

ENTREGA 1: PROYECTO DE ANALÍTICA DE DATOS
PREDICT CO2 EMISSIONS IN RWANDA.

INTEGRANTES:

LAURA CRISTINA DIAZ OSORIO.

C.C: 1018351214

JUAN FELIPE ESCOBAR RENDÓN.

C.C:1001416321

CURSO:

INTRODUCCION A LA INTELIGENCIA ARTIFICIAL

TUTOR:

RAUL RAMOS POLLAN



UNIVERSIDAD DE ANTIOQUIA

FACULTAD DE INGENIERÍA

AMALFI-ANTIOQUIA

2023

PREDICT CO2 EMISSIONS IN RWANDA.

1. Problema Predictivo:

El objetivo de este desafío es utilizar datos de emisiones de CO2 de observaciones satelitales de Sentinel-5P para crear modelos de aprendizaje automático que puedan predecir las emisiones futuras de carbono.

CO2: “El CO2, dióxido de carbono o, simplificando mucho, ‘carbono’, es el principal gas de efecto invernadero de origen humano. Significa que contribuye al calentamiento global, cuyas consecuencias notamos a diario. Es incoloro y carece de olor. Está presente de forma natural en la atmósfera y, sí, forma parte esencial de nuestro organismo.”

2. Dataset Utilizado:

El conjunto de datos utilizado en este proyecto proviene de kaggle competitions y se compone de observaciones satelitales de CO2 de Sentinel-5P. para la recolección de la información se seleccionaron aproximadamente 497 ubicaciones únicas de varias áreas en Rwanda, distribuidas alrededor de tierras de cultivo, ciudades y plantas de energía.

El dataset es un archivo en formato zip el cual está compuesto por los siguientes archivos csv:

- traind.csv: En este archivo se encuentra el porcentaje de datos que fueron seleccionados para entrenar el modelo.
- test.csv: En este archivo se encuentra el porcentaje que será usado para observar la precisión del modelo a la hora de predecir los niveles de emisión de CO2.
- sample_submission: En este archivo muestra el formato correcto en el que se deben presentar las predicciones del modelo.

Los datos contienen siete características principales que se extrajeron semanalmente desde enero de 2019 hasta noviembre de 2022. Estas características incluyen:

- Dióxido de Azufre (Sulphur Dioxide)
- Monóxido de Carbono (Carbon Monoxide)
- Dióxido de Nitrógeno (Nitrogen Dioxide)
- Formaldehído (Formaldehyde)
- Índice de Aerosol UV (UV Aerosol Index)
- Ozono (Ozone)
- Nubosidad (Cloud).

Algunas de las variables del dataset son: ubicación geoespacial (latitud y longitud), año y semana, junto con mediciones de emisiones de CO₂. El conjunto de datos contiene información histórica de emisiones de CO₂ hasta el año 2021:

- latitude y longitude: Representan las coordenadas geográficas de las ubicaciones para las cuales se han recopilado los datos. La latitud representa la posición norte-sur y la longitud la posición este-oeste en la superficie terrestre.
- year: Indica el año al que corresponden los datos.
- week_no: Esta columna indica el número de semana del año al que corresponden los datos.
- SulphurDioxide_SO2_column_number_density: Esta característica representa la densidad de columnas de dióxido de azufre (SO₂) en la atmósfera.
- CarbonMonoxide_CO_column_number_density: Densidad de columnas de monóxido de carbono (CO) en la atmósfera. proporciona información sobre la concentración de CO en la atmósfera en una ubicación y momento dados.

- NitrogenDioxide_NO2_column_number_density: densidad de columnas de dióxido de nitrógeno (NO2) en la atmósfera. Refleja la concentración de NO2 en la atmósfera en un lugar y tiempo específicos.
- Formaldehyde_tropospheric_HCHO_column_number_density: Densidad de columnas de formaldehído (HCHO) en la troposfera. El formaldehído es una sustancia química que puede tener implicaciones en la calidad del aire y la salud humana.
- UvAerosolIndex_absorbing_aerosol_index: Índice de aerosol absorbente en el ultravioleta (UV). Los aerosoles pueden afectar la calidad del aire y la visibilidad, y el índice de aerosol absorbente proporciona información sobre las partículas absorbentes en la atmósfera.
- Ozone_O3_column_number_density: la densidad de las columnas de ozono (O3) en la atmósfera. El ozono en la atmósfera desempeña un papel importante en la protección contra la radiación ultravioleta del sol.
- Cloud_cloud_fraction: Fracción de nubes en la atmósfera. Puede indicar el grado de cobertura de nubes en una ubicación y momento específicos
- emission: indica los niveles emitidos de CO2 correspondientes a cada semana en las que se tomaron los datos.

3. Métricas de Desempeño:

Root Mean Squared Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

Esta métrica mide la raíz cuadrada del error cuadrático medio entre las emisiones de CO2 predichas y los valores reales. Cuanto menor sea el valor de RMSE, mejor será el rendimiento del modelo en términos de precisión de las predicciones, la intención es predecir las emisiones de CO2 para el año 2022 a través de las observaciones satelitales.

4. Criterio de Desempeño Deseable:

Un criterio deseable de desempeño es que el modelo de predicción de emisiones de CO₂ arroje un RMSE lo más pequeño posible, con la finalidad de que las predicciones sean precisas y esto ayude a los gobiernos con la toma de decisiones respecto al cambio climático y las futuras acciones a tomar para tratar de revertir o tratar que el proceso del cambio climático sea más lento.

Bibliografía.

- Newtral, CO2: el aire que exhalamos y que está matando el planeta.

(<https://www.newtral.es/que-es-co2-peligros/20190725/>)

- Kaggle, Playground Series - Season 3, Episode 20, Predict CO2 Emissions in Rwanda.

(<https://www.kaggle.com/competitions/playground-series-s3e20/overview>)

- Evaluando el error en los modelos de regresión, 2018.

(<https://aprendeia.com/evaluando-el-error-en-los-modelos-de-regresion/>)