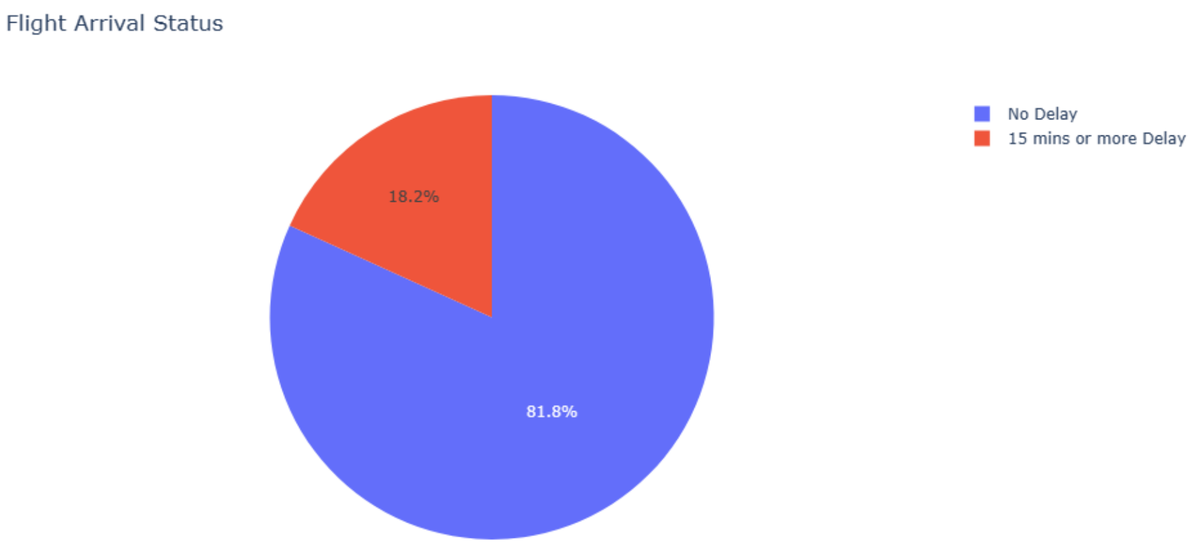
Air Transport AI-ML Case Study

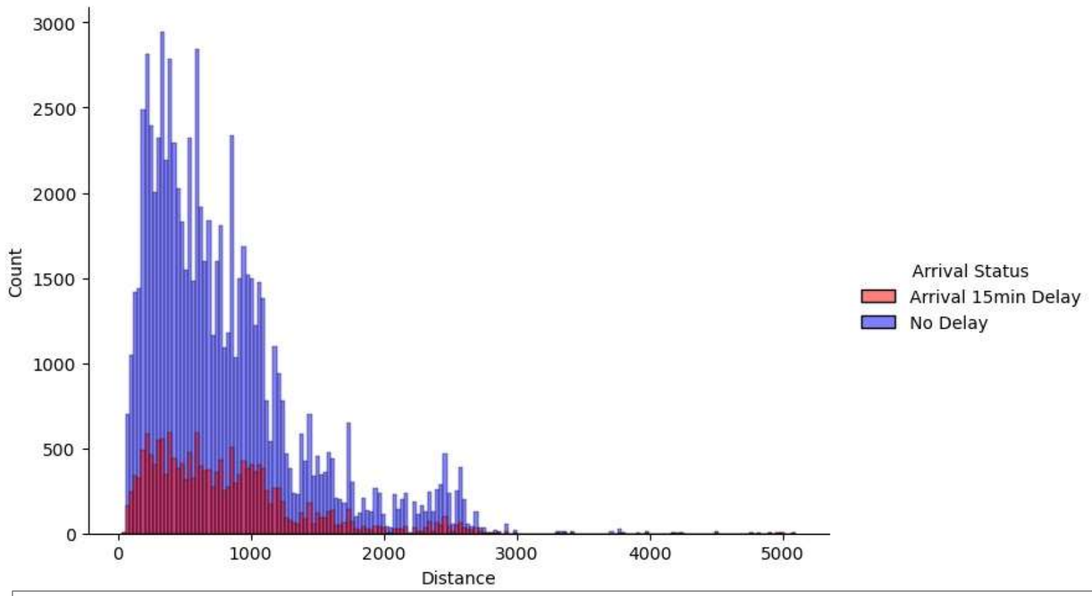
In the air transport industry, predicting the arrival status of flights accurately is a critical task that helps airlines and other stakeholders to plan and manage their operations more efficiently. Traditionally, this task is performed by human operators who manually analyze various factors such as weather conditions, air traffic, and historical data. However, this process can be time-consuming, costly, and prone to errors.

Auto-ML techniques offer a promising solution to automate the classification of flight arrival status. By analyzing large datasets of flight information and historical data, Auto-ML models can learn to distinguish between different categories of flight arrival status with high accuracy. This can significantly reduce the time and cost required for flight prediction, while also improving the consistency and reliability of the results .In this study we aim to develop an Auto-ML model that can accurately classify flight arrival status based on various factors such as weather conditions, air traffic, and historical data. The model will be trained on a large dataset of flight information and their corresponding labels. Once trained, the model can be used by airlines, airports, and other stakeholders in the air transport industry to quickly and accurately predict flight arrival status, and optimize their operations accordingly.

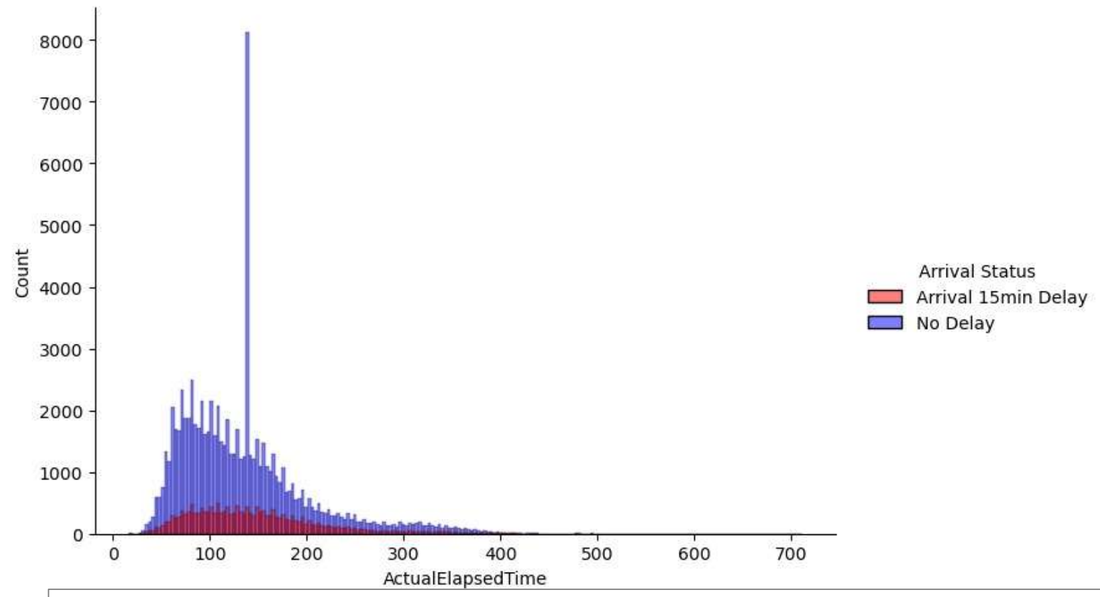


Arrival Status	Count
No Delay	81785
15 min or more delays	18215

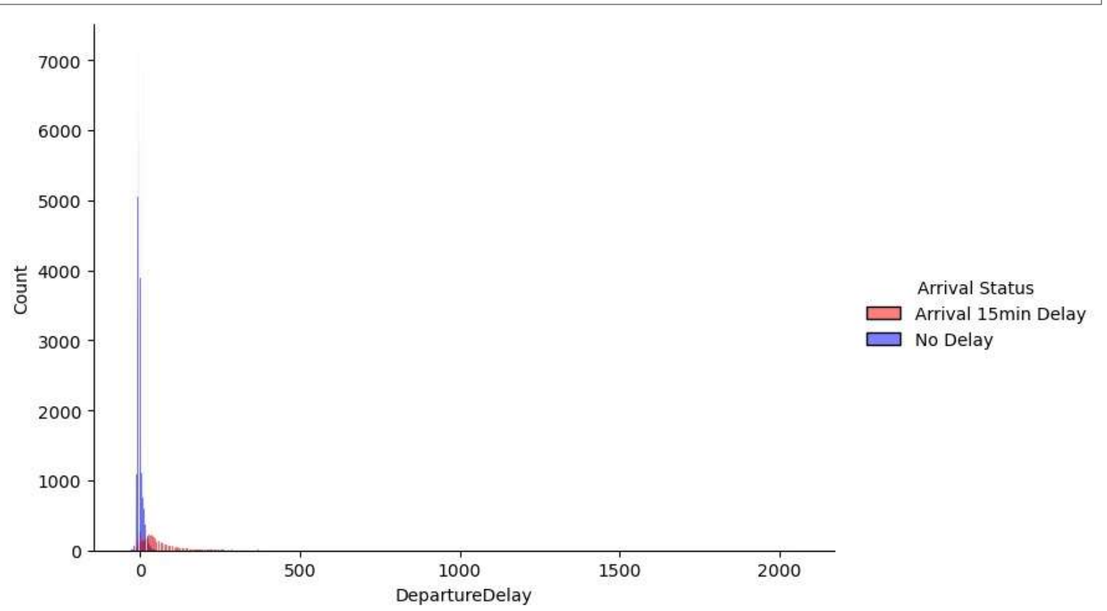
Features Responsible



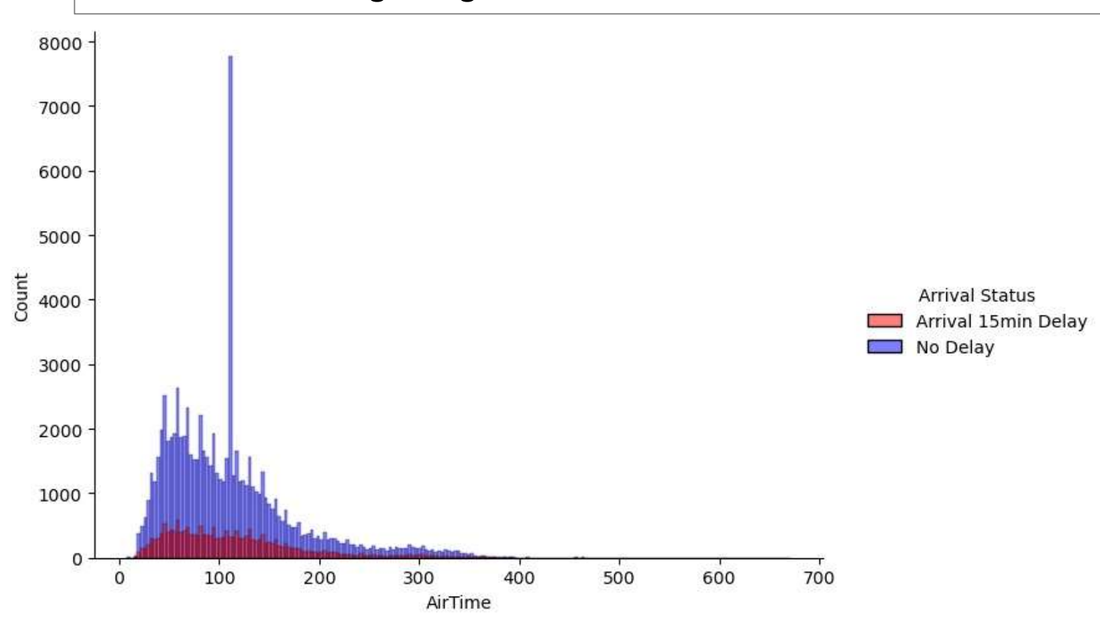
- **Distance**: Maximum delay are observed when distance < 1000 miles. Longer distance flights tend to have a higher probability of experiencing delays or other arrival status issues compared to shorter distance flights.



- **Actual Elapsed Time**: Actual elapsed time, which is the actual time taken for a flight to complete its journey from departure to arrival. The main reasons why actual elapsed time impacts the arrival status is due to the scheduling of flights.



- **Departure Delay**: When a flight is delayed at its origin airport, it can cause a ripple effect that affects the flight's schedule and arrival status throughout the day.



- **Air Time:** Air time, which is the time a flight spends in the air between takeoff and landing. Longer air times can increase the likelihood of delays, diversions, or cancellations due to various factors such as weather conditions, air traffic congestion, and mechanical issues.

Auto-ML Methodology Results

Case	Percentile	No. of Features	Random Forest	XGBoost	RNN	MLP	Decision Tree	Avg. Accuracy
Case 1	25	14	90.04	90.85	85.16	85.87	76.02	**85.58**
Case 2	50	28	91.26	91.11	78.63	88.41	76.2	**85.32**
Case 3	75	42	93.22	94.49	50	59.93	78.95	**75.31**
Case 4	90	50	93.18	94.79	50	50	78.44	**73.28**

- Based on our observation , XGBoost was the best performing algorithm with 94.79% accuracy in 75^{th} and 90^{th} percentile.
- 25^{th} percentile is the best percentile with an average accuracy of 85.58%.

Conclusion

Flight arrival status prediction is a crucial aspect of the air transport industry as it helps airlines and other stakeholders to plan and manage their operations more efficiently. The dataset contains features of the flight including information about origin airport, destination airport, airplane information, departure time and arrival time, delay time etc. of flights in the month of January 2020. The dataset contains 100000 records with 21 Categorical features and 36 Numerical features. 81.8% of the dataset show flights with no arrival delays.

For classification, models were created with algorithms using Auto-ML techniques like Decision tree, Recurrent Neural Network, Multilayer Perceptron, Random forest and XGBoost . With these models, performance measurement values were obtained for feature sets of 14, 28, 42 and 50. The Auto-ML algorithms were able to predict flights arrival status with an average accuracy between 70% – 86% and helped to identify factors that determine the flight delays. The major factors include Distance, Departure Delay, Actual Elapsed Time and Air Time. When the results are examined, it is observed that with the addition of each new feature, the success of classification decreased. Based on the performance measurement values obtained, it is possible to say that the study achieved success in classifying in flight arrival status.