Dataset Link:

The dataset used in the project can be found on Kaggle:

Description of Project:

The project is a flight cancellation prediction model that uses four different classifiers (Logistic Regression, Random Forest, XGBoost, and LightGBM) to predict whether a flight will be cancelled or not. The project uses a dataset from the Bureau of Transportation Statistics that contains information on domestic US flights from 2003 to 2019.

A brief explanation of the outputs of the Flight Delay Prediction Project

1. Data preprocessing report: This report outlines the steps taken to clean and preprocess the raw data, including removing duplicates, handling missing values, and converting data types.

2. Feature engineering report: This report describes the features created to enhance the predictive power of the model, such as creating new features from existing data and scaling the data to improve model performance.

3. Model training and evaluation report: This report describes the performance of each of the four classifiers used in the project, including their accuracy, precision, recall, and F1 scores. It also includes visualizations of the model performance, such as ROC curves and confusion matrices.

4. Prediction report: This report provides the predictions of the model on a new dataset, indicating whether each flight is predicted to be cancelled or not.

Description of Output:

The main output of the project is the prediction report, which provides the predictions of the model on a new dataset. The report indicates whether each flight is predicted to be cancelled or not based on the features provided in the dataset. The report includes a binary classification for each flight, indicating whether it is predicted to be cancelled or not. This output is useful for airlines and passengers to anticipate possible flight cancellations and make necessary adjustments to their travel plans.

Instructions on How to Run the Code/Project/File:

1. To run the code in the notebook, you will need to have Jupyter Notebook installed on your computer. Once you have installed the Jupyter Notebook, you can download the notebook from the Kaggle website and open it in the Jupyter Notebook.
2. Before running the code, you will need to make sure that you have downloaded the necessary data files and saved them in the correct directory. The notebook provides instructions on how to download the data files and where to save them.
3. Once you have downloaded the data files and opened the notebook in Jupyter Notebook, you can run each cell of the notebook by clicking on the cell and then clicking the "Run" button in the toolbar or by using the keyboard shortcut "Shift + Enter".
4. It is recommended that you run the code cells in order, as some cells depend on the output of earlier cells. The notebook also provides explanations and comments for each code cell, so beginners can follow along and understand the code.

Note: Make sure to update the file paths in the code cells to match the location of the downloaded dataset and kernel files on your local machine.