**Procesamiento Natural del Lenguaje (NPL)**

Actividad 1

**Contexto**

La concesionaria AutoFlex, donde trabajamos como parte de una célula Data Science, está preocupada por la aceptación de la nueva versión de su marca estrella. Se han recibido muchos comentarios en redes sociales al respecto, y claramente se quiere optimizar el costo de procesamiento de esas opiniones. Para ello, desea interpretar esos comentarios utilizando el preprocesamiento de texto e implementar un modelo para resumir las opiniones de manera automática. El equipo de trabajo propone realizar un análisis de sentimientos que clasifique cada texto escrito, ya que la gerencia comercial tiene un conjunto de comentarios clasificados como “bueno”, “malo” o “info” (necesita más información). Debemos realizar, primero, una descarga de la base, y terminar de sustentar metodológicamente el proyecto.

**Consignas**

1. En primer lugar, debemos sustentar metodológicamente:
2. ¿Qué tipo de aplicación es un análisis de sentimientos? Por favor, en un párrafo de no más 6 líneas, explique en qué consiste, de manera que la gerencia de analítica lo comprenda fácilmente.
3. ¿Qué tipo de procesamiento es necesario realizar primero? Exponga en un dibujo los pasos que va a realizar sobre los comentarios con el fin de convertirlos en data estructurada.
4. La gerencia de *marketing* le ha entregado una tabla con diferentes comentarios de los clientes, la cual se llama “comentarios.csv”. Primeramente, realice una lectura de los datos. Para ello, use el método read\_csv en vez de read\_table; utilice como separador la coma (,). Indique cuántos registros tiene la tabla y cuántas columnas; visualice los 20 primeros registros.
5. Realice un análisis exploratorio de esta data encontrando el porcentaje de tipo de comentarios que han sido clasificados como malos, buenos o información, aplique el código Python que considere necesario y exprese el resultado en una tabla.
6. Diseñe un patrón de expresión regular para utilizarlo como tokenizador más adelante, que además de las palabras en idioma español, lea los emojis como 🥰❤. También considere la expresión como “:)” como un solo token, pero que excluya la puntuación punto “.”, coma “,” y punto y coma “;”.
7. Defina el conjunto X como los comentarios del *data frame*, y el *target*, y como la columna tipo.

**Formato de entrega**

****

Se debe entregar esta parte en formato *notebook* de Jupyter. Escriba el código de cada parte solicitada en una celda “code”, y mantenga las salidas dentro del mismo documento. Sus comentarios y respuestas textuales las puede escribir en una celda tipo “markdown”.

**Cómo cargar la actividad**

**¡Felicitaciones por llegar al final de esta actividad!**

Para concluir tu entrega sigue estos pasos:

1. Haz clic en el botón “Enviar tareas” para comenzar el proceso de carga de tu archivo con las respuestas.

2. Adjunta el archivo que contiene tus respuestas a través de la interfaz proporcionada.

3. Una vez que hayas adjuntado el archivo, confirma tu entrega haciendo clic en el botón correspondiente.

Recuerda completar todos los puntos y tener cuenta las condiciones de entrega.

**¡Éxitos!**