**US NATIONAL FLIGHTS DELAY**

**JANUARY-JUNE- 2024**

This is a repository created for the ML as an individual delivery of the bootcamp in datascience.

**1.INTRODUCCION**

One of the biggest frustrations these days is flight delays. In the following dataset we have information on US domestic flights between January-June 2024 with variables such as: when the flights depart, the scheduled and actual arrival time, how long they are delayed, reasons for the delay, etc.

**2.OBJECTIVE:**

In this ML project we will try to predict the current arrival time of flights, and with this we will be able to know the delay of domestic flights

**3.METHODOLOGY:**

This project deals with a regression problem, since the objective is to predict a numerical value: the arrival time of flights.

**4.METRIC:**

We will focus on the Arrival Time

**5.BUSINESS IMPACT:**

* Help to improve flight- air connections that present a long delays
* Find inefficiencies between flights and improve customer satisfaction
* Improve airport efficiency.
* Reduce operating costs for airlines.

**6.ML CONTENT -NOTEBOOK**

The dataset for this ML is comprehensive, with 3,461,319 entries and 8 columns, detailing national flights within the US for the period from January to June 2024. The Raw data has been taken from Bureau of Transportation Statistics from US Department of Transportation.

Step 1 | Overview

- Step 1.1 | Data

- Step 1.2 | Variables

Step 2 | Dataset

- Step 2.1 | Libraries

- Step 2.2 | Dataset Information

- Step 2.2.1 | Dataset Information | Treatment of Nulls

- Step 2.2.2 | Dataset Information | Treatment of Variables

Step 3 | Machine Learning

- Step 3.1 | Split & Test

- Step 3.2.1 | Split & Test | Split & Test

- Step 3.2.2 | Split & Test | Target

- Step 3.2 | Mini EDA

- Step 3.2.1 | Mini EDA | Division of categorical and numerical variables

- Step 3.2.2 | Mini EDA | Categorical columns

- Step 3.2.3 | Mini EDA | Numerical Columns

\*Step 3.2.3.1 | Mini EDA | Numerical Columns | Treatment of Numerical

- Step 3.3 | Scaling

- Step 3.4 | Modeling

- Step 3.4.1 | Modeling | Random Forest

- Step 3.4.2 | Modeling | XGBoost

- Step 3.4.3 | Modeling | LightGBM

- Step 3.4.4 | Modeling | Best Model

- Step 3.5 | Graphs

- Step 3.5.1 | Graphs | Real vs Prediction

- Step 3.5.2 | Graphs | Distribution of residuals

Step 4 | Data Analysis

- Step 4.1 | Data Analysis

- Step 4.2 | Conclusions