**EL SUICIDIO COMO PROBLEMA DE SALUD PÚBLICA**

El suicidio es un problema de salud pública significativo a nivel mundial y Colombia no es la excepción. Según la Organización Mundial de la Salud (OMS), cerca de 800,000 personas se quitan la vida cada año, lo que equivale a una persona cada 40 segundos. En Colombia, las cifras también son preocupantes, con tasas de suicidio que varían según el género, la edad y la región. Este problema no solamente afecta al circulo social de la persona que se quita la vida sino también al sistema de salud que, por falta de acciones preventivas, debe asumir los sobrecostos de la atención.

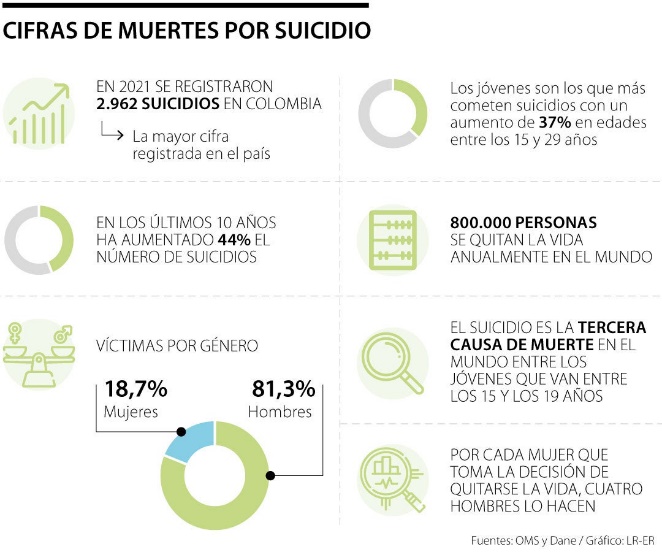
Utilizaremos el conjunto de datos "Suicide Rates Overview 1985 to 2016" de ZDataset, que combina información socioeconómica con tasas de suicidio por año y país. Complementaremos este conjunto de datos con información específica de Colombia obtenida de fuentes oficiales como el DANE (Departamento Administrativo Nacional de Estadística) y el Ministerio de Salud para darle contexto al presente trabajo.

Datos globales:



Fuente: <https://fundacionelementos.org/2019/10/09/prevencion-del-suicidio-el-foco-del-dia-mundial-de-la-salud-mental-2019/>

Datos de Colombia:



Fuente: <https://www.larepublica.co/economia/en-los-ultimos-10-anos-el-numero-de-casos-de-suicidios-ha-aumentado-44-en-el-pais-3448483>

**PROBLEMA DE NEGOCIO**: ¿Cómo podemos predecir y comprender los factores que influyen en las tasas de suicidio en Colombia para implementar estrategias de prevención más efectivas?

**PROBLEMA ANALITICO**

Desarrollar un modelo de aprendizaje automático que prediga las tasas de suicidio en Colombia con base en indicadores socioeconómicos y demográficos.

Además, se busca identificar los factores que tienen mayor impacto en las tasas de suicidio para orientar las intervenciones de salud pública.

**Diseño de Solución Propuesto**

**Preparación de Datos:**

**Limpieza y transformación de datos:** Manejar valores faltantes, codificar variables categóricas y escalar variables numéricas. Usaré Label Enconding y One-Hot Enconding. Usaré técnicas de escalado como StandarScaler y probaré una nueva: RobustScaler vista en otra materia.

**Análisis exploratorio de datos**: Visualizar y analizar las distribuciones, relaciones y patrones en los datos. Usaré gráficos de dispersión para ver el comportamiento de los datos, matrices de correlación y probaré la T-Student para aplicar lo aprendido en el curso de la profe Lorena Blum.

**Modelado**:

**Selección de variables:** Identificar las variables más relevantes para la predicción de las tasas de suicidio. Usaré arboles de decisión y bosques aleatorios.

**Selección de algoritmos**: Evaluar y comparar diferentes algoritmos de aprendizaje automático, como regresión lineal, SVM, árboles de decisión o redes neuronales. También probaré Gradient Boosting para saber cual ajusta mejor los datos y proporciona una buena generalización.

**Entrenamiento y evaluación del modelo**: Entrenar el modelo seleccionado, optimizar sus hiperparámetros y evaluar su rendimiento utilizando métricas apropiadas. Probaré con GridSearch. Para precisión RMSE o R2.

**Análisis e Interpretación:**

**Identificar los factores de riesgo**: Analizar la importancia de las variables en el modelo para comprender qué factores tienen mayor impacto en las tasas de suicidio. Cómo afecta cada variable en las tasas de suicidio? Es positiva o es negativa?. Se usarán pruebas de hipótesis.

**Visualización de resultados**: Presentar los hallazgos de manera clara y comprensible para las partes interesadas. Trataré de usar diagramas de calor, boxplots, diagramas de dispersión. Quiero saber cuáles son los factores de riesgo más relevantes. La edad? El género? El estrato? Quiero proponer como hacer programas para la prevención a un público específico, hacer segmentación para ser más asertivos

**Implementación y Monitoreo:**

**Desarrollar una herramienta de predicción:** Implementar el modelo en un sistema que permita predecir las tasas de suicidio en diferentes escenarios.

**Monitorear el rendimiento del modelo:** Evaluar la efectividad del modelo a lo largo del tiempo y realizar ajustes según sea necesario. Quiero establecer métricas de rendimiento como R2 o RMSE

Me gustaría que se pudieran recopilar sugerencias de los usuarios del sistema de predicción (Personal de salud o familiares) para poderlo hacer más asertivo.

**Beneficios Esperados**

Mejorar la comprensión de los factores de riesgo del suicidio en Colombia.

Predecir las tasas de suicidio con mayor precisión.

Identificar grupos de población con mayor riesgo.

Diseñar intervenciones de salud pública más efectivas para la prevención del suicidio: Quizás esta es para mi la más importante, ya que he sido afectada como familiar. Como una situación personal y como profesional de la salud me gustaría que se diseñaran programas de prevención efectivas.