datos	El problema que el negocio busca resolver es determinar la elegibilidad de determinado paciente para pruebas de cáncer a partir de texto descriptivo. Para tal efecto, el ingeniero de datos, en un principio, realizó una transformación del texto que se recibe buscando limpiarlo. En dicha transformación se quitaron caracteres ASCII, se cambiaron mayúsculas a minúsculas, se removieron los números y se quitaron signos de puntuación. Posteriormente, se implementó el modelo que predice la elegibilidad. La decisión sobre cual algoritmo de ML implementar se tomó teniendo en cuenta cual fue el que dio mejor resultado en la entrega previa del proyecto; en este caso se usa Naive Bayes.

Detalles de la actividad de minería de datos:

Tarea	Técnica	Algoritmo e hiperparámetros
Preprocesado de datos	Limpieza, eliminación de stop words	remove_stop_words, NLTK stopwords
Preprocesado de datos	Reducir palabras a su raíz	lemmatize_words, NLTK lemmatizer
Preprocesado de datos	Eliminación de la puntuación	text = "".join([c for c in text if c not in punctuation])
Modelamiento	Predicción	Naive Bayes

2 Descripción del usuario/rol, conexión con el proceso de negocio e importancia para el rol de la existencia de la aplicación.

El usuario potencial de la aplicación es un doctor que desee agilizar el proceso de selección de pacientes elegibles para pruebas de cáncer. Una aplicación del estilo, teniendo en cuenta que el modelo predictivo utilizado tiene una precisión alta, permite al doctor usuario ahorrar una inmensa cantidad de tiempo leyendo y analizado la elegibilidad paciente por paciente, más aún si hay un volumen masivo de texto de pacientes por examinar. Por consiguiente, la aplicación aporta gran valor al proceso de negocio, pues no solo lo agiliza, sino que también arroja resultados precisos y hace eficiente/eficaz el proceso relacionado con la selección de pacientes elegibles para pruebas de cáncer.

A continuación, se presenta la interfaz de la aplicación web que permite al doctor interactuar con el modelo predictivo:

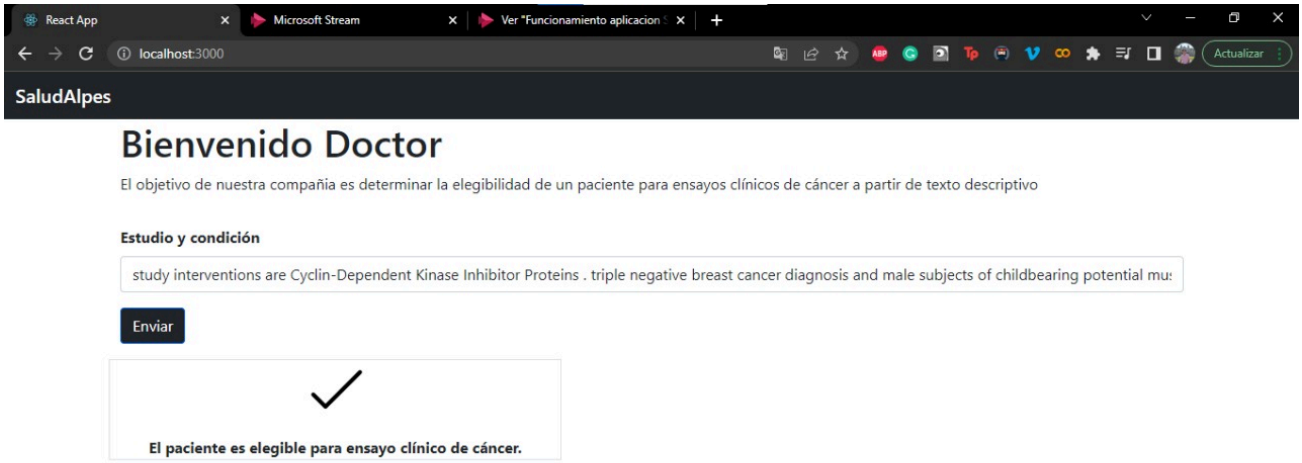

Figura 1: Interfaz con paciente elegible

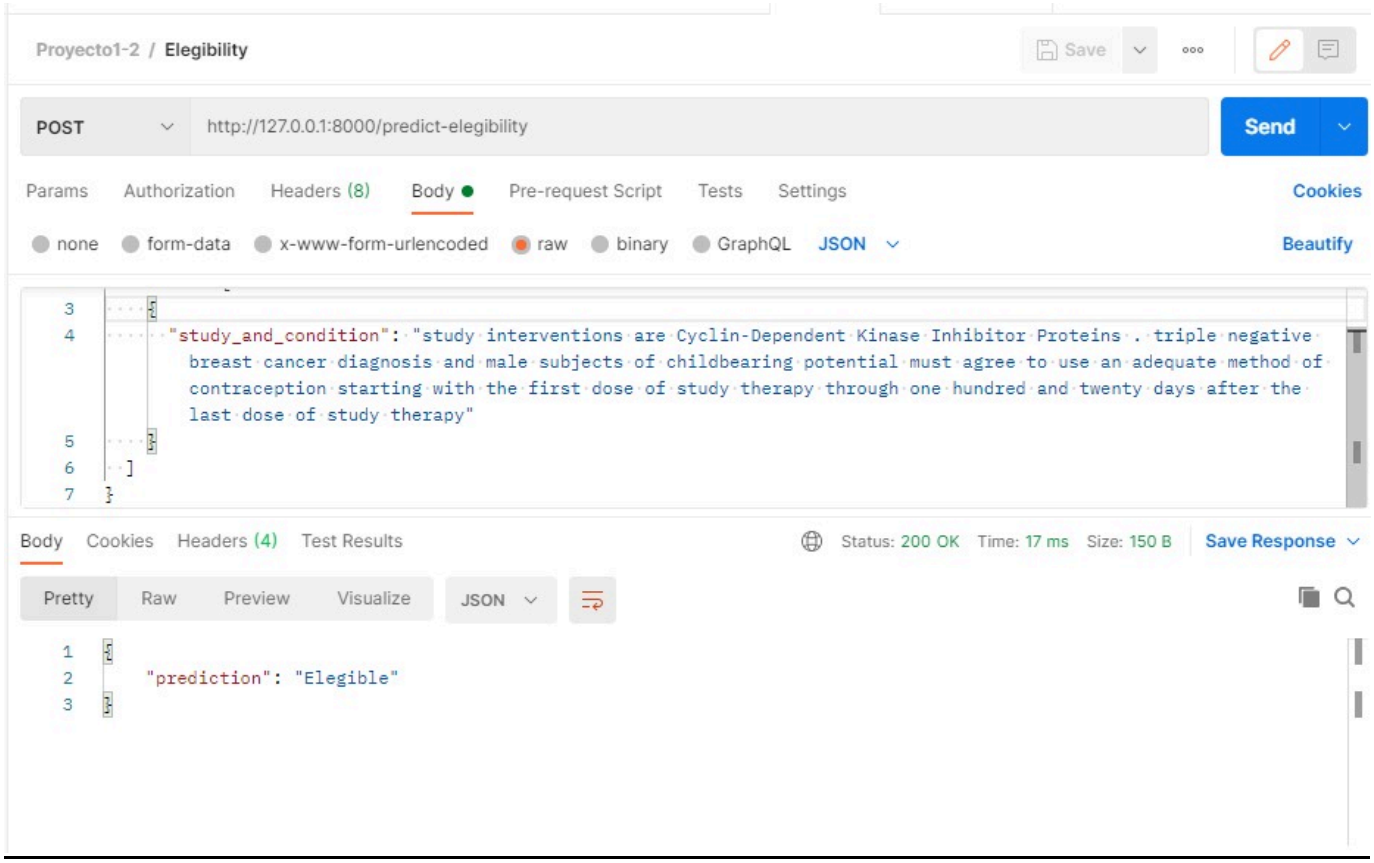

Figura 2: Predicción en Postman

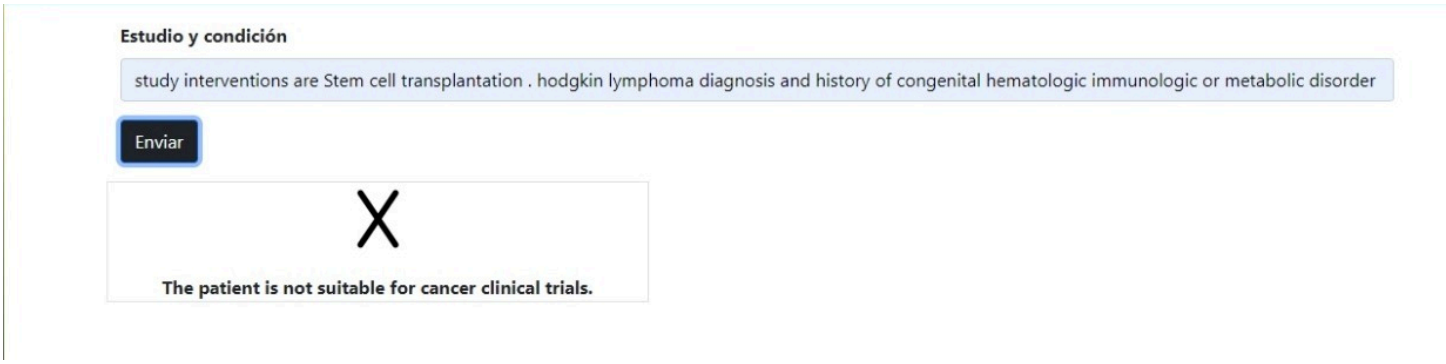

Figura 2: Interfaz con paciente no elegible

3 Resultados

[VIDEO EN EL PADLET]

4 Trabajo en Equipo

- **Líder de proyecto:** Juan David Becerra Romero
- **Ingeniero de datos:** Juan Andrés Santiago
- **Ingeniero de Software responsable de diseño y resultados:** Juan Andrés Santiago
- **Ingeniero de Software responsable de desarrollo aplicación:** Nicolás Guerrero

Los mayores retos enfrentados en el desarrollo del proyecto fueron, en primer lugar, lograr implementar la conexión entre el modelo desarrollado en JupyterLab con un front que permitiera al usuario interactuar con el algoritmo. Asimismo, se nos dificulto la implementación de un preprocesamiento óptimo de los datos debido a todas las consideraciones que se debían hacer para dar una entrada optima al algoritmo. De la misma forma, debido a la naturaleza del problema que se quería resolver alinear lo desarrollado con los objetivos de negocio fue un reto para el equipo.

Cada estudiante estuvo muy pendiente de sus responsabilidades según el rol asignado. Nicolás estuvo muy pendiente del desarrollo del back de la aplicación, trabajando de la mano con Juan Andrés, quien estuvo más pendiente de la conexión con el front y su correcto funcionamiento. Juan David fue la persona encargada de planear las reuniones de equipo y apoyar todas las partes del desarrollo si era necesario.

Repartición de puntos:

- Juan Andrés Santiago: 42,5
- Juan David Becerra: 32,5
- Nicolas Chalee: 25

Esto debido a que Juan Andrés invirtió más tiempo al desarrollo del proyecto en relación a los otros dos integrantes.