1.DOMAIN BACKGROUND

Flight delay is a challenging problem for all airline companies , which will lead to financial losses and negative impact on their business reputation .Airline industry incurs an average cost of $11300 per delayed flight , based on 61000 delayed flight per month average .We can build some predictive model which can predict flight delay in advance which can help airline companies to optimize operation and reduce further loss.

2.PROBLEM STATEMENT

The real problem lies in predicting the flight delay in advance with good accuracy .One of the biggest challenge could be to incorporate dynamic factors like weather and wind speed for model training and predicting.

3.DATASET AND INPUT

Currently i am scraping data from flightradar24.com which provides real time update on status of flights .I am capturing following details for a particular flight:
3.1 tail_id
3.2 flight number
3.3 origin
3.4 destination
3.5 scheduled departure
3.6 actual time of departure
3.7 scheduled arrival
3.8 actual time of arrival
3.9 data of event
3.10 time
3.11 duration of flight

4. SOLUTION STATEMENT

The above described problem can be be solved by building a predictive model using supervised learning techniques. Algorithms like Naive Bayes , Decision Tree classifier ,Support Vector Machines can be tested on training data to verify the performance .

5. BENCHMARK MODEL

The benchmark model for predicting flight delay could be Naive Bayes Algorithm which considers Bayes Rule of conditional probability while making prediction on unseen data.

6.EVALUATION METRIC

The proposed evaluation metric could be accuracy of the model and while training the model we can look for precision and recall to optimize.

7.OUTLINE OF PROJECT DESIGN

First step is to gather past data and perform data wrangling techniques to check for validity of the data . Once we have preprocessed data we can train some supervised machine learning models and compare the result .we can also check how neural networks can be used for training such model . Once final model is decided we can test it on test dataset .