Acoustic Extinguisher Fire Dataset

Flame Extinction Status Prediction

Inteligência Artificial

L.EIC - Grupo 58

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

André Leonor – up201806860
Luís Diogo – up201806340
Renato Leite – up201908633

Problem Specification

The effectiveness of a sound wave fire-extinguishing system can be predicted. The objective of this project is to apply machine learning models and algorithms related to supervised learning to a given dataset. Specifically, the goal is to classify examples in terms of a target concept using at least three supervised learning (classification) algorithms, such as Decision Trees, Neural Networks, K-NN, SVM and Random Forest and compare their performance using appropriate evaluation metrics, such as precision, recall, accuracy, F1 measure, etc. The project also involves dataset analysis to check for the need for data pre-processing, definition of training and test sets, selection and parameterization of learning algorithms, and evaluation of the learning process on the test set.

Problem Specification

The dataset contains the following data:

- Distance between the fuel container and measurment devices;
- Airflow;
- Sound intensity (decibel);
- Frequency of the sound waves;
- Temperature of the flame;
- Fuel type;
- Size of the fuel container;
- Flame extinction/non-extinction states.

Description of the tools and algorithms

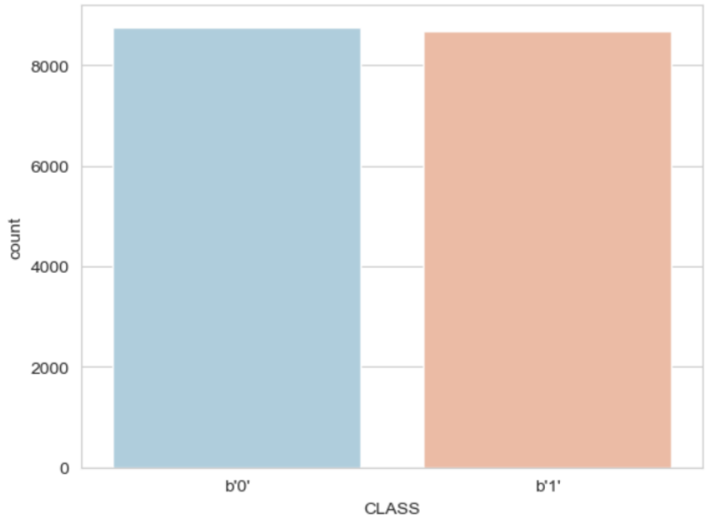
Tools - We used Python as programming language, programming in Python/Jupyter Notebook environment. For machine learning algorithms, we used Numpy to process our data, Pandas to read and handle the data and Seaborn and Matplotlib to visualize it.

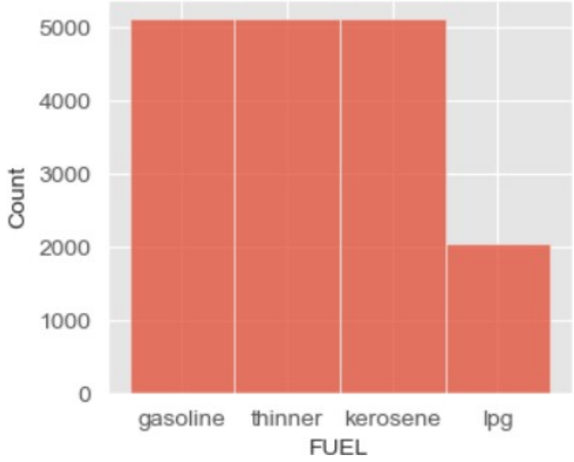
Algorithms – We used different types of algorithms like Random Forest, Neural Networks, Decision Trees, Support Vector Machine and K-Nearest Neighbours.

Data Pre-processing

The data pre-processing used in this project consists of the following steps:

- Initial exploration: Since the dataset is too large, the first few rows were displayed to get an overview of the data.
- Checking data information: Check the data types and non-null counts of each column in the DataFrame.
- Summary statistics: Numerical summary statistics (mean, standard deviation, minimum, maximum, quartiles) were calculated.
- Visualizing class distribution: Using a countplot, which displayed the number of extinguished (CLASS=1) and unextinguished fires (CLASS=0).
- Visualizing fuel usage: A histogram was plotted to analyze the distribution of fuel usage. It showed that LPG fuel was less used compared to other fuels.





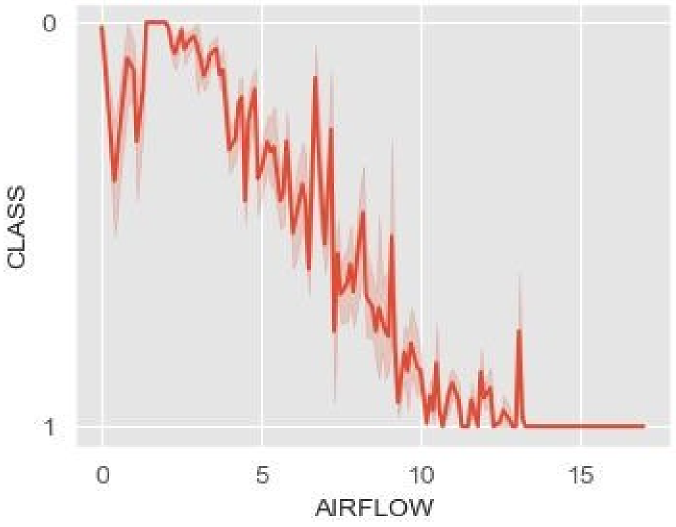
U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

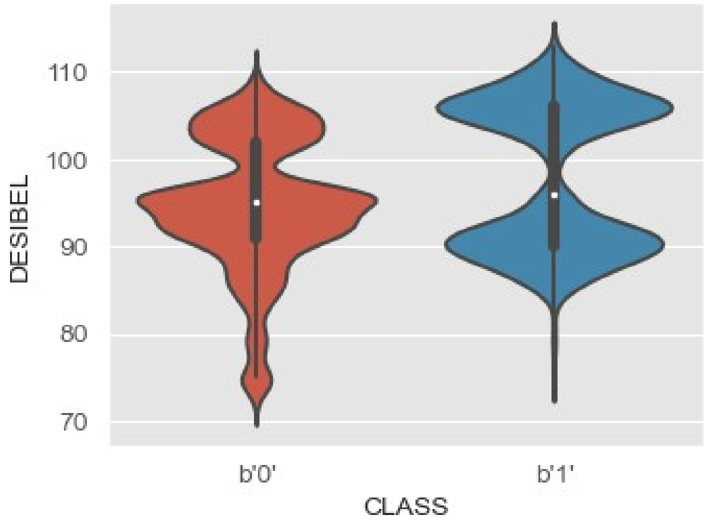
Data Pre-processing

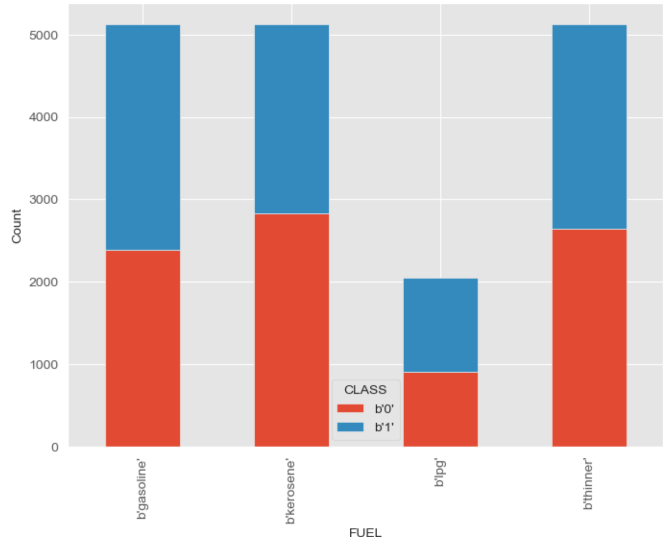
- Analyzing pairs of variables - Several plots were created to analyze the relationship between variables:
  - A line plot showed how the flame class (extinguished or not) varied with the airflow. When airflow was lower, the flame class was non-extinguished (CLASS=0), and as airflow increased, the flame class shifted to extinguished (CLASS=1).
  - A violin plot displayed the distribution of decibel levels (DESIBEL) for each flame class. It indicated that lower decibel levels were more likely for extinguished flames (CLASS=1).
  - A stacked bar chart was used to visualize the relationship between fuel type and flame class, showing the count of extinguished and unextinguished fires for each fuel type.
  - A bar plot compared the size of flames for each fuel type. It revealed that flames using LPG fuel had a higher size compared to other fuels.
  - Box plots were created to analyze the relationship between flame class and distance as well as flame class and frequency. The plots showed that lower distances and certain frequency ranges were associated with extinguished flames.

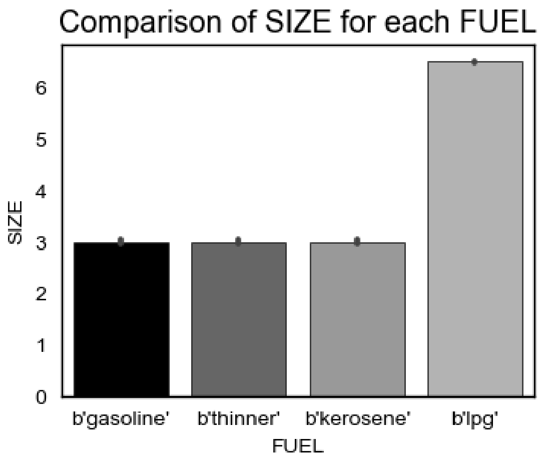
U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

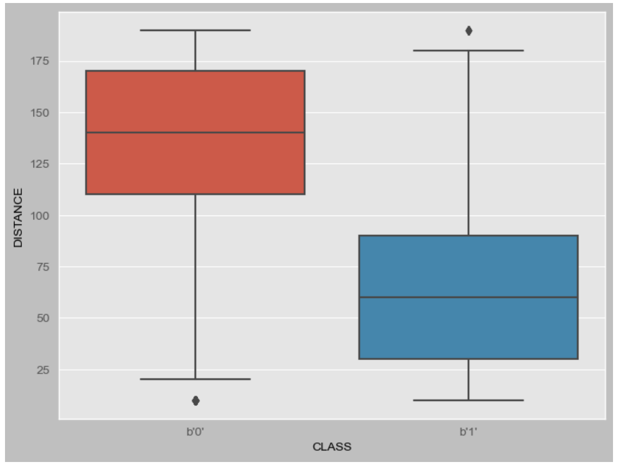
Data Pre-processing

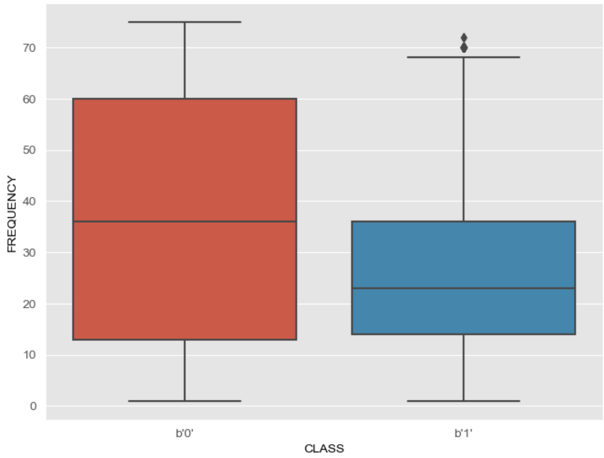








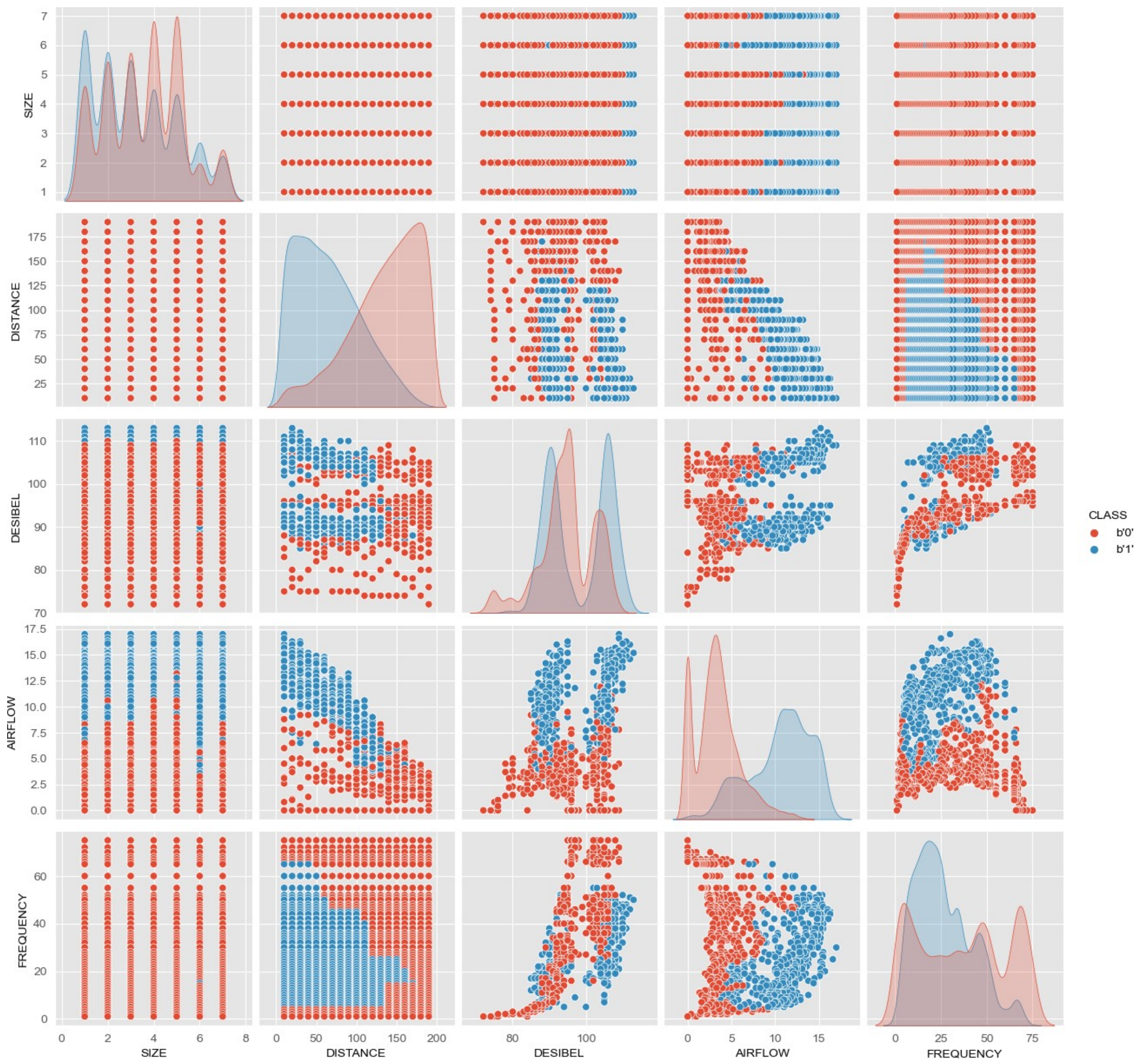




U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Data Pre-processing

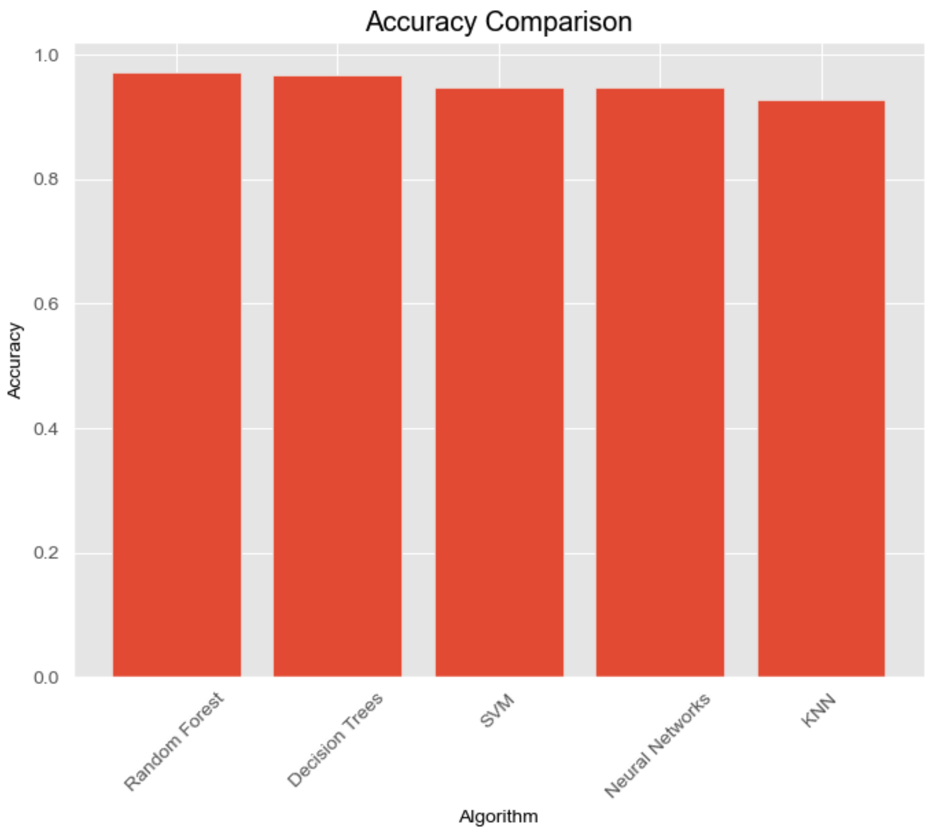
- Pairwise relationships: A pairplot was generated using the pairplot() function from seaborn to visualize the pairwise relationships between all numerical variables. The hue parameter was set to 'CLASS' to differentiate extinguished and unextinguished flames.

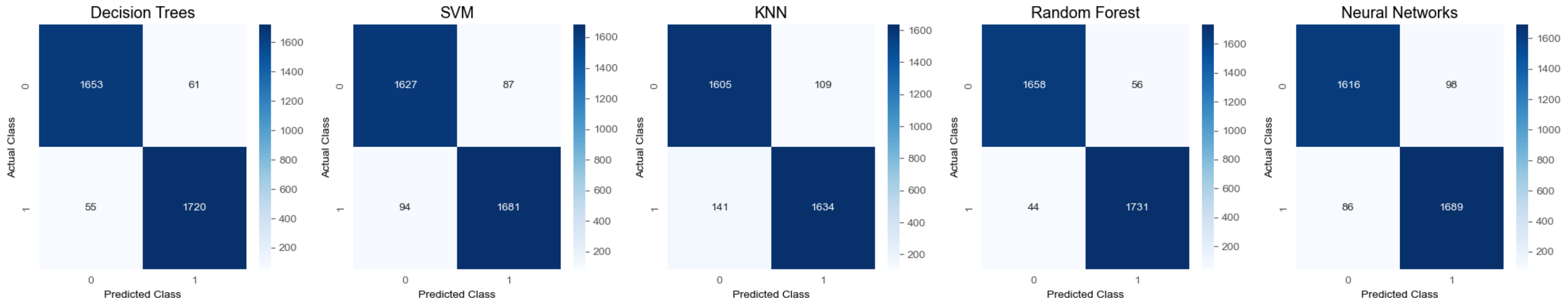


U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Developed Models

In order to try to determine the best model for this problem, we tried 5 different models and compared its respective accuracies and confusion matrixes as we can observe in the following graphics:





U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Conclusion

- There was no need to perform a great "cleaning" of our dataset as we didn't find any outliers nor missing/duplicated values.
- We had to convert our 'FUEL' parameter to numerical values so we could be able to work with it. We also had to convert our 'Class' values to integers because we couldn't find another way to handle it.
- Since the dataset used is very large, we expected algorithms like Random Forest and Support Vector Machine to perform well and that turned out to be true.
- We were expecting better results from Neural Networks in comparison with the other algorithms used, due to its excellent performance on large datasets, but such wasn't verified.
- Clearly the worst one, in terms of accuracy, was K-Nearest Neighbours as we suspected because it works poorly then the others when handling datasets of this size.
- Overall, we had better results then we expected at the beginning, having accuracies between 92,8% and 97,1%, and, therefore, we believe have completed our goals successfully.

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO