Project Report

ITCS 6112: Software System Design and Implementation

Professor Name: Ali Server

Submitted by Group-3

Group Name: Team Wizards

Team Members:

Roshna Sanjana Kommareddy (801308381)

Srija Bandlamudi (801307500)

Kashmeera Akkenapally (801317282)

Rishika Annu (801311827)

Mounika Kasaragadda (801310485)

AIRLINE RECOMMENDATION SYSTEM

ABSTRACT

Nowadays, consumers increasingly rely on online sources to guide them take decisions about their shopping selections. Online tools can aid businesses in gaining insights from customer reviews, the priorities of various customer types, and how customers' choices impact their reviews. In the United States, 82% of adults report consulting Internet reviews and customer evaluations before making their first product purchase, with 40% reporting doing so constantly or almost constantly. The data that have been gathered by various internet sites have been subjected to predictive analysis in the suggested study for qualitative reviews. In addition, we looked at how different airline service categories were interpreted. According to the results of our investigation, the majority of passengers traveling in business class give the food quality and friendliness of the personnel very good ratings, while most passengers traveling in economy class give the legroom and comfort of their seats very high ratings.

CHAPTER 1

OVERVIEW

Online reviews have filled the gap between traditional word-of-mouth and an emergent form of expression, which has opened up new avenues in marketing and communication. Analyzingclient feedback gives businesses information about customer satisfaction and insight into what customers really desire. Businesses can enhance their customer service by immediately and effectively addressing customer complaints, resulting in a great customer experience, and prioritizing their requirements by using this information as feedback. Positive internet customer reviews are extremely valuable and can be more successful than traditional marketing initiatives.

Online evaluations consistently convey a positive impression of a good or service to prospective clients, helping the brand become more well-known and benefiting both current and future enterprises. 82% of American adults examine internet reviews or ratings before making a first-time purchase of any items, according to a Pew Research Center research, with 40% of respondents saying they do so almost regularly or usually. More than half of people between the ages of 18 and 29 and 47% of people between the ages of 30 and 49 said they almost usually read internet reviews before making a first-time purchase. The utilization of online reviews is influenced by age because older people tend to use the internet less regularly. The percentage of people who make purchases based on internet evaluations declines with age, from 34% of persons in the 50–64 age range to 23% of people 65 and beyond. While there is a modest age variation in the likelihood of providing feedback, Americans under 50 are more likely than other age groups to share their product assessments.

Online evaluations have made it possible for firms to understand customer reactions and pinpoint areas for development in the fiercely competitive airline industry, where there are multiple options for the same routes. But the unstructuredness, irregularities, and size of review data might make them difficult to assess. This paper used natural language processing and machine learning methods to analyze the reviews and derive insights in order to get around these problems. These methods included stemming, normalization, and n-gram data calculations. Predictive analysis was also used to ascertain whether or not a user would promote the product using a variety of models.

A. CUSTOMER FEEDBACK AND ENDORSEMENT

One quantifiable way to gauge client happiness is through their ratings. Since ratings act as feedback for businesses to ascertain if customers are likely to purchase a product, they are crucial for performance evaluations and enhancing the customer experience. One of the most trustworthy ways to assess a company's potential for expansion is by looking at its Net Promoter Score (NPS). Where Net Promoter Score represents the distinction between advocates andopponents. NPS is a measure of the success of promotional operations, according to numerous other academics. The ability to maintain a customer's interest in a product can be offered by offering a satisfactory purchasing experience that can result in repeated purchases. Numerous scholars are interested in comprehending consumer purchasing behavior due to the significance of client happiness. The quantity and quality of the product play a role in the consumer's decisionto purchase it. If a product receives a greater percentage of good reviews, this will encourage buyers to buy it.

B. ONLINE EVALUATIONS

Customers now rely on reviews for their decision-making in the quickly expanding internet era and growing popularity of online services to eliminate any confusion related to the purchase experience. You can find client testimonials on their Facebook and Twitter sites on online travel portals. For online shops, favorable reviews are extremely valuable since they draw customers who may end up purchasing the services that were the subject of the review. If the review ratingsrise by more than 5%, There has been a 10% spike in online hotel reservations. Thus, online reviews have an impact on the service being reviewed. One of the concerns with internet reviews is their authenticity. Various factors, such as the author's emotional state at the time of writing, can have an impact. We only need to know what consumers say to their friends—or what is known as word of mouth—in order to gauge customer happiness or retention. One of the finest ways to determine consumer happiness is through word of mouth. Travelers primarily use social media to connect and organize their excursions. They can also assist tourism service providers, and overcome the limitations of conventional information sources (e.g., by offering pertinent information). According to surveys on tourism, between 20 and 45 percent of visitors usereviews to help them make decisions, and between 5 and 30 percent use them to share their experiences.

Examination Of The Text Of Internet Reviews

Text analysis is the automated process of extracting information from unstructured text data, such as emails, tweets, and product reviews. Unstructured text data contain enormous amounts of information. Text analysis transforms the text into information that computers can use. In this study, data are extracted from text using text mining techniques, then machine learning models are used to analyze the data. Each word in a corpus or document is evaluated using the TF-IDF metric. The TF-IDF is determined by the frequency with which a word appears in a document, and it is then normalized by the proportion of documents in the corpus that contain that word. An automated technique for identifying opinions in text is sentiment analysis. Finding out if a review is positive or negative and the reasons influencing it can be done by analyzing the sentiment of the reviews.

CHAPTER 2

LITERATURE REVIEW

A summary of the literature review is provided in this chapter. Some of the researchers' pertinent work is represented in this chapter. Researchers have examined numerous methods for predicting airline prices and delays, but our work is the first to propose a method for predicting airline reviews from social media material. A few of these methods are detailed below.

Flight Delay Prediction Using Machine Learning and Big Data in Aviation

IEEE Transactions on Vehicular Technology, vol. 69, no. 1, pp. 140–150, Jan. 2020; doi: 10.1109/TVT.2019.2954094; G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou, and D. Zhao.

Accurately forecasting flight delays is essential for increasing the efficiency of the airline sector. Utilizing machine learning strategies to forecast flight delays has been the focus of recent investigations. Prior prediction techniques, however, were restricted to particular airports or routes. While taking into account a larger variety of factors that may affect flight delays, this research attempts to assess multiple machine learning models for predicting flight delays in a generalized environment. To do this, a comprehensive dataset is created by combining acquired automated dependent surveillance-broadcast (ADS-B) messages with other information including weather, flight schedules, and airport information. Experimental findings indicate that long short-term memory (LSTM) can manage airplane sequence data and that it can handle a variety of classification and regression tasks. However, overfitting is an issue with small datasets. With regard to binary classification, the recommended random forest-based model outperforms earlier systems in terms of prediction accuracy (90.2%), and it also addresses the overfitting issue.

Using machine learning methods, predict flight delays caused by the weather

2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), pp. 1-6, doi: 10.1109/DASC.2016.7777956, S. Choi, Y. J. Kim, S. Briceno, and D. Mavris.

The main objective of this study is to forecast aircraft delays caused by inclement weather utilizing supervised machine learning and data mining approaches. Domestic US flight data and meteorological data from 2005 to 2015 were merged to train the model. Unbalanced training data was addressed using sampling techniques. For each individual flight delay prediction, different models were created, such as AdaBoost, k-Nearest-Neighbors, decision trees, and random forests. We compared the receiver operating characteristic (ROC) curve and prediction accuracy of the models. Weather forecasts and flight data were gathered and included into the model during the prediction step. The model was developed to do binary classification in order to predict whether a planned aircraft will take off on time or be delayed.

A New Method: Machine Learning-Based Prediction of Airline Delays

2018 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 1081–1086. V. Natarajan, S. Meenakshisundaram, G. Balasubramanian, and S. Sinha.

Passengers may become agitated as a result of ineffective airport management of aircraft delays. Researchers are tackling this problem as a result of recent developments in machine learning. To decrease delays and improve customer satisfaction in the air transportation system, an efficient decision-making process is needed. 19% of domestic flights in the US land with an average delay of 15 minutes. However, given the complexity of the aviation industry and the stochastic nature of delays, it is challenging to estimate flight delays with accuracy. By examining operational data from departure and arrival as well as historical weather data, this study seeks to qualitatively anticipate aircraft delays. To ascertain the delay status and performance, respectively, a decision tree model's performance and a logistic regression model's performance are evaluated. According to the study, logistic regression is not as good as the decision tree method. The simulation results imply that big airports may achieve low delays by taking into account a variety of variables, including the time of day and weather.

Machine learning for cost-sensitive airline delay prediction

2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC), S. Choi, Y. J. Kim, S. Briceno, and D. Mavris, pp. 1–8, doi: 10.1109/DASC.2017.8102035.

The costing sampling methodology and supervised machine learning techniques are combined in this study to create a framework for forecasting individual flight delays. In order to convert cost-insensitive classifiers into cost-sensitive ones, the costing strategy entails choosing instances from the original training dataset according to the costs associated with misclassification. The effectiveness of the model taking into account misclassification costs has recently been assessed using a weighted error function. The function considers a variety of cost ratios for false positive and false negative mistakes, which shows how heavily late and on-time classes are weighted. The model can achieve a lower weighted error rate when the cost ratio is set to 10, according to the relationship between the cost ratio and the weighted error rate.

Review of Automatic Ticket Price and Demand Detection for Airlines

2018 International Conference on Innovations in Information Technology (IIT), J. A. Abdella, N. Zaki, and K. Shuaib, pp. 169–174, doi: 10.1109/INNOVATIONS.2018.8606022.

Demand and price predictions for airline tickets are exceedingly difficult to make since they depend on a number of internal and external variables that can change quickly. The goal of the many ticket price/demand prediction models put forth by researchers is to either help the customer predict ticket pricing or help the airline predict demand. We review customer-side and airline-side prediction models in this research. The analysis carried out in our study demonstrates that both types of models rely on a small number of features, such as historical ticket pricing, the day the ticket was purchased, and the departure date.

Complex machine learning methods are not being combined with outside factors like search engine inquiries and social media data.

CHAPTER 3

ANALYSIS

ANALYZING A SYSTEM

Systems analysis is a technique for solving problems that entails dissecting systems into their component elements and evaluating how well those parts function both individually and collectively to achieve the intended goal.

AIM OF THE STUDY

Any airline system can make advantage of the scope of work. These days, airline firms evaluate their ratings and evaluations using sophisticated methodologies and procedures. Customer reviews, ratings for food and seating, and other factors are taken into account via machine learning and deep learning techniques. For customers to select the best airline suggested by our prediction technology, there are several social aspects that are closely related.

OBJECTIVES

The work's goal is to present a deep learning and machine learning-based recommendation system for airlines based on reviews and ratings. Convolutional neural networks and Random forests were employed in the project's prediction work. The study's goal is to provide customers with recommendations for airlines as user-friendly results when they select a source and destination. Utilizing the Random Forest method, this is accomplished.

PROBLEM STATEMENT

The airline sector provides a wide range of options on comparable routes, making it a very competitive market. Many businesses use online reviews to obtain insight into customer feedback and pinpoint areas where they can improve in order to acquire an advantage in this market. However, it is a difficult operation that is frequently impossible to complete manually to review such enormous amounts of unstructured data, which may contain anomalies and ambiguities.

REQUIREMENT ASSESSMENT

A dynamic study field that is quickly expanding in the online review detection sector is computer aided learning. Recent advances in machine learning research offer increased review prediction perception accuracy. In this case, the computers are given the ability to think through growing intelligent. Deep learning techniques come in a variety of forms and are employed to categorize data sets.

CONDITIONS FOR FUNCTION

The suggested application should be able to categorize and rate user ratings in order to suggest the best airline for customers. Functional requirements specify how a system and its parts will function and behave. Outlining a function's inputs, behavior, and outputs will help you describe it.

DATA GATHERING

The process of collecting data must include carefully selecting high-quality data for analysis. Here, we used Airlinequality.com to web scrape data for the application of machine learning. A data analyst's task is to identify methods and resources for gathering accurate and complete data, interpreting that data, and using statistical methods to analyze the outcomes.

DATA PREPARATION

Preprocessing is the process of transforming unstructured input into a form that machine learning algorithms can understand. A data scientist can produce more exact answers by combining structured, clean data with an applied machine learning model. Image sizing, color correction, data organization, data cleansing, and sampling are all parts of the process.

DATASET DIVISION

To use a dataset for machine learning, it must be divided into three subsets: training, testing, and validation sets.

To train a model and find its ideal parameters, a data scientist needs a training set.

A test set is required to assess a trained model's generalizability, or its capacity to find patterns in fresh, untrained data following training on a training set. To avoid overfitting the model, which might cause poor generalization as discussed above, it is crucial to employ different subsets for training and testing.

MODEL EDUCATION

A data scientist can start building a model by putting the training data into an algorithm after pre-processing the acquired data and dividing it into training and testing sets. The algorithm analyzes the information and builds a predictive model that can locate a desired property in fresh information. The creation of a model is the goal of training.

EXAMINATION AND TESTING OF MODELS

This step's objective is to create the most straightforward model that can rapidly and accurately forecast a target value. This objective can be accomplished by a data scientist through model tuning, which entails tweaking model parameters to produce the optimal algorithm performance.

FUNCTIONALITY

Future enhancements will be simple to implement because to the program's design. In order to demand little maintenance, the project was designed. Open-source and simple to install, the programs used are. User-friendly installation and use are required for the final application.

RELIABILITY

Critical factors include dependability, fault tolerance, and recoverability. Regardless of the volume of user input and training data, the system is reliable.

USABILITY

The computer program is simple to use, comprehend, and learn. Users can input their source and destination information to receive algorithm-generated airline choices.

The system must address safety issues relating to integrity levels. The computer system is protected by a password.

SECURITY

The Windows firewall does not prevent access to some open ports, therefore they stay open.

ROBUSTNESS

Future enhancements will be simple to implement because to the program's design. In order to demand little maintenance, the project was designed. Open-source and simple to install, the programs used are. User-friendly installation and use are required for the final application.

REQUIREMENTS THAT ARE NOT FUNCTIONAL

System performance elements including resources, time intervals, transaction rates, throughput, and others are determined by non-functional needs.

MAINTAINABILITY

After downloading, maintaining the system is easy because it doesn't call for any particular steps. Only when the user is made aware of an update is it necessary. One of the most useful elements of this idea is how simple maintenance is.

PORTABILITY

The program should be simple to install in a different setting and portable on a standard PC. From the place where the system was installed, the user has access to the computer.

PERFORMANCE

Less time is spent on algorithmic airline recommendations.

ACCURACY

Our work outperforms all other existing models in terms of accuracy.

Software Evaluation Face detection and identification are the project's main areas of focus. We used Python 2.6 for our implementation. Installing the necessary libraries is necessary before starting the project. We set up Pandas, Tensorflow, Keras, etc.

Hardware and software requirements are part of the system requirements, and they are listed below.

HARDWARE REQUIREMENTS

Any processor with a frequency of 500 MHz or above.

Ram : 4 GB

4 GB on a hard drive

standard keyboard and mouse for input.

High resolution monitor and VGA are the output devices.

SOFTWARE SPECIFICATIONS

Windows 10 or newer Operating System Programming: Python 3.6

Pandas, Flask, Matplotlib, Keras, and Tensorflow are a few libraries.

DESCRIPTION OF THE SOFTWARE

PYTHON

Python is a high-level, general-purpose programming language that may be interpreted. It was originally made available in 1991 and was designed by Guido van Rossum. Python places a strong focus on the readability of its code, which is accomplished in part via the usage of whitespace. Python may be used for programming projects of all kinds, from little to large, thanks to its adaptable collection of tools.

Python has an autonomous memory management system and a dynamic type system. It supports a wide range of programming paradigms, including imperative, functional, procedural, and object-oriented programming, and has a large and comprehensive standard library.

There are Python interpreters available for a variety of different operating systems. The majority of the various Python versions, including the default implementation, CPython, employ the collaborative development paradigm. The Python Software Foundation, a nonprofit organization, is in charge of overseeing CPython.

SCIKIT-LEARN

For data mining and analysis, Scikit-learn is an easy-to-use yet effective tool. For usage with Python, it is a free machine-learning library. Many algorithms, including support vector machines, random forests, gradient boosting, k-means, and DBSCAN, are available in Scikit-learn for classification, regression, and clustering. The scientific and numerical libraries for Python, such as NumPy and SciPy, are also effortlessly compatible with it.

Scikit-learn provides a selection of well-liked model categories, including:

Clustering methods like KMeans can be used to organize unsupervised data.

Cross Validation: It is a technique for assessing how well supervised models perform on fresh, untested data.

Datasets: to produce datasets with specific properties for the investigation of model behavior and to make use of them as test datasets.

By reducing the amount of features in a dataset, the technique of dimensionality reduction seeks to streamline data and make it simpler to summarize, visualize, and identify key traits.

Predictions from different supervised models can be combined using ensemble

methods.

For defining properties in picture and text data, use feature extraction.

The right attributes that can be used to build supervised models can be found through feature

selection.

To maximize the performance of supervised models, use parameter tuning.

To represent and simplify complex multi-dimensional data, manifold learning can be used.

Generalized linear models, discriminant analysis, naive Bayes, lazy learning, neural networks, support vector machines, and decision trees are only a few examples of the numerous supervised model types that are accessible.

POSSIBILITY ANALYSIS

A business proposal that comprises a basic project plan and cost estimates is submitted during the project's initial stage of feasibility assessment. An evaluation of the proposed system's viability is done to make sure it won't burden the company by doing system analysis. When doing a feasibility analysis, it's critical to understand the fundamental system needs.

TECHNICAL POSSIBILITIES

This study's objective is to evaluate the system's technical feasibility or requirements. The system must not place a considerable strain on the available technological resources because doing so could cause these resources to become overworked and place a lot of demands on the client. To prevent this, the developed system should require few or no modifications during implementation, putting little strain on the available resources.

FINANCIAL VIABILITY

This study's goal is to evaluate the costs and benefits of the system's implementation in the organization. The expenses must be backed up by reliable evidence because the firm can only invest a finite amount of money in the creation and research of the system. The majority of the technology used is derived from the public domain in order to ensure that the system is constructed within the budget, with only the specialist goods necessitating acquisition.

SOCIAL REALISTICISM

The study's objective is to determine the degree to which users have accepted the system, which entails assessing how simple and effective the user's instructions are to follow. Instead of intimidating or overwhelming the user, the system should be seen as a useful tool.

CHAPTER 4

SYSTEM DESIGN

IMPLEMENTATION METHODOLOGY

The suggested work is implemented in Python 3.6.4 using the necessary libraries, including OpenCV, Keras, Tensorflow, Scikit-Learn, and matplotlib. Dataset was acquired fromkaggle.com. Two classes of images—one with a mask and the other without—are included in thedata that was downloaded. MobileNet V2 network, which is based on CNN, uses a deep learning algorithm. This chapter provides an overview of the design of the architecture, the dataset utilized for implementation, the algorithm, and UML designs.

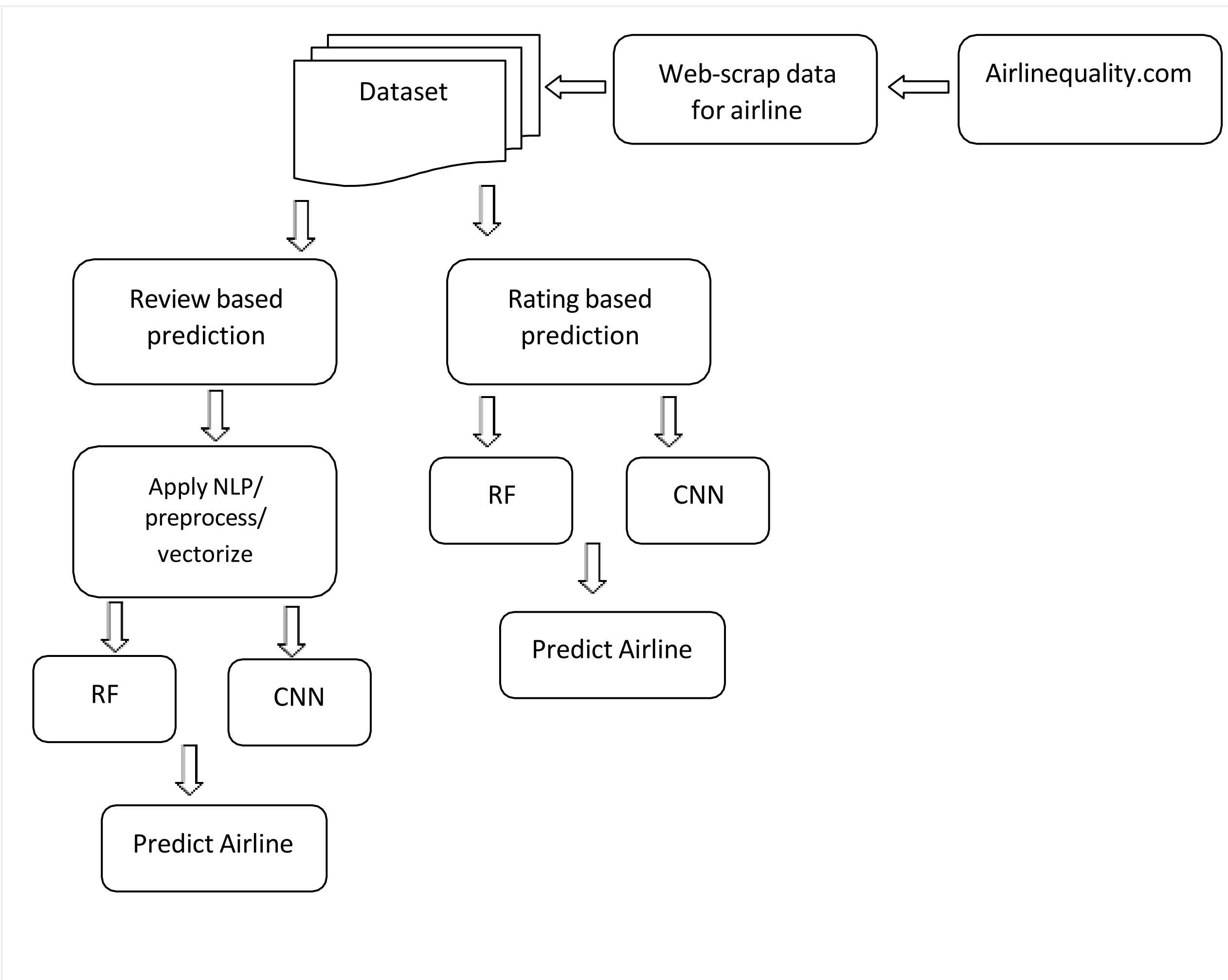


Figure: System Architecture

USE CASE DIAGRAM

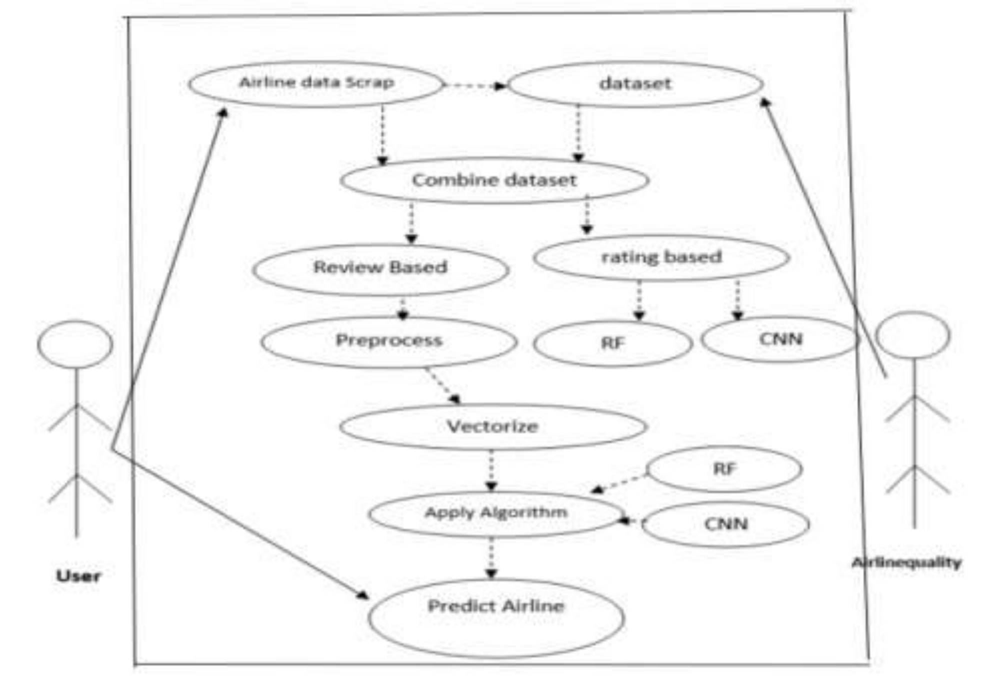


Figure: Use case Diagram

The use-case diagram shown above shows how tweets from the pre-processed Airline dataset are treated as separate train and test sets. The use of feature extraction algorithms to pre-process data. The problem of airline prediction is examined using the extracted features.

SEQUENCE DIAGRAM

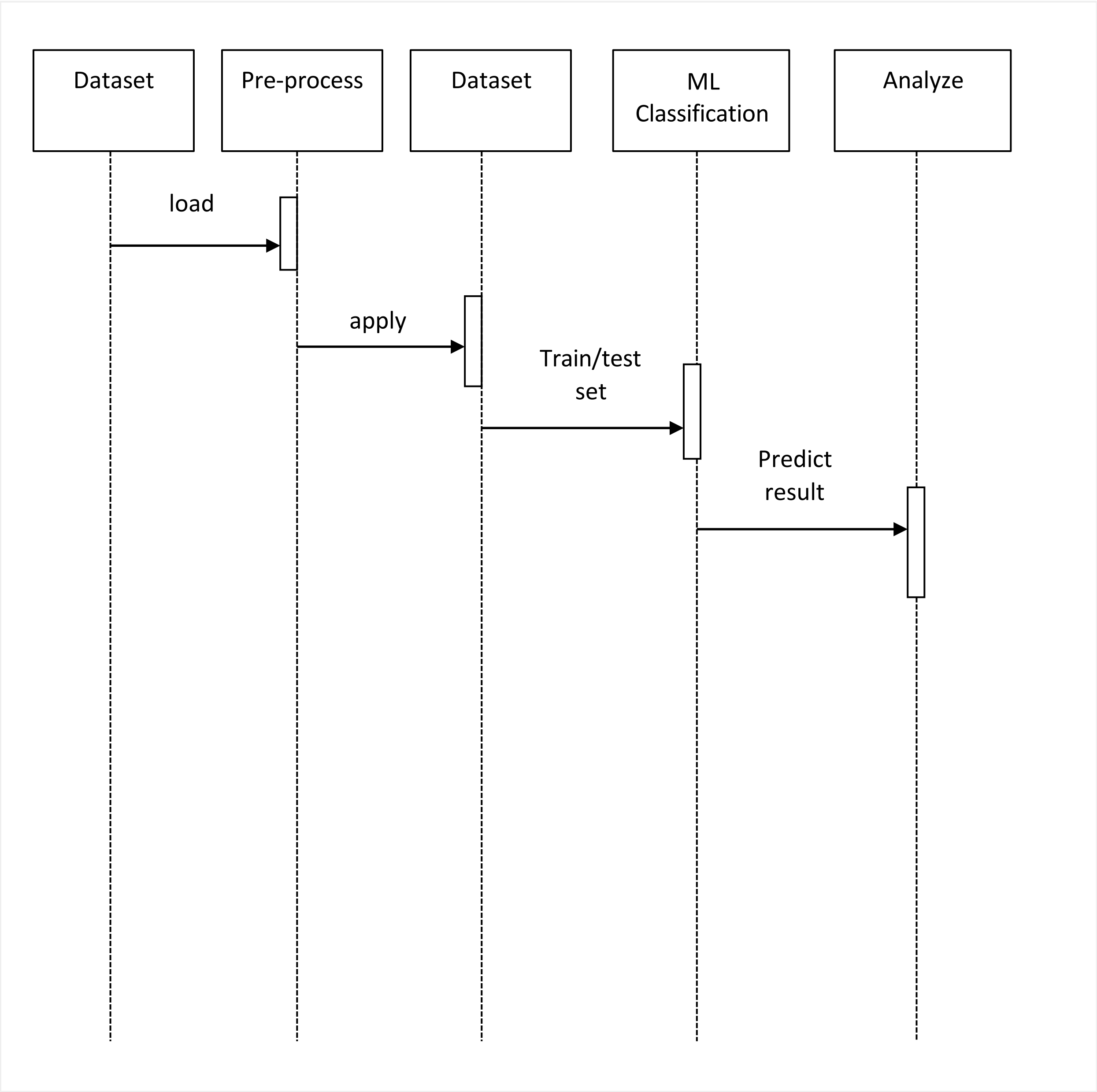
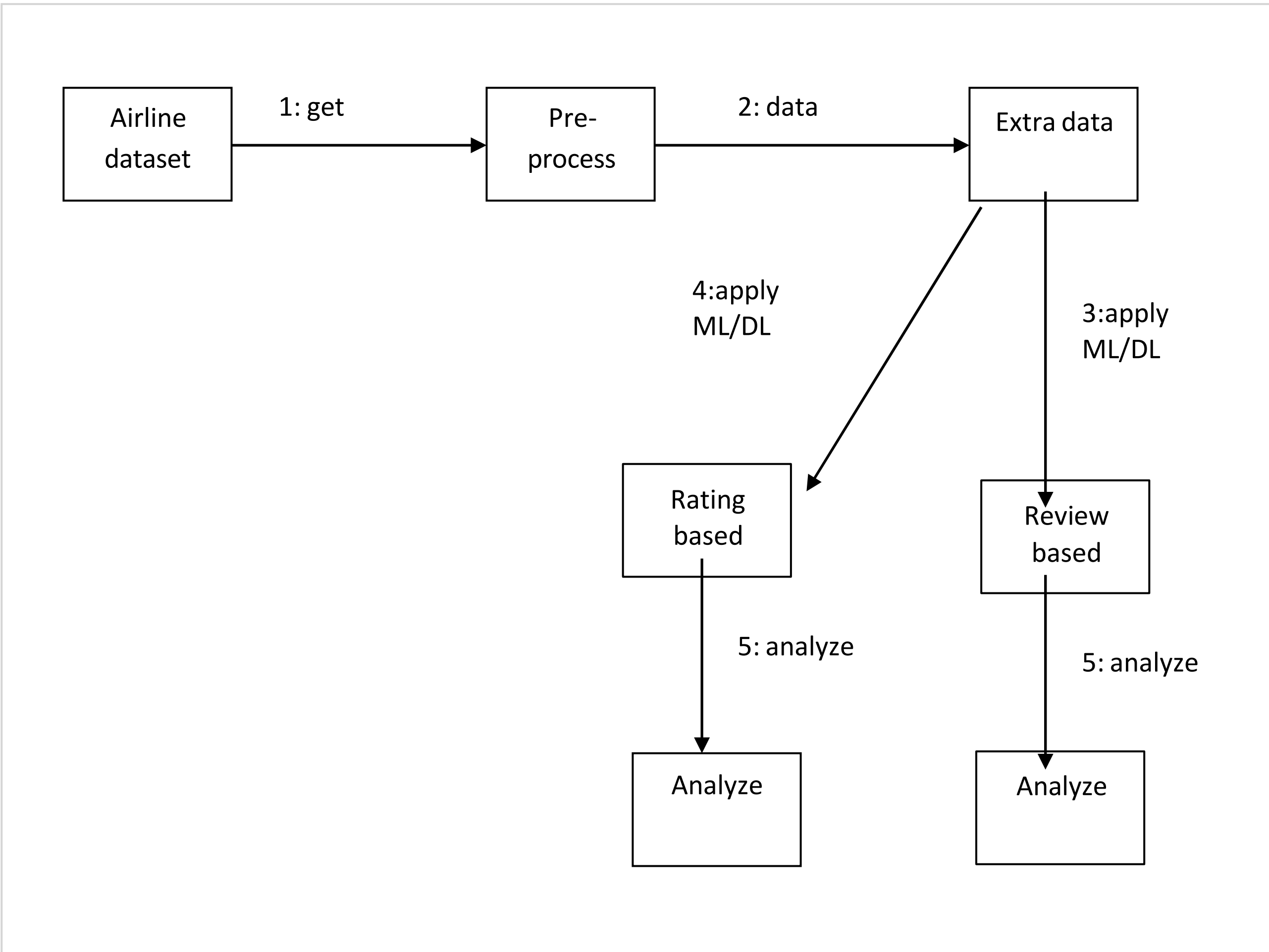


Figure: Sequence diagram

A sequence diagram shows the various processes that were implemented as vertical lines and the messages that were passed between the processes as horizontal arrows, in the order that they were implemented. The suggested system's data flow sequence is shown in the sequence diagramin the aforementioned figure.

COLLABORATION DIAGRAM



The collaboration diagram for the suggested system can be seen in the picture above. We used a sequence number to describe the interaction between the actor and function modules.

ACTIVITY DIAGRAM

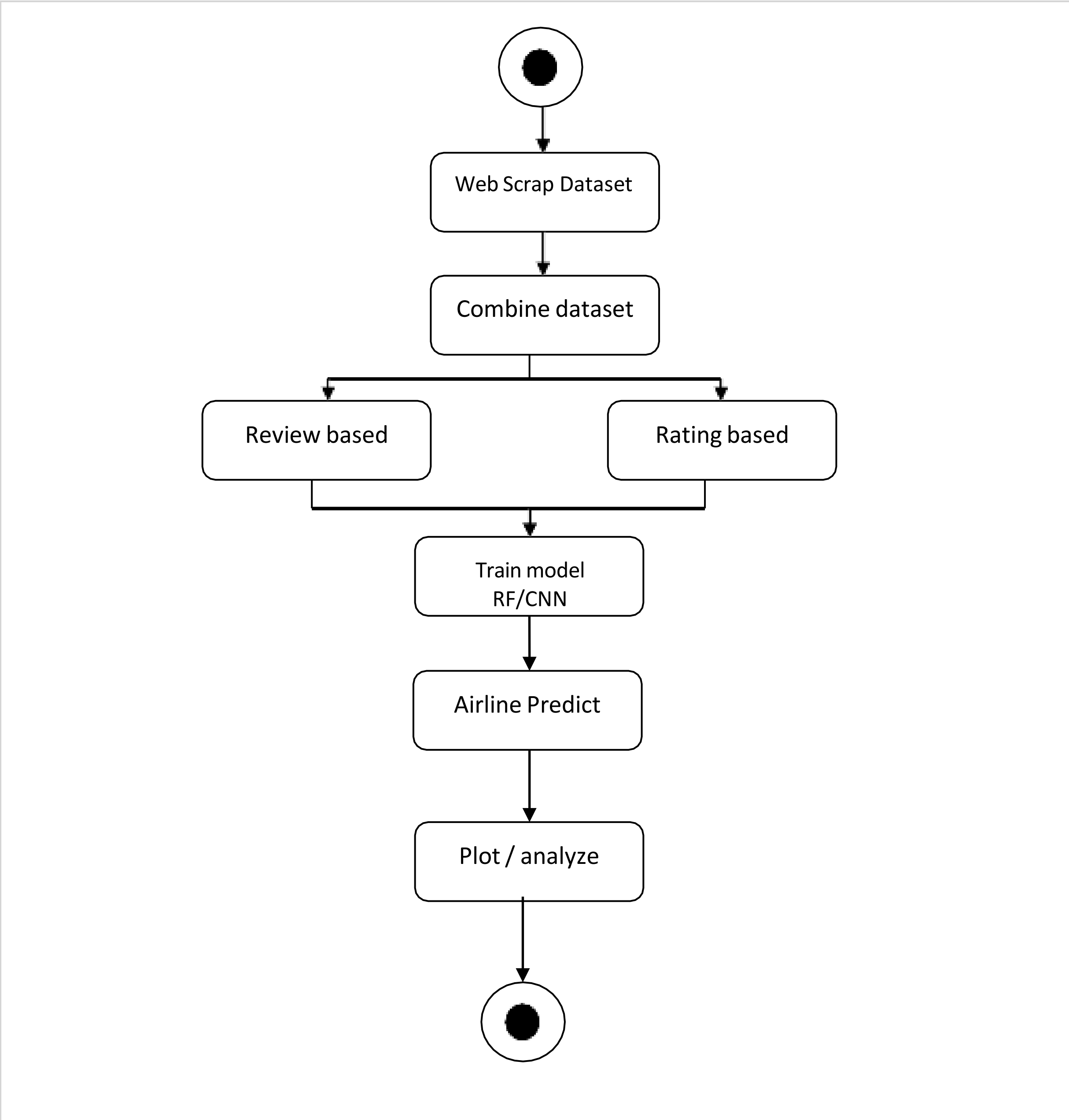


Figure: Activity Diagram

Workflows are graphically represented by activity diagrams, which show sequential actions with choice and iteration. Activity diagrams for airline forecast based on review and rating procedure are shown in the above picture.

DEPLOYMENT DIAGRAM

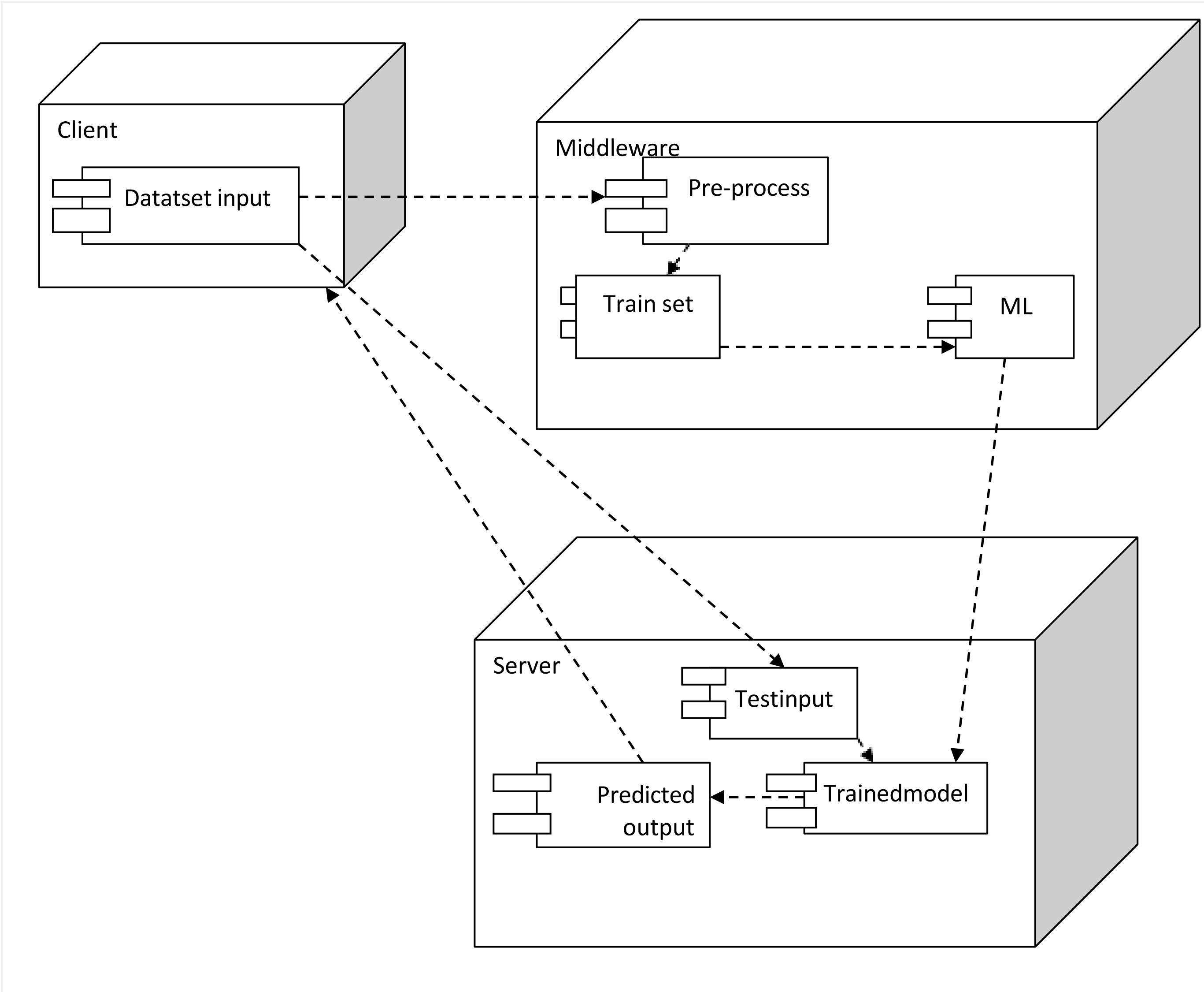


Figure: Deployment Diagram

The physical deployment of artifacts on nodes is modeled by the UML in the deployment diagram. The artifacts assigned to each node appear as rectangles inside the boxes that house the nodes. Sub-nodes, which look like nested boxes, can exist within nodes. In a deploymentdiagram, one node may conceptually represent a cluster of database servers or another set of several physical nodes.

CHAPTER 5

IMPLEMENTATION

MODULES

The modules included in our implementation are as follows

- Dataset collection
- Review based airline prediction
- Rating based airline prediction
- Predict airline based on source and destination

Dataset Collection

More than 3800 entries from the live airline dataset on airlinequality.com are web scraped in real time. Title, review value, reviewer name, review date, review text, aircraft, traveler type, seat type, route, date of flight, seat comfort rating, cabin crew, food and beverage rating, in-flight entertainment rating, ground service rating, value for money rating, recommendation, source and destination, and airlines are among the attributes included in the dataset.

Review based airline prediction

Text from customer reviews is used to forecast flight prices. Text input is handled by a vectorization model. The text data is transformed into vector input for the method using term frequency-inverse document frequency (TF-IDF). Recommendation is the target or prediction variable. The value has been set as a binary value of the type yes or no.

The TF-IDF technique is used to analyze review text in the following bit of code.

```
vectorizer = TfidfVectorizer(use_idf=True)
x_dataset = vectorizer.fit_transform(reviews['review_text'].values.astype('U')).toarray()
y = reviews['recommendation']
```

The Random Forest and Convolutional Neural Network algorithms are the machine learning and deep learning models used for this project.

Rating based airline prediction

Reviews, seat comfort, cabin attendant service, food and beverage quality, in-flight entertainment quality, ground service quality, and value for money quality are all used aslearning criteria in this module. Recommendation is the prediction characteristic. The snippet of code is provided below. Training and test data make up 80% and 20%, respectively, of the dataset.

```
x_dataset=reviews.drop(columns=['title', 'reviewer_name','date_of_review','review_text','aircraft','traveller_type','seat_type',

reviews['recommendation'] = reviews['recommendation'].replace(['no','yes'],[0,1])
y=reviews['recommendation']
print(x_dataset)

x_train, x_test, y_train, y_test = train_test_split(x_dataset, y, test_size=0.2, random_state=42)
```

Algorithm Used: Random forest and Convolutional Neural Network

Random Forest Model

1. Assuming the training dataset has n cases. Sub-samples are selected at random with replacement from these n examples. The individual trees are constructed using these arbitrary sub-samples taken from the training dataset.

2. With the assumption that there are k input variables, a number m is selected such that m k. At each node, m variables are chosen at random from a pool of k variables. The best split among these m variables is selected to split the node. While the forest expands, the value of m remains constant.

3. Each tree is allowed to grow as big as it can without being pruned.

4. The class of the new item is projected using the total votes cast by all voters.4. Based on the results of the combination of all the decision trees, the class of the new object is projected.

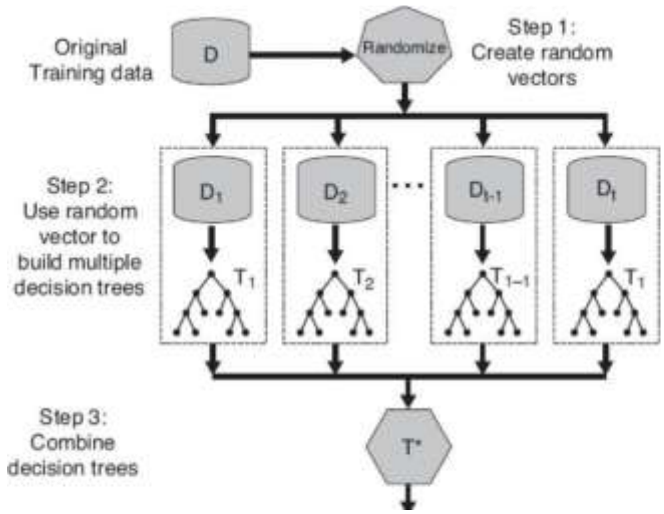
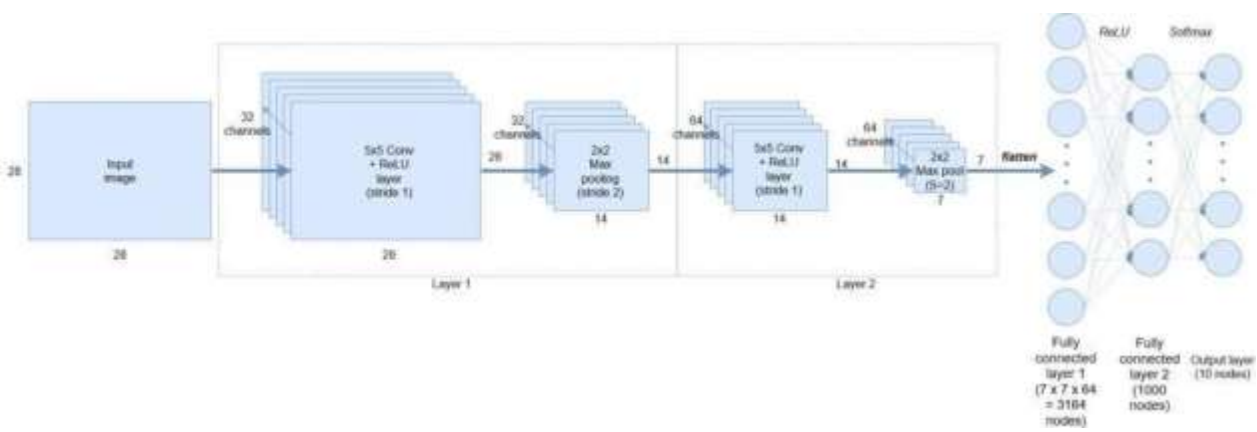


Figure: Flow chart of Random Forest

TRAINING USING CONVOLUTIONAL 2D NEURAL NETWORK

For training and testing our model, we used a convolutional 2F neural network that is available in Keras. Below is a diagram illustrating Conv2D's overall architecture.



Sequential Model

In Keras, models can be obtained in one of two ways: sequentially or through the Functional API. The Sequential model is probably applicable to the majority of deep learning networks. It enables simple stacking of the network's sequential layers (and even recurrent layers) from input to output.

The model type is declared as Sequential() in the first line.

Adding 2D Convolutional layer

A 2D convolutional layer can be added to process 2D input images. The initial input to the Conv2D() layer method is the quantity of output channels, which in this example is 32. The next input is the kernel_size, which in this case is a 5x5 moving window, and then the x and y strides of 1. The size of the input for the layer must be given to the model, and the activation function is a rectified linear unit. As Keras can determine the tensor sizes going through the model from there, only the first layer has to declare the input shape.

Adding 2D max pooling layer

Add a layer of 2D Max pooling. Simply put, we define the strides and the size of the pooling in the x and y directions, which in this case is (2, 2).

Adding another convolutional + max pooling layer

Another convolutional + maximum pooling layer with 64 output channels is then added. We may omit the strides argument because it is the default value in Keras for the Conv2D() function, which is (1, 1). In Keras, the strides argument is set to the pool size by default.The picture size for this layer is 28 x 28 pixels, and the number of output channels from the layer before is 32.The input tensor for this layer is (batch_size, 28, 28, 32).

Flatten and adding dense layer

The output from these must then be flattened before entering our fully connected layers. The following two lines declare our fully connected layers. Using Keras' Dense() layer, we describe their size as 1000 nodes, each of which is triggered by a ReLU function. The second is the magnitude of the number of our classes as determined by our soft-max classification, also known as the output layer.

Training neural network

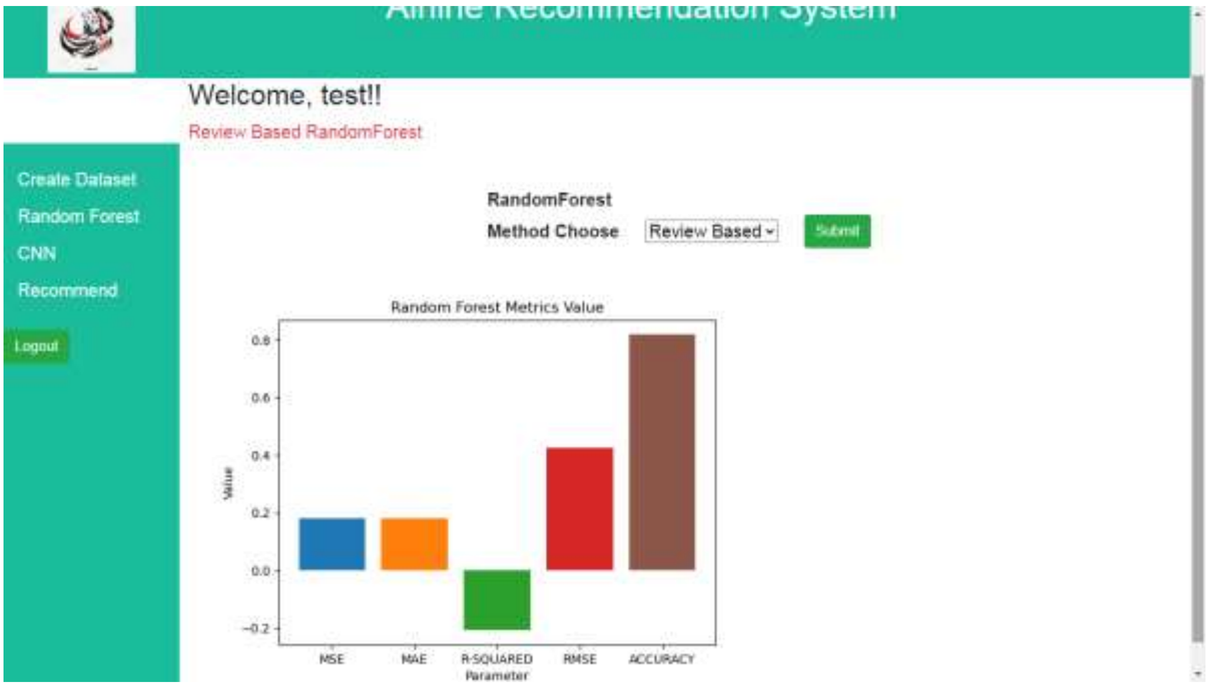
We must define the loss function or tell the framework which sort of optimizer to employ in the training model (such as gradient descent, Adam optimiser, etc.).

For categorical class classification, use the Lass function of the standard cross entropy (keras.losses.categorical_crossentropy). We employ the Keras.Optimizers.Adam optimizer. Last but not least, we have the option to specify a metric that will be computed when the model is evaluated.

All of our training data, in this case x_train and y_train, are initially passed in. The batch size is the following input. In this instance, the batch size is 32. The quantity of training epochs, in this case 2, is passed next. The verbose flag, which is set to 1 in this case, indicates whether you want certain information written in the console concerning the progress of the training.

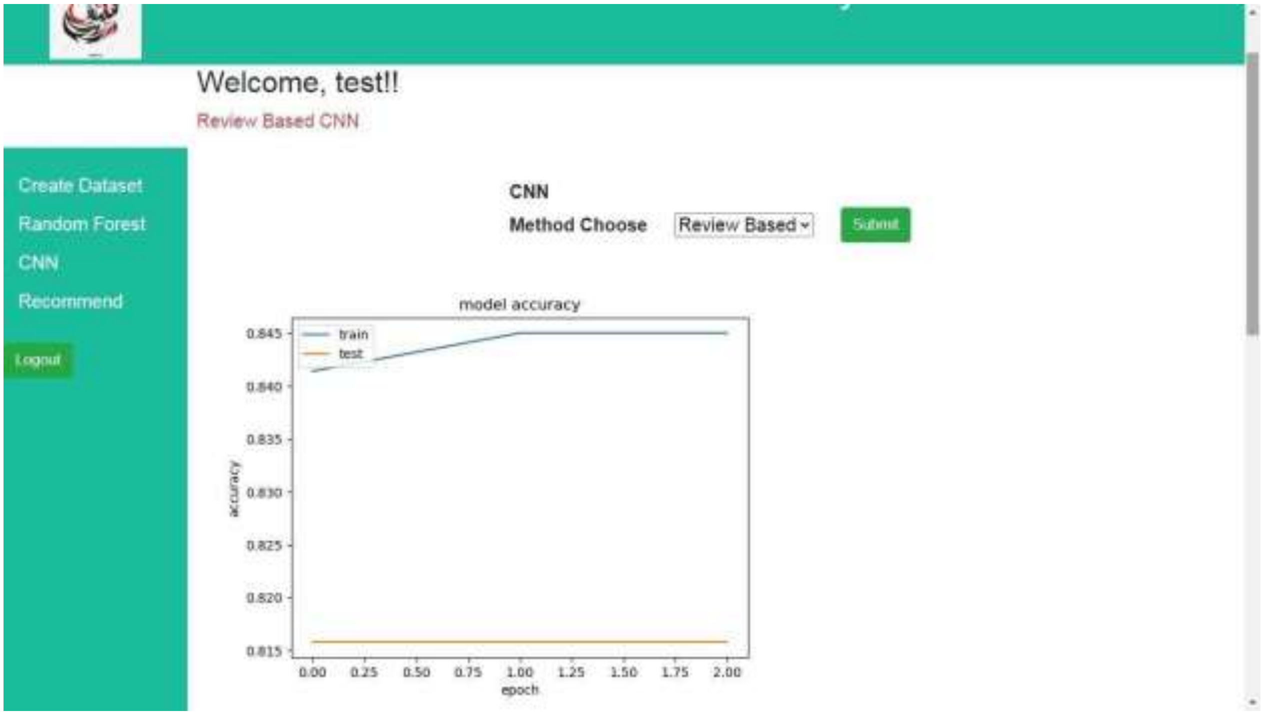
SCREEN SHOTS

The following screen shows the prediction of airline based on customer review based on random forest algorithm

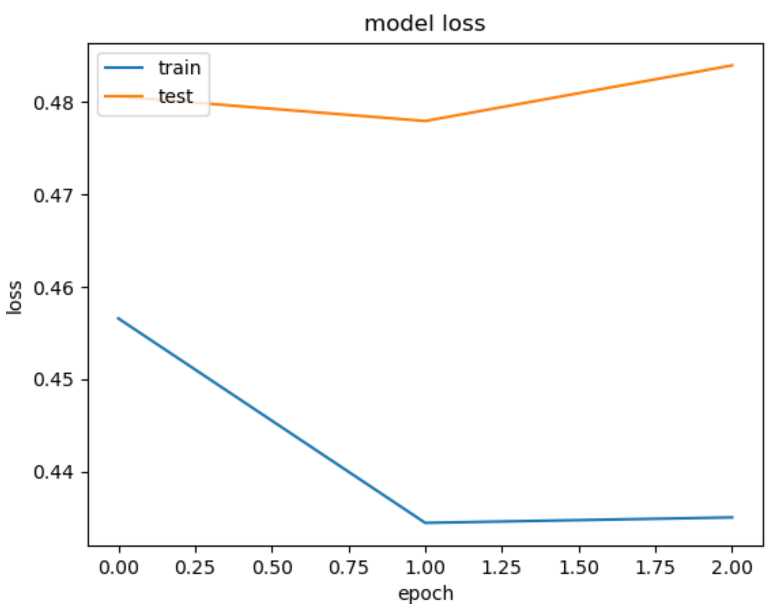


The following screen shows the prediction of airline based on customer rating based on random forest algorithm

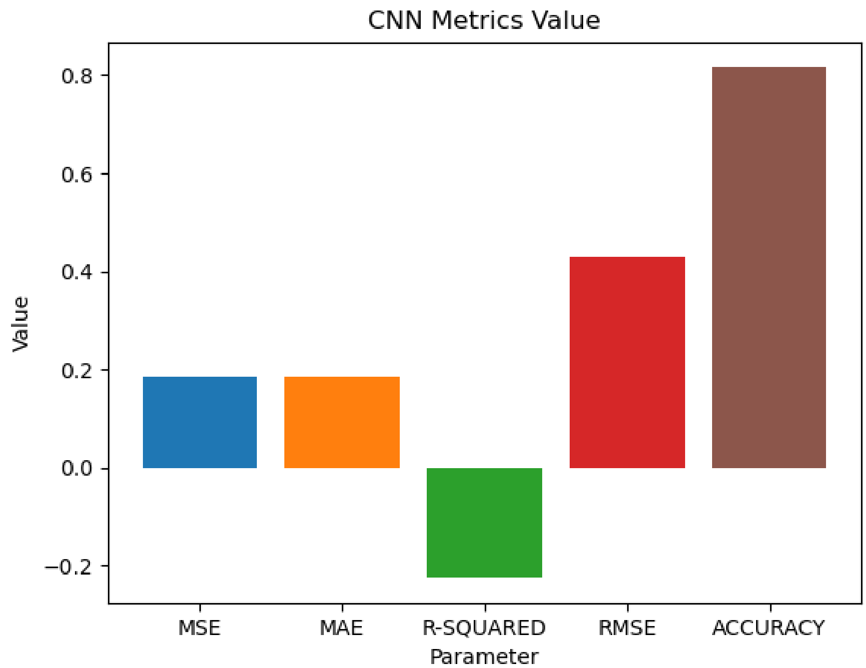
The following screen shows the accuracy of airline prediction based on customer review based on CNN algorithm



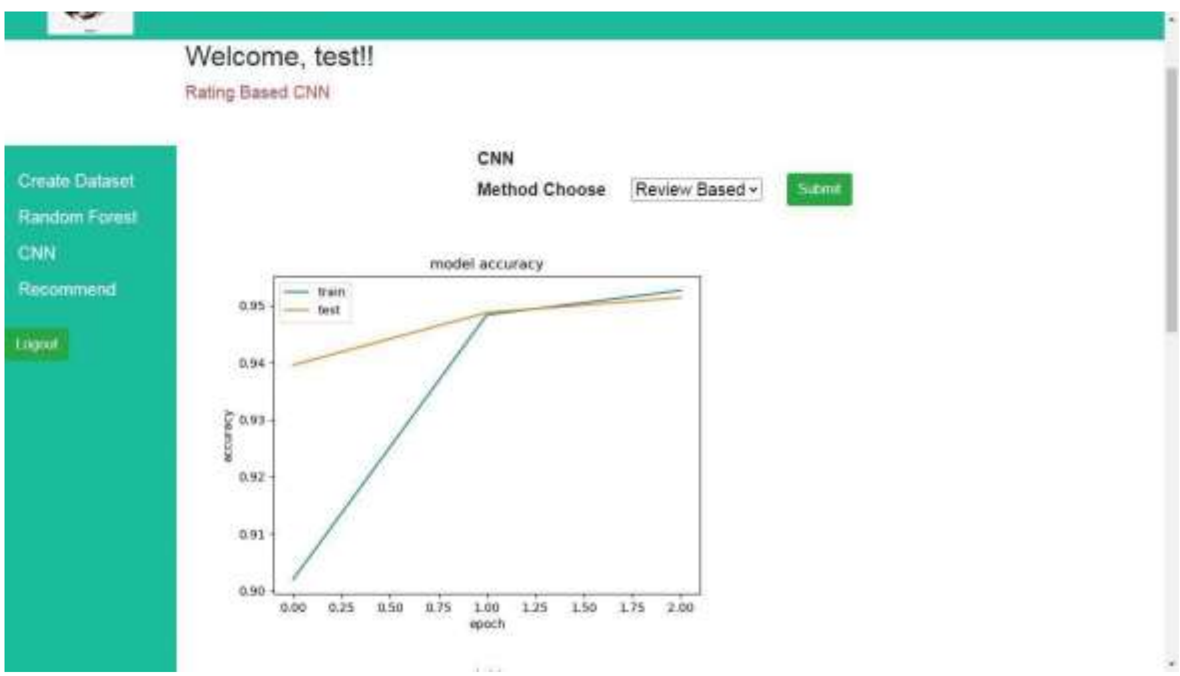
The following screen shows the loss metrics of airline prediction based on customer review based on CNN algorithm



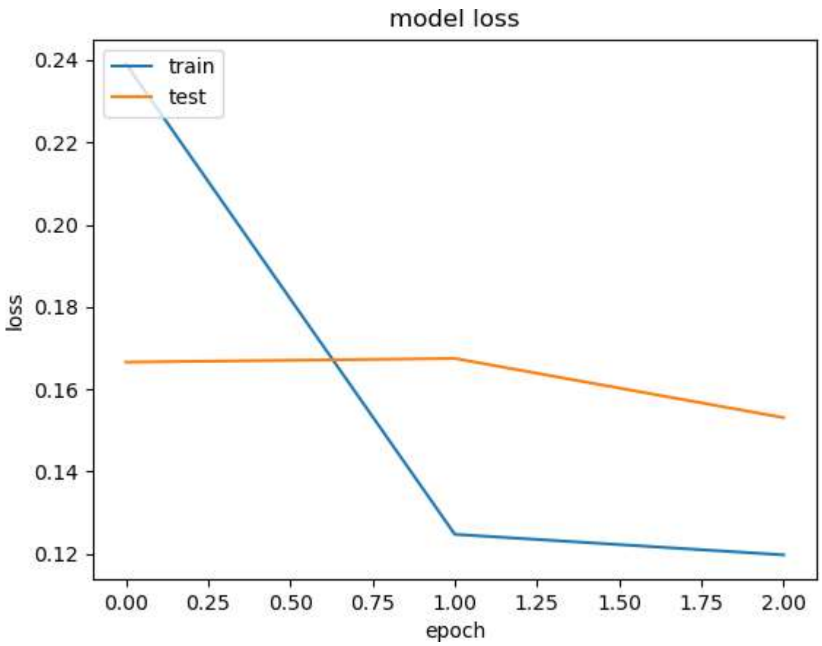
The following screen shows the CNN metric of airline prediction based on customer review



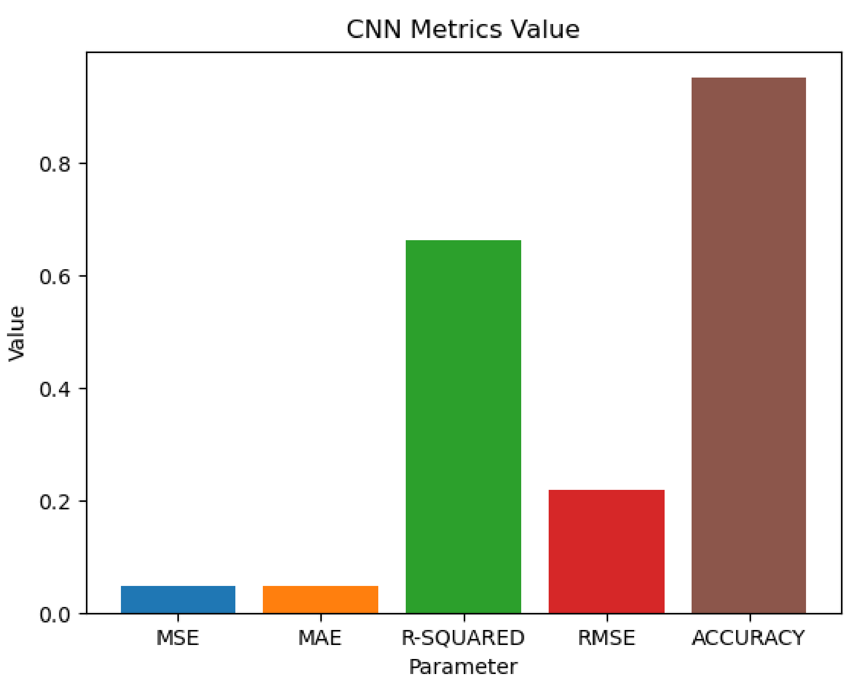
The following screen shows the accuracy of airline prediction based on customer rating based on CNN algorithm



The following screen shows the loss metrics of airline prediction based on customer review based on CNN algorithm



The following screen shows the CNN metric of airline prediction based on customer review



Recommended airline based on given source and destination

FUTURE SCOPE AND CONCLUSION

Further research might look into the factors influencing the preferences of travelers in business and economy classes, and using predictive analysis on customer review data could provide useful insights for spotting trends and areas for development. Businesses can use the research to guide the creation of their goods and services as well as their marketing plans, and online review sites may be able to offer consumers tailored recommendations. All things considered, the growing reliance on Internet sources for decision-making gives an opportunity for organizations to better understand and adapt to client preferences, with the potential for additional research and applications to increase customer happiness and competitiveness.

REFERENCES

[1] Flight Delay Prediction Using Machine Learning and Big Data in Aviation, IEEE Transactions on Vehicular Technology, vol. 69, no. 1, pp. 140–150, Jan. 2020; doi: 10.1109/TVT.2019.2954094; G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou, and D. Zhao.

[2] A New Method: Machine Learning-Based Prediction of Airline Delays, 2018 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 1081–1086. V. Natarajan, S. Meenakshisundaram, G. Balasubramanian, and S. Sinha.

[3] Machine learning for cost-sensitive airline delay prediction, 2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC), S. Choi, Y. J. Kim, S. Briceno, and D. Mavris, pp. 1–8, doi: 10.1109/DASC.2017.8102035.

[4] Review of Automatic Ticket Price and Demand Detection for Airlines, 2018 International Conference on Innovations in Information Technology (IIT), J. A. Abdella, N. Zaki, and K. Shuaib, pp. 169–174, doi: 10.1109/INNOVATIONS.2018.8606022.

[5] Using machine learning methods, predict flight delays caused by the weather, 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), pp. 1-6, doi: 10.1109/DASC.2016.7777956, S. Choi, Y. J. Kim, S. Briceno, and D. Mavris.