**PROYECTO – ETAPA 2**

**Juan José Córdoba:** 201922105

**Andrés Peña:** 201913766

**Laura Isabela Martínez Galindo:** 202012613

**Temática:** Salud Mental

**Problema de negocio:**

El tabú alrededor del suicidio o los problemas relacionados con la salud mental hace que para las personas con estos problemas sea más difícil hablarlo. Con esto en mente, muchas de estas personas utilizan las redes sociales para expresar estos sentimientos. Reddit, una red social de foros, contiene una gran variedad de comunidades que comentan este tipo de situaciones y en algunos casos sus participantes sufren de depresión o han intentado suicidarse.

**Enfoque analítico:**

Una herramienta muy útil para predecir cuales usuarios tienen depresión o han intentado suicidarse es el Machine Learning. En este caso se van a utilizar técnicas de Analítica de Textos (AT) para poder analizar comentarios de usuarios en Reddit que cumplen con estas características para poder entrenar un modelo que prediga si una persona, dependiendo de sus comentarios en esta red social, padece de alguna de estas enfermedades.

**Roles:**

* *Líder de proyecto:*Juan José Córdoba (4 hrs)

Estuvo encargado de que el proyecto se realizara de la manera más organizada posible, agendó las reuniones del grupo. Si alguno presentaba un error en el código, ayudaba a encontrar el problema y poderlo resolver rápidamente.

* *Ingeniero de datos y de software responsable del diseño de la aplicación y resultados:*Laura Isabela Martínez (4.5hr)

Analizó si era necesario hacer cambios en la preparación y entendimiento de los datos. También, verificó que la construcción del modelo fuera correcta.

* *Ingeniero de software responsable de desarrollar la aplicación final:*Andrés Peña (5.5 hrs)

Implementó la aplicación donde el usuario pueda hacer uso de este, ingresando los textos de reddit y obtener los resultados del modelo construido.

**Repartición de puntos:**

* Juan José Córdoba: 33
* Andrés Peña: 33
* Laura Isabela Martínez: 33

**Reuniones de equipo:**

*Reunión de lanzamiento y planeación:* 10 nov 2022 (0.5 hrs)

Nos reunimos para leer el enunciado de la etapa 2 del proyecto y poder definir los roles y las tareas que iba a realizar cada integrante en el grupo. Con la asignación de los roles nos definimos las tareas que se van a realizar en el proyecto.

*Reunión de seguimiento:* 11 nov 2022 (1 hr)

En esta reunión conversamos acerca de los avances que había hecho cada integrante.

*Reunión de seguimiento 2:* 12 nov 2022 (0.4 hrs)

Fue una reunión para poder verificar que no hubiera ningún error y que todo estuviera corriendo correctamente.

Reunión de finalización 1: 13 nov 2022 (1.5 hrs)

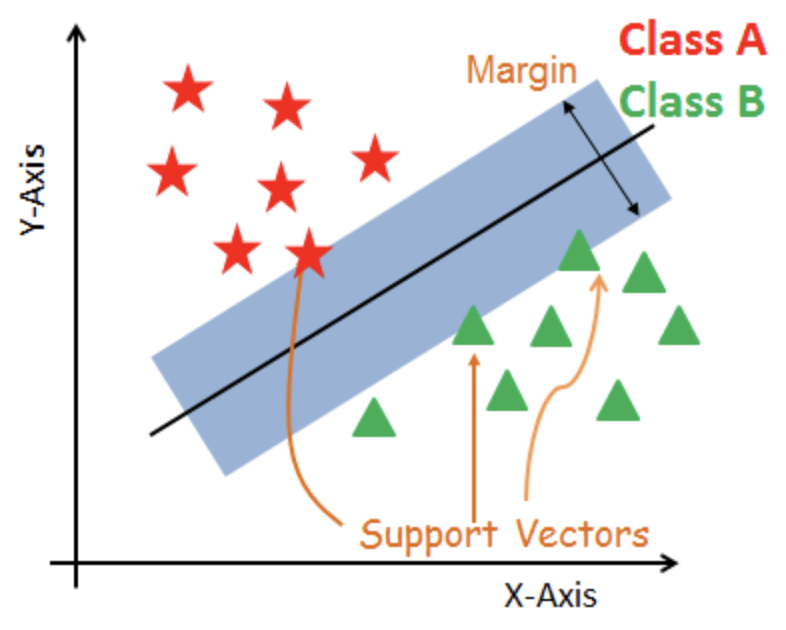
Analizamos los resultados finales de la aplicación y se realizó el video de presentación del proyecto.

**Proceso de automatización del proceso de preparación de datos**

Para el preparación y entendimiento de los datos se trabajó con la misma metodología de la etapa 1. Es decir, no añadimos ni quitamos más columnas, definimos las stopwords, usamos tokenizer para dividir las palabras en tokens y seleccionamos el mismo modelo para vectorizar (TfidfVectorizer).

**Algoritmo a realizar:** Support Vector Machines

El propósito de este algoritmo es construir de manera iterativa, un hiperplano óptimo que separe el dataset en diferentes clases.



**Construcción del modelo**

Para construir el modelo seguimos los siguientes pasos:

1. Importar el módulo de Support Vector Machines
2. Crear el objeto de Support vector llamando a Support Vector Classification que pertenece al módulo de Support Vector Machines. Y como parámetro, un kernel lineal ya que este se usa cuando hay una gran cantidad de features en el data set, en este caso como está trabajando con Text classification, lo más conveniente es usar el kernel lineal.
3. Entrenar el modelo usando los training sets.
4. Crear el pipeline y lo exportamos para poder usarlo en el API.

**Acceso por medio de API**

Para acceder por el API, importamos el pipeline y lo usamos para poder predecir las entradas que lleguen por el API para luego enviar estas entradas al API como un archivo json.

El usuario manda varias entradas para poder re-entrenar el modelo. Consecuentemente, se vuelve a crear un pipeline con el modelo re-entrenado para volver a hacer las predicciones.

**Descripción del usuario/rol:**

El usuario que va a utilizar la aplicación es una persona u organización que necesita apoyo para poder clasificar publicaciones de personas en reddit y poder detecar si son personas con tendencias suicidas o no.

Esta aplicación es importante ya que puede ser una herramienta que ayude a brindar apoyo psicológico a personas suicidas, cabe resaltar que los resultados no son 100% correctos. Sin embargo, si hay errores en la predicción no van a haber consecuencias graves, por ejemplo, si una persona se clasifica como suicida y es no suicida, se le dará ayuda a una persona emocionalmente estable y por ello, no le afectará en nada.

**Descripción de la aplicación**

La aplicación tiene dos entradas, una de ellas manda el texto de reddit para determinar si es suicida o no suicida. La segunda entreda es para volve4r a re-entrenar el modelo.