**FLIGHT DELAY PREDICTION**

**FOR**

**AVIATION INDUSTRY**

**USING**

**MACHINE LEARNING**

**SUBMITTED BY,**

**TEAM LEADER: L.KALAIYARASI.**

**TEAM MATES:**

**P.KARTHIKA**

**M.KAVYA**

**R.NARMADHA**

**INTRODUCTION**

1. **OVERVIEW**
2. **PURPOSE**

* **PROJECT FLOW**
* **PROECT STRUCTURE**

1. **ADVANTAGES**
2. **DISADVANTAGES**
3. **APPLICATIONS**
4. **CONCLUSION**
5. **FUTURE SCOPE**

***OVERVIEW***

* OVER the last twenty years, air travel has been increasingly preferred among travelers mainly because of its speed and in some cases comfort.
* This has led to phenomenal growth in air traffic and on the ground. An increases in air traffic growth has also resulted in massive levels of aircraft delays on the ground and the air.
* These delays are responsible for large economic and environmental losses. According to, taxi-out operations are responsible for 4,000 tons of hydrocarbons, 8,000 tons if nitrogen oxides and 45,000 tons of carbon monoxide emissions in the United States in 2007.
* Moreover, the economic impact of flight delays for domestic flights in the US is estimated to be more than $19 Billion per year to the airlines and over $41 Billion per year to the national economy in response to growing concerns of fuel emissions and their negative impact on health, there is active research in the aviation industry for finding techniques to predict flight delays accurately in order to optimize flight operations and minimize delays.
* Using a machine learning model, we can predict flight arrival delays. The input to our algorithm is rows of feature vector like departure date, departure delay, departure distance between the two airports , scheduled arrival time etc.
* We then use decision tree classifier to predict if the flight arrival will be delayed or not. A flight is delayed when difference between scheduled and actual arrival times is greater than 15 minutes.

Furthermore, we compare decision tree classifier with logistic regression and a simple neural network for various figures of merit. Finally, it will be integrated to web based application.

**PURPOSE**:

* Therefore, predicting flight delays can improve airline operations and passenger, which will result in a positive impact on the economy. In this study, the satisfaction main goal is to compare the performance of machine learning classification algorithms when predicting flight delays.
* By using Machine Learning (ML) Algorithms you can try to predict if your flight will be delayed in many ways. Of course, all of these different algorithms will have pitfalls and a certain degree of accuracy which will be associated to the data that they are fed.

**PROJECT FLOW:**

* User interacts with the UI to enter the input.
* Entered inputs is analysed by the model which is integrated.
* Once model analyses the input the prediction is showcased on the UI.

**To accomplish this, we have to complete all the activities listed below,**

* **Define Problem/problem Understanding**

1. **Specify the business problem**
2. **Business requirements**
3. **Literature Survey**
4. **Social or Business impact.**

* **Data Collection & Preparation**

**I. Collect the dataset**

**II. Preparation**

* **Exploratory Data Analysis**

1. **Descriptive statistical**
2. **Visual Analysis**

* **Model Building**

1. **Training the model in multiple algorithms**
2. **Testing and model**

* **Performance Testing & Hyper Tuning**

1. **Testing model with multiple evaluation metrics**
2. **Comparing model accuracy before & after applying hyperparameter tuning**

* **Model Deployment**

1. **Save the best model**
2. **Integrate with Web Framework**

* **Project Demonstration & Documentation**

1. **Record explanation Video for project end to end solution**
2. **Project Documentation Step by step project development procedure**

**PROJECT STRUCTURE**

**Create the Project folder which contains files as shown below**

* **We are building a flask application which needs HTML pages stored in the templates folder and a python script app.py for scripting**
* **flight.pkl is our saved model. Further we will use this model for flask integration.**
* **Training folder contains a model training file.**

**ADVANTAGES:**

* Therefore, predicting flight delays can improve airline operations and passenger satisfaction, which will result in a positive impact on the economy.
* In this study, the main goal is to compare the performance of machine learning classification algorithms when predicting flight delays.

**DISADVANTAGES***:*

* Flight delays not only irritate air passengers and disrupt their schedules but also cause a decrease in efficiency, an increase in capital costs, reallocation of flight crews and aircraft, and additional crew expenses.
* Delays and cancellations affect both passengers and air carriers. By resulting in increased travel time and increased expenses on food and lodging, they cause stress among passengers. Further, they disrupt the purpose of air travel — rapid, affordable and safe — and make the passengers distrust airlines.

**APPLICATIONS**:

* **Airlines** can determine efficient routes **with** minimum **delay** possibility. • Opt for secondary airports for particular routes ...
* AI in aviation has the potential to increase urban air mobility, improve airline safety, automate flight scheduling, and enable predictive maintenance of airplanes. AI algorithms analyze the flight data like route distance, altitudes, mileage, fuel use, aircraft type, weather conditions and a lot of other data points.

**FUTURE SCOPE:**

Scope of Aviation Management in India and Abroad

* In the next ten years, the aviation industry will witness huge investments of up to two lakh crore rupees, demand 935 more planes and traffic growth of 13% CAGR. This will eventually lead to a huge rise in employment opportunities.

Future scope of aviation industry in India

* To satisfy the growing demand, by 2038, the country's aircraft fleet is anticipated to grow fourfold to about 2500 aircraft. The government has also pledged to build 100 additional airports by 2024 of the Ude Desh ka Aam Nagrik (UDAN) scheme. The air fleet number is also expected to increase from 600 to 1200.

**CONCLUSION:**

* The paper performed a prediction of the occurrence of flight delays by adapting it into a machine learning problem. A supervised machine learning approach in the form of binary classification was used for the prediction. Seven algorithms were used for delay prediction, and four measures were used for algorithms performance evaluation. Due to the imbalanced nature of the data set, evaluation measures were weighted to eliminate the dominant effect of non-delayed flights over delayed flights. After applying classifiers to the delay prediction, the values of their four measures were compared to evaluate the performance of each model.
* The result shows that the highest values of accuracy, precision, recall, and f1-score are generated by the Decision Tree model (accuracy: 0.9778; precision: 0.9777; recall: 0.9778; f1-score: 0.9778). Such high values indicate that the Decision Tree performs well when predicting flight delays in the data set. Other tree-based ensemble classifiers also show good performance. Random Forest and Gradient Boosted Tree have an accuracy of 0.9240 and 0.9334, significantly higher than the rest of the models. The other four base classifiers Logistic Regression, KNN, Gaussian Naïve Bayes, and SVM, are not tree-based and did not show good performance. The KNN model is the worst performed since its precision and f1-score are the lowest among the seven models.