Optimizing Flight Booking Decisions through Machine Learning Price Predictions

Milestone 1: Define Problem / Problem Understanding

Activity 1: Specify the business problem

Business problem related to flight price prediction is to develop a real-time pricing strategy at can dynamically adjust prices based on factors such as demand, competition, and seasonality. Airlines need to analyze a wide range of data sources, such as historical pricing trends, customer behavior, and market trends, to develop an effective pricing strategy that can maximize revenue and profitability. Machine learning algorithms can assist in developing predictive models that can analyze real-time data and recommend optimal pricing strategies.

Activity 2: Business requirements

Accurate Flight Price Prediction: The primary business requirement is to develop a reliable and accurate flight price prediction model that can predict prices with a high degree of accuracy. This model should be able to incorporate various factors that affect pricing, including demand, competition, seasonality, and customer behavior.

Activity 3: Literature Survey (Student Will Write)

In the research work done by Twinkle B Ankleshwaria et.al: In the field of information processing Data mining refers to mining knowledge from large amounts of data .Similarly,Mining Data Streams is concerned with extracting knowledge from non stopping and continuous stream of information Thetopic of data streams is very recent one . A data stream is a massive, infinite, temporally ordered, continuous and rapidsequence of data elements . Research on data stream is motivated by emerging applications involving massive data sets suchas customer click streams, supermarket, telephone records, stock market, meteorological research, multimedia data, scientificand engineering experiments and sensor data. A new generation of mining algorithms are needed for real–time analysis &query response for these applications since most conventional data mining algorithms can only be applied to static data sets thatmay be updated periodically in large chunks, but not in continuous streams of data. While data mining has become a fairly wellestablished field now, the data stream problem poses a no. of unique challenges which are not easily solved by traditional datamining methods. Some of issues of data stream like dynamic nature, Infinite size and high speed, unbounded memoryrequirements, Lack of global view, handling the continuous flow of data impose a great challenge for the researchers dealingwith streaming data.

In the research work done by Vivek Srivastava: Data mining, a process better known as knowledge discovery, is the process of analyzing data fromdifferent perspectives and summarizing it into useful information-information which could be ofeconomic

value (which can increase revenue, cut costs or both). Companies use data mining as customsoftware by writing in-house code for analysis of large chunk of data. It can help us to identify patterns; itcan categorize the data and find hidden relationships in it. It is a part of the process of knowledgediscovery from relational databases.

Flight price prediction - data mining

In the research work done by K. Tziridis, Th. Kalampokas, K.I. Diamantaras: They began their paper by giving a general introduction to machine learning, and then went on to discuss the methodology they used,which consists of four phases;Feature Selection, where the most important features that influence air ticket prices are determined; Feature Selection, where the most important features that influence air ticket prices are determined; Feature Selection, where the most important features that influence air ticket Data gathering comes next, followed by the selection of an accurate ML model, and ultimately, the model's evaluation.

In the research work done by Anzeen R.A: It might have happened so many times that you or someone yours needdoctors help immediately, but they are not available due to some reason. TheHealth Prediction system is an end user support and online consultation project.Here we propose a system that allows users to get instant guidance on their healthissues through an intelligent health care system online. The system is fed withvarious symptoms and the disease/illness associated with those systems. Thesystem allows user to share their symptoms and issues. It then processes user'ssymptoms to check for various illnesses that could be associated with it. Here weuse some intelligent data mining techniques to guess the most accurate illness thatcould be associated with patient's symptoms. In doctor module when doctor loginto the system doctor can view his patient details and the report of that patient.Doctor can view details about the patient search what patient searched for according to their prediction. Doctor can view his personal details. Admin can addnew disease details by specifying the type and symptoms of the disease into thedatabase.

Knn - prediction

In the research work done by J. Agarwal, R. Nagpal and R. Sehgal: An have analyzed crime and considered homicide crime takinginto account the corresponding year and that the trendis descending from 1990 to 2011. They have used the k-means clustering technique for extracting usefulinformation from the crime dataset using Rapid Minertool because it is solid and complete package with flexible support options.

In the reseach work done by Vrushali Pednekar: Crime is one of the most predominant and alarming aspects in our society and its prevention is a vital task. Crime analysis is asystematic way of detecting and investigating patterns and trends in crime. Thus, it becomes necessary to study various reasons, factors and relationship between different crimes that are occurring and finding the most appropriate methods to control and avoid more crimes. The mainobjective of this project is to classify clustered crimes based on occurrence frequency during different years. Data mining is used broadly interms of analysis, investigation and discovery of patterns for occurrence of different crimes. In this work, various clustering approaches of datamining are used to analyze the crime data. The K-Nearest Neighbour(KNN) classification is used for crime prediction.

Decision tree - prediction

In the research word done by Asma ,*A.Aljarullah :* Earlier lots of works have been done in this field and tosummarize a few of them have been mentioned below.Jianchao Han used WEKA decision tree to build and predict type 2 diabetes data set which considered only thePlasma Insulin attribute as the main attribute while neglectingthe other attributes given in the datasetperformed completely different classification algorithmswith variable accuracies and steered improved predictionaccuracy exploitation weighted statistical method SVM.also used WEKA decision tree classifier onthe diabetes information set with

association rule being enforced to get a mix of attributes.

In the reseach work done by Hai Pham-The: In this report, the blood brain barrier (BBB) permeability prediction is carried out using a decision tree. A recently published data set of 497 compounds is selected to develop the tree model. The developed model shows an accuracy of 87.66% for training set; 86.09% in the 10-fold cross-validation procedure and 87.93% for the test set. Some structural explanation of how our model describe the passage of molecules through the BBB is given. Moreover, a comparison with other approaches is carried out showing good behaviour of our method.

Randomforest - prediction

In the research word done by Patrick Cronin: The goal of this project is to be able to predict the onset of a ovulation cycle, given the termination point (date) of the prior ovulation cycle, i.e. the actual length of a cycle, given the past history of he patient. The valuation of this process is measure by how much been a prediction could be over a naive estimate. The naive estimate, ignore all other attributes is based upon the mean cycle length, which based on the current data set is 28.75 days

In the research word done by Zutherman Rustam: The Random Forest method has been discussed since Breiman published this method in machine learning.Starts by selecting m variables randomly from several independent variables, Breiman seeks to improve the estimation process previously carried out by the Decision Tree and CART. Then many trees are grown without any pruning process like the Decision Tree or CART method. Which of these trees will be chosen, which gives the highest accuracy results. Random forest procedure depends on a number of classifiers. This helps inminimizing bias, tolerate outliers, and avoid overfitting. The random forest has been utilized in several fields, andnow we will apply random forest for breast cancer.

In the reseach work done by Dinali Jayawardana,Andrew Benya: Using the random forest as opposed to one single decision tree for the stocks is better because decision trees are unstable. A small change in a decision tree can shift the overall structureof the decision tree. The random forest uses multiple decision trees which lowers the biases thatwould occur when using a single decision tree. The random forest also adds multiple points ofview to the stock predicting problem. Since there are multiple variable that contribute to the stock price fluctuation the random forest could create multiple decision trees to tackle the differentvariables creating a better model.

Activity 4: Social or Business Impact.

Social Improved Accessibility: Accurate flight price prediction models and real-time pricing strategies can make air travel more accessible to a wider range of people, including those with limited financial resources. By making air travel more affordable, these models can enable people to travel more frequently, and to access opportunities such as education, work, and leisure.

Business Improved Revenue Management: Accurate flight price prediction models and real-time pricing strategies can help airlines optimize their revenue management. By identifying demand patterns and adjusting prices accordingly, airlines can maximize their revenue and profitability.