

# Salud Mental

## PROYECTO 1

-Juan Carlos Marin  
-Kevin Cohen  
-Juan Felipe Castro



# Publicaciones de jóvenes en Reddit

## **SOBRE LA EMPRESA**

La empresa busca apoyar la detección de casos de posible suicidio mediante la información histórica de miles de usuarios de la plataforma Reddit, los cuales hacen publicaciones de diferentes tipos. Esto lo hacen con el fin de poder ayudar a esas personas que se encuentran en una situación crítica antes de que sea tarde, mediante un análisis automático que sea lo más rápido posible.



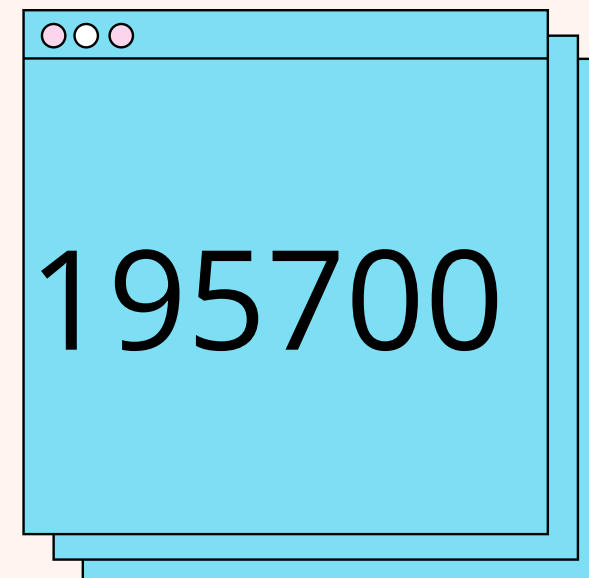
# ● Enfoque Analítico

## -Machine Learning para el modelo:



Lo que se busca es que mediante una base de datos que contiene las publicaciones de 195,700 usuarios y su destino final (suicidio o no) se construya un modelo de clasificación (mediante el procesamiento de lenguaje natural) que sea capaz de interpretar las publicaciones de futuros usuarios y determinar con cierto grado de precisión si esa persona tiene o no tendencias suicidas.

Total de publicaciones:



# Entendimiento y preparación de los datos:

Dentro de los datos podemos ver que son totalmente completos, debido a que no contiene valores nulos, sin embargo, los textos deben ser procesados pues tienen elementos que deben ser tratados, es por esto que la preparación de los datos se realiza el siguiendo estos pasos:

**Transformar la  
columna 'class' a  
una variable  
numérica de 1s y 0s**

**Pasar todos los  
textos a minúsculas**

**Tokenizar los  
textos:**

**Se eliminan los  
artículos, las  
conjunciones y las  
preposiciones**

# Modelado y Evaluación

Tres algoritmos de clasificación utilizados:

## Regresión Logística



hace uso de `CountVectorizer`, lo cual propone una matriz binaria, con la aparición de cada token en cada texto, y a partir de esto ejecutamos el algoritmo de regresión logística para la clasificación de los textos.

## Naive Bayes



Este algoritmo aplica el teorema de Bayes y probabilidad condicional según las palabras tokenizadas y su influencia que tienen en la decisión de clasificación.

## SVM



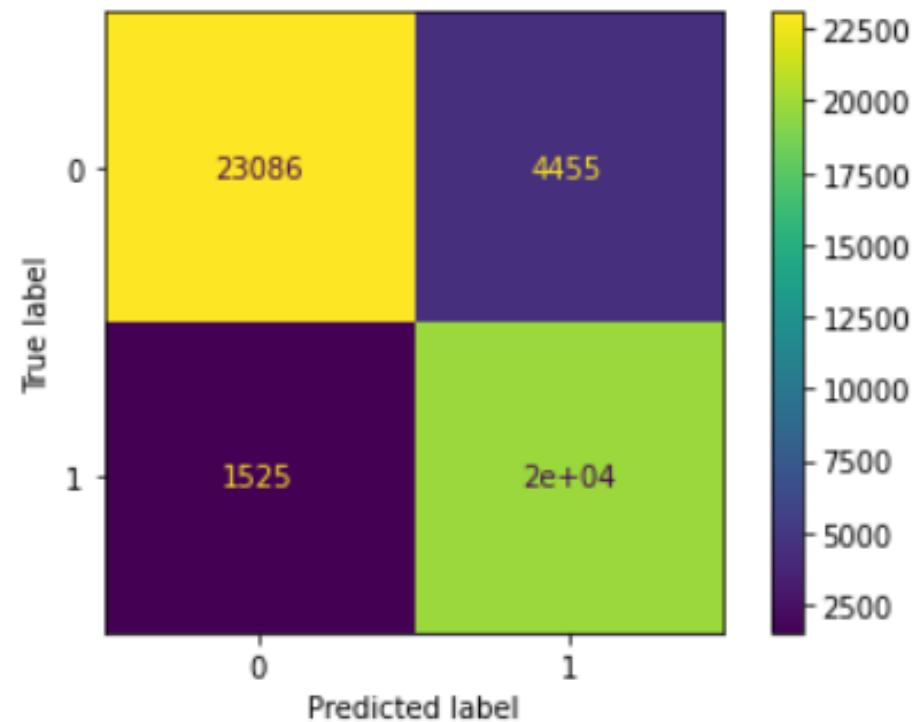
El algoritmo SVM funciona mediante la correlación de los diferentes datos en un espacio de grandes dimensiones, de forma que los puntos de datos se pueden categorizar, incluso si los datos no se pueden separar linealmente de otro modo.



# ANALISIS DE RESULTADOS

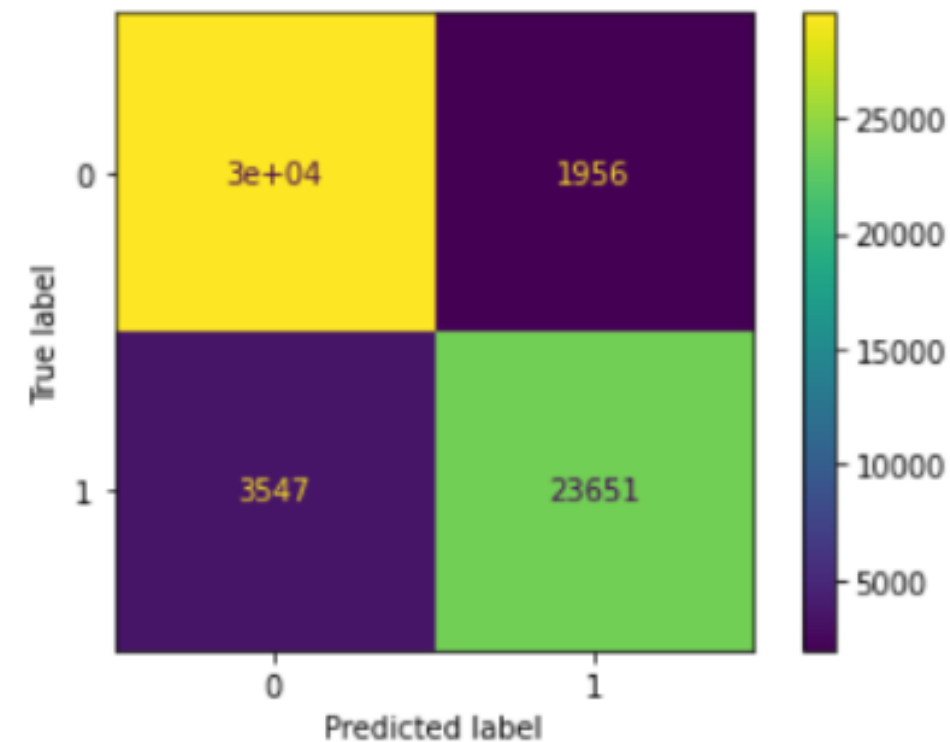
## Regresión Logística

- Accuracy: 87.7%
- Precision 81.68 %
- Recall: 92.86%
- F1 Score: 86.91 %



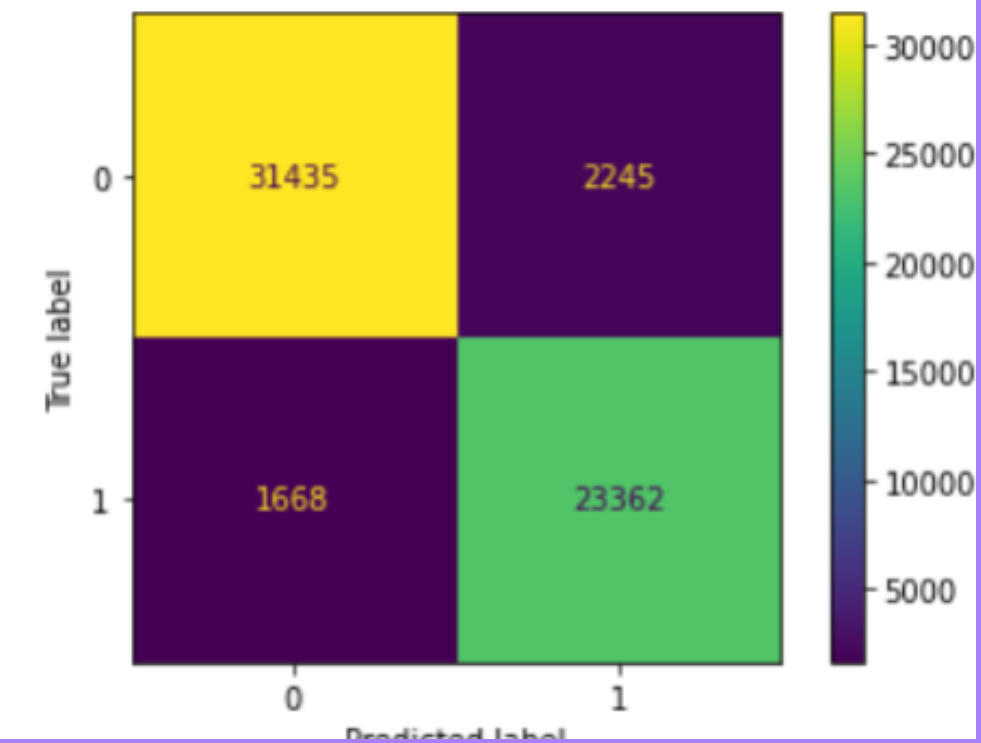
## Naive Bayes

- Accuracy Score: 90.62
- Precision Score: 92.36
- Recall Score: 86.95
- f1 Score: 89.57



## SMV

- Accuracy Score: 93.33
- Precision Score: 91.23
- Recall Score: 93.33
- f1 Score: 92.27



Para la ejecución de los 3 algoritmos se utilizaron los datos previamente preparados.

## Elección del mejor algoritmo:

### SVM

Después del proceso realizado con los distintos algoritmos y metodologías, llegamos a la conclusión de que el mejor para cumplir con los objetivos del negocio es el modelo de SVM con un F1 Score de 92.27 %, el cual es el más alto entre todos los modelos probados. Esto es bastante beneficioso, dado que nos permite predecir con bastante certeza posibles casos de suicidio a partir de publicaciones en la plataforma de reddit, lo cual es el objetivo del proyecto.

# Posibles aplicaciones

## Para empresas

01

CENTROS MEDICOS: Los hospitales podrían utilizar esta información para entender que pasa por la cabeza de sus pacientes y así ayudarlos mas fácilmente.

02

PSICOLOGOS: Estos modelos serian de gran ayuda para estos doctores, ayudando a identificar mas fácilmente posibles casos de suicidio.

03

ESCUELAS: Los colegios podrían beneficiarse de esta información compartiéndola con sus padres, para que estén pendientes de sus redes sociales.

04

REDES SOCIALES: Estas plataformas podrian adaptar estos modelos asus algoritmos, para que dentro de la misma se detecten estas situaciones.



# Conclusiones

Todos nuestros modelos presentaron muy buenos resultados

nuestro modelo SMV está en la capacidad de ayudar a las personas a detectar posibles casos de suicidio

Siempre se necesitará la opinión de un experto

El modelo no reemplaza la labor de los especialistas..

