

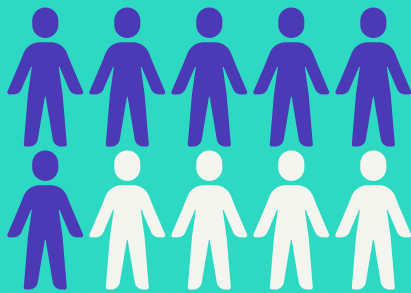
# Predicción del Suicidio

## OBJETIVO

Realizar un análisis de clasificación de texto para predecir ideación suicida, comparando diferentes modelos mediante diferentes métricas para evaluar los resultados.

## INTRODUCCIÓN

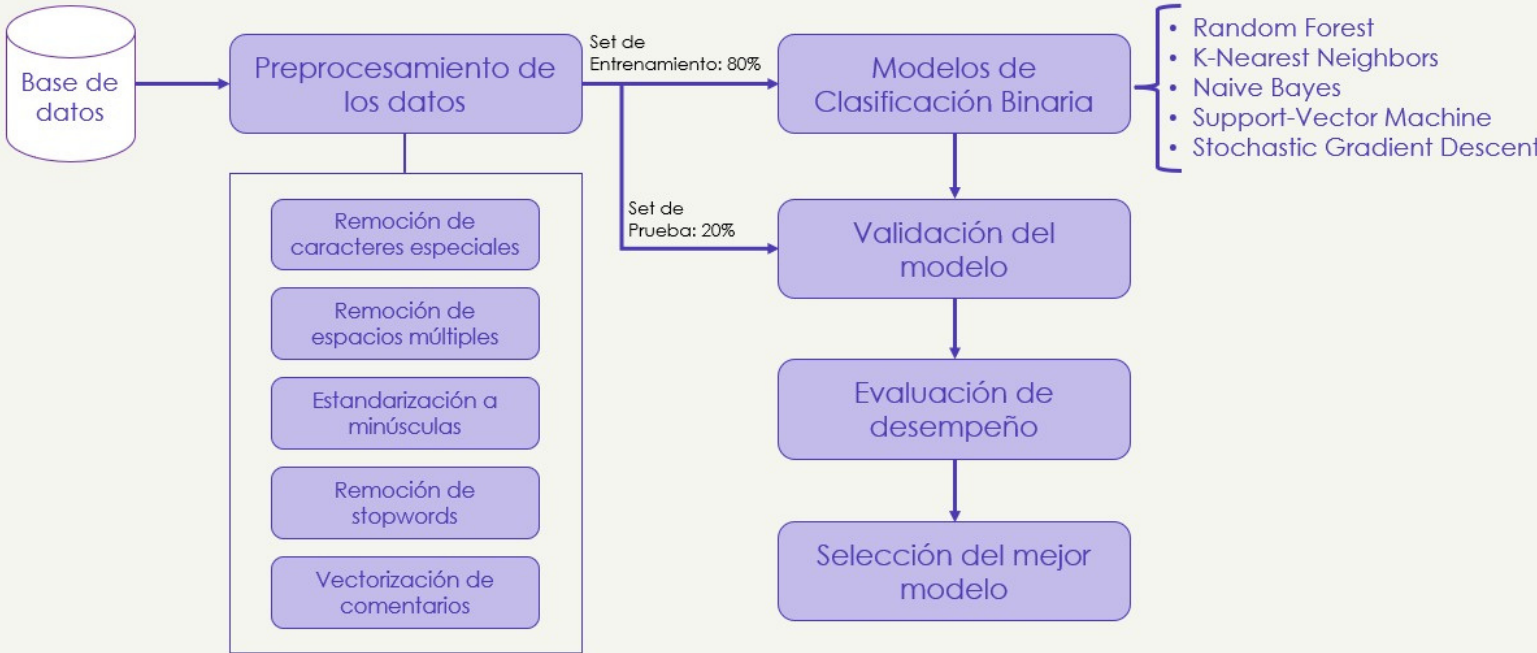
Es la **segunda causa de muerte** en el mundo, entre la población **adolescente** y **adultos jóvenes** de entre 15 a 29 años. Se calcula que cada año **800 mil** personas se quitan la vida y muchas más lo intentan. Entre sus causas existen factores biológicos, psicológicos, sociales y culturales, los cuales al conocerse podrían aumentar las probabilidades de **prevención** de suicidio.



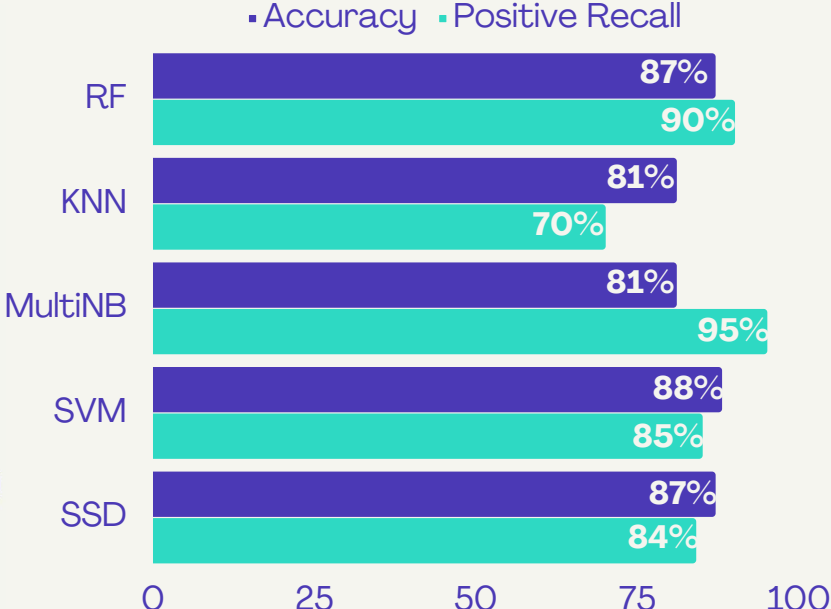
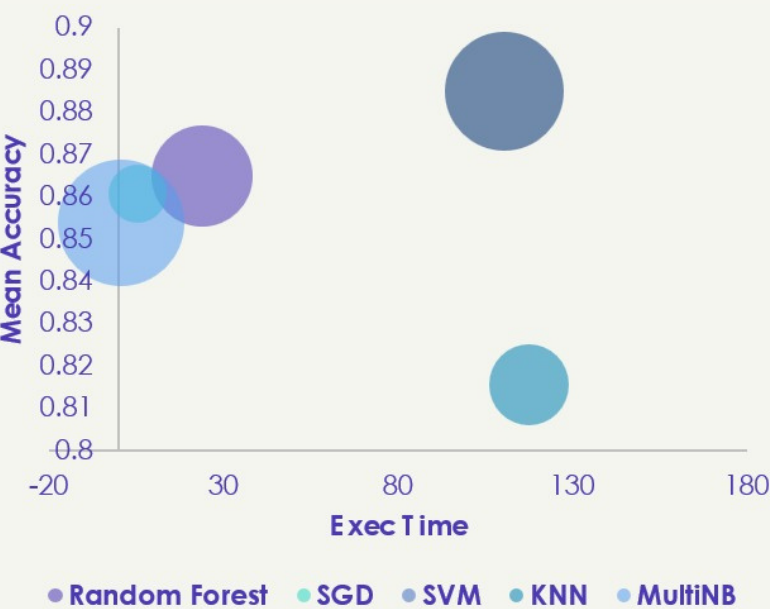
## CONJUNTO DE DATOS

Kaggle - Colección de publicaciones clasificadas en ideación suicida/no suicida de los subreddits "**Suicide Watch**" desde dic 2008 hasta ene 2021 y "**Depression**" desde ene 2009 hasta ene 2021.

## METODOLOGÍA



## RESULTADOS



## CONCLUSIONES

- Los modelos con una exactitud más alta fueron SVM y Random Forest, ambas mayores a 87%.
- Debido al contexto del problema la métrica elegida es la exactitud, esto ya que tanto los falsos positivos como los falsos negativos tienen gran importancia.
- También puede utilizarse la métrica de recall en la categoría positiva ya que esta minimizaría los falsos negativos.
- Tomando en cuenta ambas métricas el modelo con mejor desempeño es el de Random Forest.



### Referencias

[Suicidio, segunda causa de muerte en el mundo | Fundación UNAM](#)

[Tiroteos masivos en Estados Unidos | Wikipedia](#)

[Tiroteos masivos en Estados Unidos muestran un nuevo patrón entre los perpetradores | Ojo Público](#)

[Suicide Watch - Dataset | Kaggle](#)

