INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 6.078
(Volume 6, Issue 3)
Available online at: www.ijariit.com

Air traffic Control of US domestic flights

Prachi Dhariwal
prachidhariwal24@gmail.com
Bharati Vidyapeeth College of Engineering, New Delhi, Delhi

Preeti Verma
cse.preeti20@gmail.com
Bharati Vidyapeeth College of Engineering, New Delhi, Delhi

Anusha Ramachandran
anusha98iyer@gmail.com
Bharati Vidyapeeth College of Engineering, New Delhi, Delhi

Parul Yadav
parulyadav0111@gmail.com
Bharati Vidyapeeth College of Engineering, New Delhi, Delhi

ABSTRACT

With the increase in data and especially air traffic data, it is increasingly becoming difficult to handle the same. It is important not just to handle the data to get the necessary output, but also to maintain the integrity of the data. Hadoop is a framework which we have used as it provides reliable services for the data in concern. MapReduce can be used to provide apt management of large amount of data and it can be used as a filter to generate useful insights. Machine Learning on the other hand is used for the recommendation portion of the project. Machine Learning helps in identifying the patterns in the data which might not be visible to the naked eye. It then is used to make useful predictions and correct recommendations based on the data in consideration. For instance, a machine learning program may successfully pinpoint a causal relationship between two events. This makes technology highly effective at data mining. Recommender systems aim at suggesting relevant items to users to support them in various decision- making processes, on the basis of available information on items or users. In the latter, the customer's interests and tastes can be learnt and expressed using historical browsing data, purchase histories, and even other nontraditional data sources such as social networks. Despite its proven success in the online retailing industry, in electronic commerce and, even tourism, recommender systems have been less popular in-flight itinerary selection processes.

Keywords— *Air Traffic Control (ATC), Air Traffic Management (ATM), Hadoop, Hadoop Distributed File System (HDFS), Big Data, MapReduce, Machine Learning, Python, Recommendation system.*

1. INTRODUCTION

1.1 Congestion Control System

Traffic congestion is the problem we are dealing with, in this project. There can be congestion ranging from areas of air traffic to freight carriers. Alongside this, the problems faced in the current infrastructure, due to the constructions of roads, runways, etc. are added. So, there are a lot of risks involved in trying to maintain the stability of the traffic and henceforth, controlling the air congestion. Thus, we have tried to make an application which will help passengers choose wisely and also the controllers to maintain stability in the system with as much ease as possible.

The enhanced ability of the system to help control network traffic is additional. The METAR data is used to help monitor the real time air traffic and thus to give accurate results. This tracked information aids in the prediction of the traffic density and then the network working is enhanced. This renders valuable information for controlling traffic flow, prediction of congestion and reducing the number of accidents in that network.

The development of knowledge management has influenced many areas. However, there are two opportunities to scientific community: how lead with an amount of data so big in real-time and achieve useful results; and with Big Data available how improve the real-time decision support systems using historical information. There has been an increase in unstructured data off late. To find which of them is useful and insightful is the task of the government. This huge amount of data created is called Big data and an ATM environment has abundance of it, which is studied with focus on the following subjects:

(a) Digital data which includes communication information between the radio staff and the pilots, the flight schedules, delays, details of passengers, etc.
(b) ATM utilizes information from external sources as well. These data include monitor images, METAR data (meteorological data), GPS guidance data and many others.

1.2 Recommendation Systems

Recommendation systems have been evolving in the recent years. While people used to associate it with online shopping

portals, now it can be applied to anything ranging from webpages to blogs, search engines and even websites. The two main approaches to build these recommendation systems are:

1.2.1 Collaborative Filtering: CF maintains a database of ratings given by several users for several products or flights (in this case). It then looks out for ratings which are similar to other ratings and clubs them together. When a new rating is entered, it is clubbed with the rating with which it is strongly related. When a user asks for the rating or recommendation for a particular flight, the one with which is rated highly by several users is recommended but not by the current user himself. Collaborative Filtering is very extensively used by many companies and to build such a system, one needs to use not only the user data but also the product data.

1.2.2 Content Based: Content based systems, as the name suggests, uses data only from the products or items in concern. CB looks into the features of the items, and then recommends an item with similar features to the one which is being searched. The content of these products can be classified under structured or unstructured. ATM data is usually highly unstructured, and unlike structured data, these data can be difficult to handle.

CF systems can again be classified further into three broad heads:

(a) Active CF: In an Active CF system, the users actively are involved in the process of building the recommendation system by giving direct ratings or direct feedbacks on a particular flight. The profile data of these users might also be needed in these Active CF systems.

(b) Inactive CF: These systems do not use an users' profile data, instead just analyse their activities like particular flight choice, frequency of booking, access nature, etc. While it might seem that the data collected from Inactive CF is not very useful, but it is. It helps the system understand the nature of access of the user and makes a recommendation based on that.

(c) Item Based CF: It is a system which is almost like a mix of the Active and Inactive types. This system gives a direct recommendation based on the data of similar products.

(d) Item to item-based recommendation can have its own set of advantages and disadvantages, in fact, all the types of filtering have their pros and cons.

2. SCRUTINY OF WORK

The Big data technology provides a newer technical approach to deal with the problems mentioned above. Advantages of applying Big data are:

The large quantity of data can be easily managed by Congestion management system with big data technology. Big data resolved three major problems: data storage, data analysis, and data management. Hadoop have the default property of handling massive amounts of data where data is segmented and is stored at different nodes. A huge task is divided into numerous tiny tasks and is processed in a MapReduce model. The system's stability and fault tolerance are seen at that time while handling data. Widely, the efficiency of transporting data can be sharpened easily by Big data. The transportation industry, including various aspects of work, need to deal with the Big data each day, need a more controlled mode of application and have a great deal of equipment. In the aspect of improving transport efficiency, improving the threshold capacity of the road network, adjusting traffic demands, big data technology contributes at a large scale.

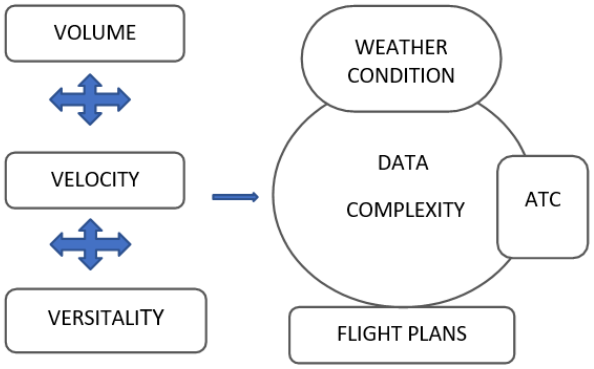

Fig. 1: Big Data Technology application

One of the most important concern of using Big Data in real-time situations is how fast one can have the final result. It is necessary to analyze the so big and not-structured data, that too from many data sources. Although, till now there is no accurate definition of Big data, it can be described formally for knowledge discovery in so big data structures. The other method to explain this definition it is a manner used by companies to define strategies and tools to structure, handle, analyze and present the achieved results, expected or not, which was ascertained from big data structures. The ramification of analyzing big data is based on three factors: capability i.e., volume, velocity, and versatility so that they can make a base for business specialists to continue decision process.

-Volume: The size of data.

-Velocity: The change of speed from the oldest data to the newest one.

-Versatility: The total measure of sources of data Executing the prospective job(work) at the very initial stage needs to make Data set for Airport Data as it accommodates a set of the table to such an expansion that Flight detail, user detail, payment mode, ticket info for convinced period.

Hadoop is the main contrivance of this paper which is used to complete the scheduled work. For meticulous data, Hive and MapReduce techniques are used. Assuming all the tools of Hadoop have been installed and having semi-structured information on airport data. The aloft revealed queries have to be engrossed in the expression as expected below. Accession used is as follows:

(a) Establish table with imperative properties

(b) Distillate half methodized data into the table using the load command

(c)Look over data for the coming queries:

- List of airports operating in the country
- List of airlines having zero stops
- Meteorological information
- Arrival-destination airports
- List of operating airlines in-country

3. GROUNDING AND RESEARCH

3.1 Consumer Recommendations

For associations, customer's assessment is a must, not only for dynamic business performance and growth but also for product and service notion as well as for improvising customer participation. Managers traditionally rely on customer feedback metrics inclusive of mean customer comfort, the proportion of customers complaining, repurchase likelihood of customers as well as word-of-mouth conversation such as the sum of endorsement or promoters. This proposition has received much analysis by academics and practitioners alike

over the past decade. Previous research has found that the rise in NPS is correlated with growth in business. In business, promoter or followers scores have been leveraged to gain perceptivity into the customer base and in turn to drive market share advancement. Thus, curriculum based on promoter scores is especially compelling in underscoring the global strategy of attending to customers and substantiating adjustments transparently by attributing them to admissible testimonies. Accustomed to the importance of customer engagement, loyalty, and feedback, as deliberated in promoter scores, it is imperative to understand the factors that influence positive word-of-mouth too, i.e. service or product recommendations. As in today's world, consumers express their experiences via digital platforms, it is one of the essential factors for enterprises to monitor this form of electronic word-of-mouth so that they can understand consumers and identify the aspects influencing the recommendation decision.

Moreover, many businesses struggle to utilize online reviews to create business value. This is due to most reviews are not directly tied to a service or product recommendation score, i.e., the issue of whether a customer recommends the service to another potential customer. Thus, a research space exists in such an area of automatically extracting propaganda from online reviews to derive indirect recommendation intentions accurately, disentangling the recommendation decision into various service aspects expressed, and to ultimately incorporate them within a promoter score.

Notably, promoter scores based on consumer recommendations are different from a typical online recommendations including the major variable of interest: Whereas the research stream on online recommendations focuses on the question of how a particular product or service can be specified from a set of products or services and be recommended to a consumer, the research stream on consumer recommendations examined whether consumers recommend a specific product or service of interest. Whereas the number of studies in the field of product recommendation agents providing online recommendations does exist, there is a research break or block in the field of extracting consumer recommendations from online reviews.

3.2 Predictive Analysis

In evaluating the powerfulness of a review's text subjects to carefully counter a reviewer's recommendation to other travelers, we construct different predictive models and assign their performance. Towards that termination phase, we propose specific model configurations incorporating various attributes to predict the airline recommendation. Moreover, we assess different machine learning techniques. In the end, we validate our results with a recommended evaluation methodology in the machine learning field to assure that the results are realistic and not an ordnance of model overfitting.

3.2.1 Machine Learning Techniques: We evaluate the performance measure of various machine learning techniques in predicting the recommendation decision. Towards that goal, we perform supervised learning and learn from illustrated examples, i.e. the digital reviews and the indication of the reviewer expressing the airline recommendation. In this context, we majorly concentrate on Naïve Bayes (NB) as a rather simple learning algorithm as well as Neural Network (NN) and Support Vector Machine (SVM) representing more space and time consuming learning algorithms.

Naïve Bayes presents a very simple machine learning technology based on the Bayes theorem. Classifiers built on the Bayes theorem are assumed to be naïve as they assume the independence of the different input variables. In Naïve Bayes classifiers, instances are separated based on joint probabilities of their input values. However, Naïve Bayes classifiers are rather simple and rely on potentially imaginary assumptions, they have nevertheless been proven to normally perform well and have the merit of requiring less computational effort and thus being more time-efficient. Neural Networks consist of a variety of (computational) neurons appearing in interconnected input, hidden, and output layers, and are intended to copy the behavior of human neural networks. To achieve this aspect, weights are assigned to the connections between different neurons. Moreover, every single neuron has an activation function that is used to process the input value of the neuron.

The output neurons, as input, the weighted measure of outputs from neurons in the previous layer (or input variables in case of the initial input layer), and implies the activation method to the input. When a neural network is trained, the weights of the different neurons are updated so that the final or total neural network's output corresponds to the general classification. In this study, we apply a feed- forward neural network using reverse propagation. As activation function, we adopt the most commonly used function i.e., sigmoid function. Support Vector Machine represents another machine learning technique that is based on the principle of searching for the highest-margin hyperplane that maximizes the distances between instances of different classes. As the singular separation of observations is not always possible, transformations are conducted employing kernel functions which enable a separation of the observations according to the classes they were assigned.

4. LITERATURE REVIEW

4.1 Related works

The ATM environment can be classified into 3 different sectors:

- Air Space Management: ASM focuses on increasing the capability of aircraft in the airspace, to provide sufficient services for demand within the available structure.
- Air Traffic Control: ATC focuses on controlling the aircraft flight, providing essential information which preserves the clause of safety.
- Air Traffic Flow Management: ATFM focuses on providing information to maintain the air traffic flow with safety and a reduced impact on future ways of doing work.

5. RESULTS AND CONCLUSIONS

This paper underlines on data analysis on airline data set. The paper addresses the usage of modern analytical tool. Hive on Big Data set which focuses on general services and requirements of any airport. Some of the instances are marked and highlighted with the images.

Figure 1 represents the usage of big data in the air traffic controlling system. The paper also provides an introduction to Map Reduce techniques that are internally taken care of by the underlying tools of the Hadoop System. Sample queries that have been executed with Hive on Hadoop, will be shown. It is found that Hive is effective in terms of handling huge data sets when compared to traditional databases concerning time and data volume.

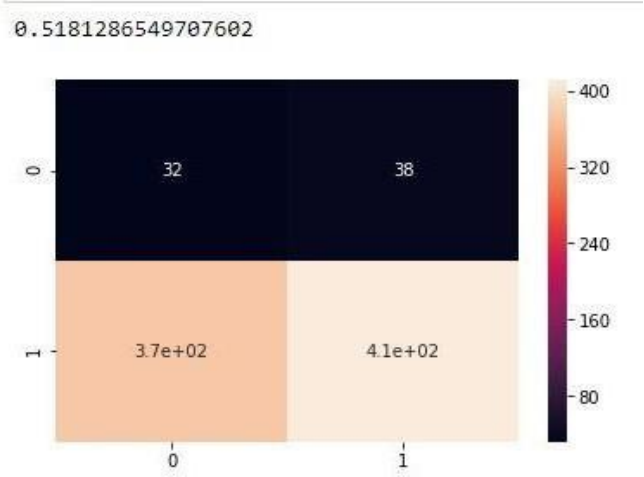

Fig. 2: Gaussian Naïve bayes

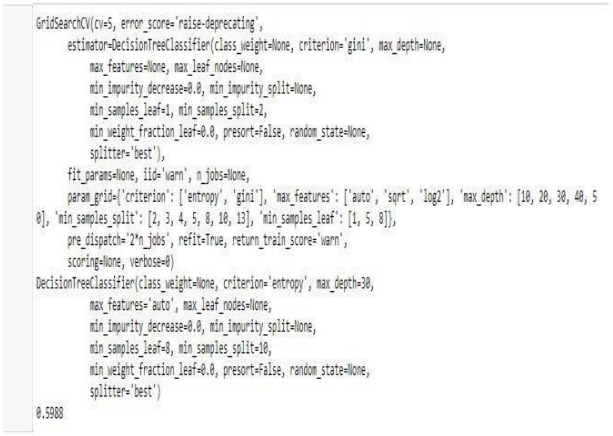

Fig. 3: Decision Tree Classifier

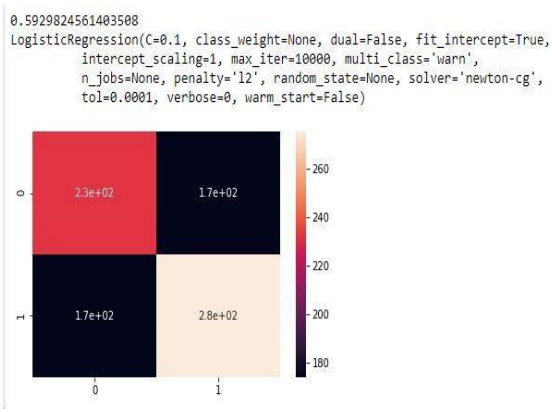

Fig. 4: Logistic Regression

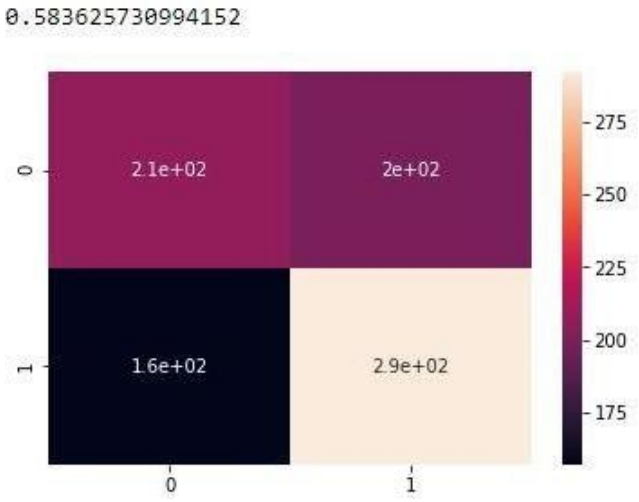

Fig. 5: Gradiant Booster Classifier

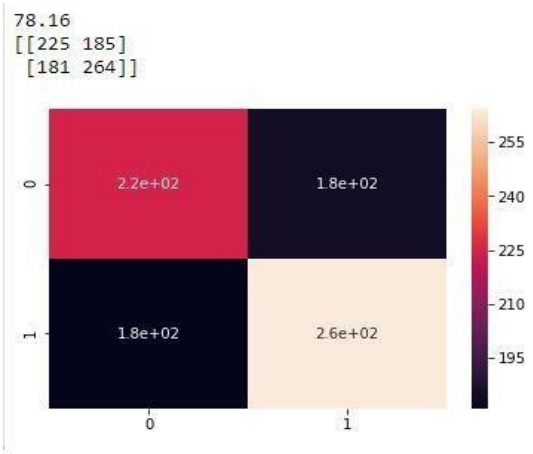

Fig. 6: Kmeans Classifier

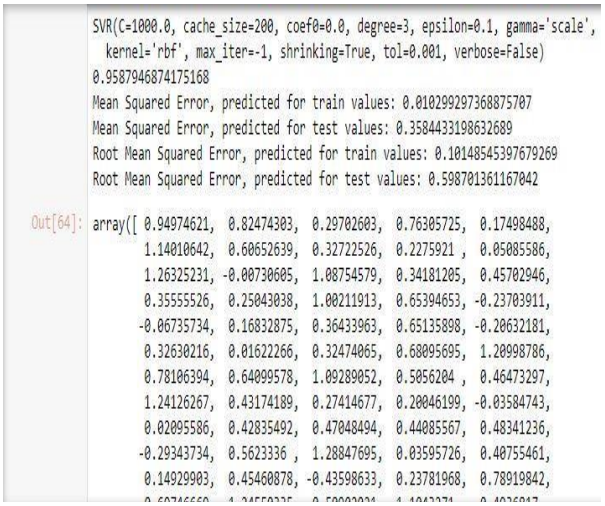

Fig. 7: Support Vector Regression

Fig. 8: Random Forest

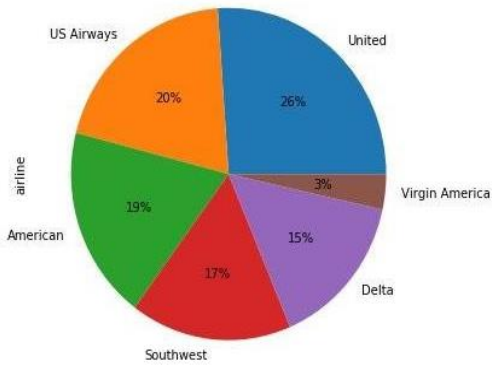

Fig. 9: Percentage distribution of airlines based on Social account presence

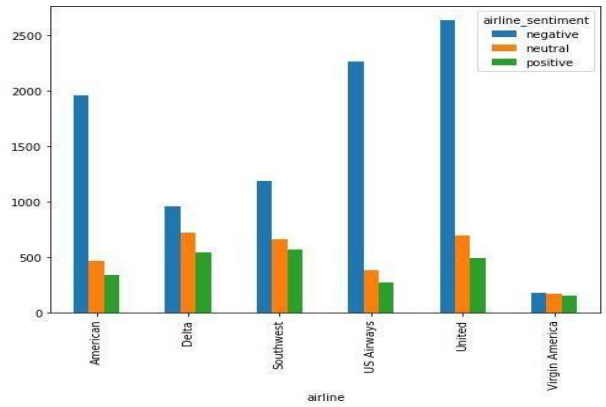

Fig. 10: The sentiments presented on social accounts with respect to the airlines

As authors of this paper, we have tried to filter out data which we need, using Hadoop. After that, we have implemented several algorithms, like explained and shown above, to get the accurate results for our data in concern. Lastly, we have tried to plot the different airlines with respect to their standings on their social accounts, just as an additional information for the reader and in order for them to make better choices. Air traffic control is here to stay and through this research, we have analyzed only US Domestic flights. But we strongly believe and hope to try the same with flights from all around, for people from all over the world to make better choices.

6. REFERENCES

[1] H.T. Rhee, S.-B. Yang, "How does hotel attribute importance vary among different travelers? An exploratory case study based on a conjoint analysis.

[2] M. Abdel-Aty, C. Lee, Y. Bai, X. Li, and M. Michalak. "Detecting periodic patterns of arrival delay", Journal of Air Transport Management, 13(6):355–361, Nov. 2007.

[3] K. F. Abdelghany, S. S. Shah, S. Raina, and A. F. Abdelghany, "A model for projecting flight delays during irregular operation conditions", Journal of Air Transport Management, 10(6):385–394, Nov. 2004.

[4] S. Ahmad Beygi, A. Cohn, Y. Guan, and P. Belobaba, "Analysis of the potential for delay propagation in passenger airline networks", Journal of Air Transport Management,14(5):221–236, Sept. 2008.

[5] F.F. Reichheld, "The one number you need to grow", Harvard Business Review 81 (12) (2003) 46–55.

[6] B.A. Sparks, H.E. Perkins, R. Buckley, "Online travel reviews as persuasive communication: the effects of content type, source, and certification logos on consumer behavior", Tourism Management 39 (2013) 1–9.

[7] Q. Ye, R. Law, B. Gu, W. Chen, "The Influence of user-generated content on travel behavior: an empirical investigation on the effects of e-word-of-mouth to hotel online bookings", Computers in Human Behavior (2011) 634–639.

