Project Design Phase-I
Proposed Solution Template

Date	26 September 2022
Team ID	PNT2022TMID24528
Project Name	Developing A Flight Delay Prediction Model Using Machine Learning
Maximum Marks	2 Marks

Proposed Solution Template:

The following information in proposed solution template.

S. No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Air travel has been increasingly preferred among travellers, 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
3.	Novelty / Uniqueness	Machine learning is the designation of algorithms that enable the computer to analyse the data, obtain potential patterns, and then use them to predict. Learning algorithms can give insight into the relative difficulty of learning in different environments. Machine learning algorithms are divided into several categories, and the two most common types are supervised learning and unsupervised learning. Algorithms of supervised learning generated a function that translates inputs to desired outputs. The primary forms of supervised learning algorithms include regression and classification. Unsupervised learning models a collection of inputs in the absence of labelled examples.

4.	Social Impact / Customer Satisfaction	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
5.	Business Model (Revenue Model)	Knowing the probability of flight delay or cancellation is a crucial tool for travellers, so we set about creating a **model to predict long-term flight delays**. Rather than looking at disruptions caused by punctual factors like weather, we wanted to see which flights and itineraries had the highest probability of delays or cancellations over time.
6.	Scalability of the Solution	Airline flight and weather observation datasets have been analysed and mined using parallel algorithms implemented as MapReduce programs executed on a Cloud platform. The results show a high accuracy in predicting delays above a given threshold. For instance, with a delay threshold of 15min, we achieve an accuracy of 74.2% and 71.8% recall on delayed flights, while with a threshold of 60min, the accuracy is 85.8% and the delay recall is 86.9%. Furthermore, the experimental results demonstrate the predictor scalability that can be achieved performing data preparation and mining tasks as MapReduce applications on the Cloud.