**Project Design Phase-I**

**Proposed Solution Template**

|  |  |
| --- | --- |
| Date | 19 September 2022 |
| Team ID | PNT2022TMID46892 |
| Project Name | Project - Developing a Flight Delay Prediction Model using Machine Learning |
| Maximum Marks | 2 Marks |

**Proposed Solution Template:**

Project team shall fill the following information in proposed solution template.

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Parameter** | **Description** |
|  | Problem Statement (Problem to be solved) | My case study was about LaGuardia Airport in New York, Logan International Airport in Boston, San Francisco International Airport in San Francisco which are four major airports in the United States of America. But we focused the idea and research on LaGuardia International Airport. Compared with the data produced by all airports in USA, the data which we gathered was very limited, but it gave us a great direction on how weather plays a part in flight delays. In this project, the goal is to use exploratory analysis and to build machine learning models to predict airline departure and arrival delays. |
|  | Idea / Solution description | As discussed, weather condition plays an important role in proper and timely functioning of flights. We propose a flight delay prediction system which focuses mainly on predicting delay of a flight based on the weather situation. To make the system more scalable it is necessary to choose an algorithm which considers all the parameters to be independent. Supervised learning as the name indicates a presence of supervisor as teacher. Essentially supervised learning could be a learning that within which we tend to teach or train the machine exploitation data which is well tagged which means some data is already labeled with correct answer. After that, machine is given new set of examples(data) so supervised learning algorithmic rule analyses the coaching knowledge(set of training ) and produces an correct outcome from tagged data Using supervised machine learning approach, the labelled data gives it authenticity. Naïve bayed model is one of the algorithm which is proven to be efficient for real time prediction as well as the fact that it considers every attribute to be independent from each other makes it an apt algorithm for the concerned project |
|  | Novelty / Uniqueness | The main objective of this project is to predict if a ﬂight will be delayed or not, hence we chose the following 13 out of 30 features which are usually known in advance: Month, Day, Day of the week, Flight Number, Origin airport, Destination Airport, Scheduled departure, departure delay, taxi-out, distance, Scheduled Arrival. |
|  | 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 (Britton et al., 2014; Yablonsky et al., 2017). Flight delay is inevitable and it plays an important role in both profits and loss of the airlines. An accurate estimation of flight delay is critical for airlines because the results can be applied to increase customer satisfaction and incomes of airline agencies. |
|  | Business Model (Revenue Model) | In this project, we were able to successfully apply machine learning algorithms to predict flight arrival-delay and show simple classier like decision tree and logistic regression can predict if a ﬂight’s arrival will be delayed or not fairly accurately. |
|  | Scalability of the Solution | we used 1000 thousand samples to train and test the three classiﬁers with the recommended 700-300 split. We performed 100fold cross-validation for decision tree and neural network classiﬁers and used the scikit and keras API where ever necessary. For the decision tree algorithm we parameterized the depth of the tree for better accuracy, and for logistic regression and neural network we applied L2 regularization to prevent the model from over-ﬁtting. In addition to traditional classiﬁcation ﬁgures-of-merit like AUC, precision, accuracy, recall, we also like to know tree depth and the total leaf nodes for the decision tree classiﬁer. |