PROJECT DESIGN

PHASE-I

PROPOSED SOLUTION

Date	24 September 2022
Team ID	PNT2022TMID00047
Project Name	Developing a Flight Delay Prediction Model using Machine Learning

S.NO.	PARAMETER	DESCRIPTION
1.	**Problem statement**	• 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 increase in air traffic growth has also resulted in massive levels of aircraft delays on the ground and in the air. • These delays are responsible for large economic and environmental losses. • The main objective of the model is to predict flight delays accurately in order to optimize flight operations and minimize delays.
2.	**Idea / solution description**	• 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, 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 considered to be 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. •Provides a help and support corner.
3.	**Novelty/Uniqueness**	• Sending notifications about flight delays to the user. • Simultaneous access by a number of users at a time.

4.	**Social impact**	• This model is beneficial for both aviation industry and passenger travel. • Delays are calculated against scheduled block times as well as against more idealized feasible flight times. • Based on econometric estimations, welfare impacts of flight delays are calculated. • We find that flight delays on a route reduce passenger demand and raise airfares, producing significant decreases in both consumer and producer welfare. • Since producer welfare effects are estimated to be three times as large as consumer welfare effects
5.	**Business model**	• Flight delay has become widespread in the United States with nearly one-quarter of all flights delayed by more than 15 minutes in 2007. • US net welfare would increase by $17.6 billion for a 10 per cent reduction in flight delay and by $38.5 billion for a 30 per cent reduction.
6.	**Feasibility of idea**	• Compatibility with all devices. • The assessment of all the contributing factors is proposed. • This model can be used to obtain future flight fluctuations before scheduling future flights, then guide the allocation of airport resources such as parking spaces and optimize resource utilization.
7.	**Scalability of solution**	• Two open datasets of airline flights and weather observations have been collected and exploratory data analysis has been performed to discover initial insights, evaluate the quality of data, and identify potentially interesting subsets. • Then, data pre-processing and transformation (joining and balancing operations) have been performed to make data ready for modelling. Finally, a parallel version of the Random Forest data classification algorithm has been implemented, iteratively calibrating its settings to optimize results in terms of accuracy and recall. • The data preparation and mining tasks have been implemented on a Cloud infrastructure. • Other than providing the necessary computing resources for our experiments, the Cloud makes the proposed process more general: in fact, if the amount of data increases (e.g., by extending the analysis to many years of flight and weather data), the Cloud can provide the required resources with a high level of elasticity, reliability, and scalability