

Entrega 1



LAURA CECILIA TOBÓN OSPINA

JUAN PABLO ARENAS VÉLEZ

JUAN DAVID ARISMENDY PULGARÍN

Tutor: RAUL RAMOS POLLAN

Introducción Inteligencia Artificial 2022-1

Universidad de Antioquia

2022

1. Planteamiento del problema

Conocer las causas de muerte de la humanidad es de vital importancia, ya que esto nos puede permitir estudiar cuales son las principales afecciones que se pueden presentar y por qué no, prevenir por medio de tratamientos o prácticas sanas. Pero para llegar a ello se necesitan recursos de investigación, financiación de proyectos de salud preventiva y predictiva. Para ello se pretende analizar cuál causa de muerte puede ser potencialmente más peligrosa, y con ello desarrollar un modelo que permita mejorar las estimaciones para así generar interés en dichos desarrollos e investigaciones que permitan ejecutar acciones que mejoren la calidad de vida de las personas y mitiguen esas estadísticas de muerte.

2. Dataset

El dataset a utilizar proviene de una competencia de kaggle en la cual se proporcionan datos del número anual de muertes por diferentes causas alrededor del mundo. El dataset está compuesto por un conjunto de archivos .csv que proporcionan la información requerida, El archivo que contiene los datos del edificio es nombrado como:

annual-number-of-deaths-by-cause.csv y contiene la siguiente información principal:

<https://www.kaggle.com/datasets/programmerrdai/cancer?select=annual-number-of-deaths-by-cause.csv>

Entity: País,

Code,Código del país

Year, Año,

Number of executions (Amnesty International), Número de ejecuciones (Amnistía Internacional),

Deaths - Meningitis - Sex: Both - Age: All Ages (Number), Muertes - Meningitis - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Alzheimer's disease and other dementias - Sex: Both - Age: All Ages (Number), Muertes - Enfermedad de Alzheimer y otras demencias - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Parkinson's disease - Sex: Both - Age: All Ages (Number), Muertes - Enfermedad de Parkinson - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Nutritional deficiencies - Sex: Both - Age: All Ages (Number), Muertes - Deficiencias nutricionales - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Malaria - Sex: Both - Age: All Ages (Number), Muertes - Malaria - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Drowning - Sex: Both - Age: All Ages (Number), Muertes - Ahogamiento - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Interpersonal violence - Sex: Both - Age: All Ages (Number), Muertes - Violencia interpersonal - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Maternal disorders - Sex: Both - Age: All Ages (Number), Muertes - Trastornos maternos - Sexo: Ambos - Edad: Todas las edades ,

Deaths - HIV/AIDS - Sex: Both - Age: All Ages (Number), Muertes - VIH/SIDA - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Drug use disorders - Sex: Both - Age: All Ages (Number), Muertes - Trastornos por consumo de drogas - Sexo: Ambos - Edad: Todas las edades

Deaths - Tuberculosis - Sex: Both - Age: All Ages (Number), Muertes - Tuberculosis - Sexo: Ambos - Edad: Todas las edades),

Deaths - Cardiovascular diseases - Sex: Both - Age: All Ages (Number), Muertes - Enfermedades cardiovasculares - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Lower respiratory infections - Sex: Both - Age: All Ages (Number), Muertes - Infecciones de las vías respiratorias bajas - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Neonatal disorders - Sex: Both - Age: All Ages (Number), Muertes - Trastornos neonatales - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Alcohol use disorders - Sex: Both - Age: All Ages (Number), Muertes - Trastorno por consumo de alcohol - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Self-harm - Sex: Both - Age: All Ages (Number), Muertes - Autolesiones - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Exposure to forces of nature - Sex: Both - Age: All Ages (Number), Muertes - Exposición a las fuerzas de la naturaleza - Sexo: Ambos - Edad: Todas las edades),

Deaths - Diarrheal diseases - Sex: Both - Age: All Ages (Number), Muertes - Enfermedades diarreicas - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Environmental heat and cold exposure - Sex: Both - Age: All Ages (Number), Muertes - Exposición al calor y al frío ambiental - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Neoplasms - Sex: Both - Age: All Ages (Number), Muertes - Neoplasias - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Conflict and terrorism - Sex: Both - Age: All Ages (Number), Muertes - Conflicto y terrorismo - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Diabetes mellitus - Sex: Both - Age: All Ages (Number), Muertes - Diabetes mellitus - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Chronic kidney disease - Sex: Both - Age: All Ages (Number), Muertes - Enfermedad renal crónica - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Poisonings - Sex: Both - Age: All Ages (Number), Muertes - Intoxicaciones - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Protein-energy malnutrition - Sex: Both - Age: All Ages (Number), Terrorism (deaths), Muertes - Desnutrición proteico-energética - Sexo: Ambos - Edad: Todas las edades , Terrorismo (muertes),

Deaths - Road injuries - Sex: Both - Age: All Ages (Number), Muertes - Lesiones viales - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Chronic respiratory diseases - Sex: Both - Age: All Ages (Number), Muertes - Enfermedades respiratorias crónicas - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Cirrhosis and other chronic liver diseases - Sex: Both - Age: All Ages (Number), Muertes - Cirrosis y otras enfermedades crónicas del hígado - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Digestive diseases - Sex: Both - Age: All Ages (Number), Muertes - Enfermedades digestivas - Sexo: Ambos - Edad: Todas las edades ,

Deaths - Fire, heat, and hot substances - Sex: Both - Age: All Ages (Number), Muertes - Fuego, calor y sustancias calientes - Sexo: Ambos - Edad: Todas las edades

Deaths - Acute hepatitis - Sex: Both - Age: All Ages (Number), Muertes - Hepatitis aguda - Sexo: Ambos - Edad: Todas las edades

3. Métricas

La métrica con la cual evaluaremos principalmente a nuestro modelo será el ECM (Error cuadrático medio), el cual calcula el valor medio de la diferencia al cuadrado entre el valor real y el predicho para todos los puntos de datos

$$MSE = \frac{1}{n} \sum_{t=1}^n e_t^2, \text{ where } e_t = \text{original}_t - \text{predict}_t$$

en cuanto a la métrica de negocio queremos determinar si las predicciones son lo suficientemente confiables para conocer las causas de muerte más recurrentes en los diferentes países y ver si la inversión en la prevención de estas causas vale la pena, y no solo sean causas que no puedan remediarse, se busca llegar al análisis de esta viabilidad, si vale la pena o no invertir en la prevención.

4. Desempeño

El criterio más importante sería predecir que una de las más altas probabilidades de muerte esté relacionada con causas que puedan prevenirse, para garantizar la inversión en un proyecto de prevención. El ideal es que las causas de muerte prevenibles estén por encima del 70%.

5. Bibliografía

| Kaggle. (2021). Retrieved 16 December 2021, from <https://www.kaggle.com/datasets/programmerdai/cancer?select=annual-number-of-deaths-by-cause.csv>

https://fayrix.com/machine-learning-metrics_es