

## **Primera entrega - Introducción a la Inteligencia Artificial: Predecir las emisiones de CO2 en Ruanda**

Integrantes.

Anderson Estiven Villa Sierra <estiven.villa@udea.edu.co>

Sebastián Ochoa González <sebastian.ochoa1@udea.edu.co>

Simón Alfredo Narváez Restrepo <simon.narvaez@udea.edu.co>

### **1. Descripción del problema predictivo a resolver**

El problema que se pretende resolver es la predicción de la emisión de CO2 o emisiones de carbono, y se pretende abordar creando un modelo de aprendizaje a partir de datos de fuentes abiertas sobre emisiones de CO2, el cual busca determinar si se producirán emisiones de CO2 en Ruanda o no, esta clasificación se realiza usando información relacionada con las ubicaciones en múltiples zonas de Ruanda, y su distribución con respecto a tierras de cultivo y centrales eléctricas.

### **2. Descripción del dataset seleccionado**

Para el desarrollo del ejercicio previamente mencionado, se hará uso del siguiente [dataset Predict CO2 Emissions](#) tomando de Kaggle, de la sesión de Competitions y está compuesto por 497 ubicaciones únicas de múltiples zonas de Ruanda y como estas ubicaciones se encuentran distribuidas con respecto a zonas de cultivos, ciudades y plantas eléctricas. Estos datos hacen referencia al rango años de 2019 a 2021, y lo que se busca es predecir las emisiones de CO2 para el siguiente año, o sea, noviembre. Este dataset contiene un total de 79024 filas y un total de 76 columnas, que nos cuenten distintas características con respecto a nuestro conjunto de datos, como lo son Densidad de la columna de dióxido de azufre, índice de aerosoles absorbentes para el NO2 y demás características, la que nos encargaremos de predecir será la última de las 76 columnas que se llama “emisión”, que indica si se generó emisión de CO2 o no, en la ubicación. El data set cuenta con los siguientes datos faltantes

```
[17] k = d.isna().sum()
      k[k!=0]
```

SulphurDioxide_SO2_column_number_density	14609
SulphurDioxide_SO2_column_number_density_amf	14609
SulphurDioxide_SO2_slant_column_number_density	14609
SulphurDioxide_cloud_fraction	14609
SulphurDioxide_sensor_azimuth_angle	14609
...	
Cloud_surface_albedo	484
Cloud_sensor_azimuth_angle	484
Cloud_sensor_zenith_angle	484
Cloud_solar_azimuth_angle	484
Cloud_solar_zenith_angle	484
Length: 70, dtype: int64	

```
d.isnull().sum()
```

ID_LAT_LON_YEAR_WEEK	0
latitude	0
longitude	0
year	0
week_no	0
...	
Cloud_sensor_azimuth_angle	484
Cloud_sensor_zenith_angle	484
Cloud_solar_azimuth_angle	484
Cloud_solar_zenith_angle	484
emission	0
Length: 76, dtype: int64	

### 3. Las métricas de desempeño requeridas

Para este modelo predictivo, la métrica seleccionada para medir el desempeño del mismo, será RMSE (Root mean squared error) que es la propuesta por la competencia en la cual se encuentra nuestro dataset, que básicamente es la raíz cuadrada del MSE , que nos permitirá medir la magnitud de los errores. Por lo mismo, la MSE también hace parte de las métricas requeridas ya que la usaremos para obtener la RMSE.

### 4. Primer criterio sobre cuál sería el desempeño deseable en producción

Los criterios deseables para el desempeño con las métricas propuestas, son valores cercanos a 0, sabemos que no es posible alcanzar un valor = 0 en lo práctico, ya que siempre existe presencia en el ruido y por la complejidad de los problemas, por eso buscamos cifras que sean lo más cercanas posibles, ya que nos sugieren que las predicciones del modelo, son muy similares a los valores reales.