

# **Primera entrega de proyecto**

**POR:**

Daniel David Sierra Moreno

**MATERIA:**

Introducción a la inteligencia artificial

**PROFESOR:**

Raul Ramos Pollan



UNIVERSIDAD DE ANTIOQUIA

FACULTAD DE INGENIERÍA

MEDELLÍN 2021

## 1. Planteamiento del problema

Es importante cuantificar el gasto energético que implican los edificios debido al uso de agua caliente, vapor, electricidad, etc., ya es necesario implementar mejoras que permitan lograr mayores eficiencias energéticas para reducir el impacto ambiental y los costes. Es por esto por lo que se desea desarrollar un modelo que permitan cuantificar el uso de la energía en edificios, ya que, al tener mejores estimaciones, se podrá generar un mayor interés por parte de inversores e instituciones financieras para la realización de inversiones que impliquen el ahorro energético.

## 2. Dataset

El dataset a utilizar proviene de una competencia de kaggle en la cual se proporcionan datos de tres años de lecturas en contadores de más de mil edificios en diferentes lugares del mundo. El dataset este compuesto por un conjunto de archivos .csv que proporcionan la información requerida, tal como los datos meteorológicos del lugar y los datos del edificio mismo.

El archivo que contiene los datos del edificio es nombrado como *building\_meta* y contiene la siguiente información:

- **site\_id** – Identificador para los archivos meteorológicos
- **building\_id** – Identificador para el archivo *training.csv*
- **primary\_use** – Identificador de la categoría principal de actividades para el edificio, basado en la clasificación de *EnergyStar property type definitions*
- **square\_feet** – Área neta del edificio
- **year\_built** – Año de apertura del edificio
- **floor\_count** – Número de pisos del edificio

En cuanto al archivo con los datos meteorológicos, este se divide en dos, un archivo que será para entrenamiento de algoritmos y otro para pruebas, llamados *weather\_train* y *weather\_test*, respectivamente. La información que contienen ambos archivos es la siguiente:

- **site\_id**
- **air\_temperature** – Temperatura del aire en grados Celsius
- **cloud\_coverage** – Porción del cielo cubierto por nubes en oktas
- **dew\_temperature** – Temperatura de rocío en grados Celsius
- **precip\_depth\_1\_hr** - Milímetros

- ***sea\_level\_pressure*** - Millibar/hectopascales
- ***wind\_direction*** – Dirección de la brújula (0-360)
- ***wind\_speed*** – Metros por segundo

También hay un archivo llamado *train.csv* el cual tiene la siguiente información:

- ***building\_id*** – Identificación para el archivo *building\_meta*.
- ***meter*** – Identificador del contador. Leído como: {0: electricidad, 1: agua fría, 2: vapor, 3: agua caliente}. No todos los edificios tienen todos los tipos de contador.
- ***timestamp*** – Cuándo se hizo la medición
- ***meter\_reading*** – Variable objetivo. Consumo de energía en kWh (o equivalente).

Por último, se tiene un archivo llamado *test.csv* que sirve para organizar correctamente las predicciones. Este archivo tiene la siguiente información:

- ***row\_id*** – Identificador de fila en el archivo de entrega
- ***building\_id*** – Código de identificación del edificio
- ***meter*** – Código de identificación del medidor
- ***timestamp*** – Pasos temporales de los datos de prueba

### 3. Métricas

La métrica de evaluación principal para el modelo será el error logarítmico medio cuadrático (RMSLE), el cual se calcula mediante la siguiente expresión:

$$\epsilon = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(p_i + 1) - \log(a_i + 1))^2}$$

Donde  $\epsilon$  es el valor del RMSLE,  $n$  es el número total de observaciones en el dataset,  $p_i$  es la predicción de la variable objetivo y  $a_i$  es el valor real de la variable objetivo

Por otra parte, en cuanto a la métrica de negocio, se tiene interés en que las predicciones sean lo suficientemente confiables para saber el gasto energético de los edificios y así cuantificar los costos relacionados. Con esta

información se pueden realizar análisis financieros para determinar que tan viables pueden ser los proyectos relacionados con el ahorro energético.

## **4. Desempeño**

Lo que se esperaría de un modelo de este tipo es obtener la predicción del gasto energético relacionado con los procesos de calentamiento y enfriamiento de agua, así como el uso de vapor y energía en edificios. Con esta información se desearía obtener mejores análisis del gasto energético relacionado con todos los procesos anteriores, y con esta información, determinar que tanto se pueden reducir los costos mediante la implementación de proyectos de ahorro energético, así como analizar los proyectos ya existentes.

## **5. Bibliografía**

- ASHRAE - Great Energy Predictor III | Kaggle. (2021). Retrieved 16 December 2021, from <https://www.kaggle.com/c/ashrae-energy-prediction/overview/description>