www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 7 July 2015, Page No. 13311-13318

Improving Classification Accuracy Using Weighted Multiple Regression

Sonali S. Salunke[1], Swati Deshpande[2]

[1] Department of Information Technology, Shah and Anchor Kutchhi Engineering College, Mahavir Education Trust Chowk, W.T.Patil Marg, Chembur, Mumbai-400088, India.
Sal.sonali21@gmail.com

[2] Department of Information Technology, Shah and Anchor Kutchhi Engineering College, Mahavir Education Trust Chowk, W.T.Patil Marg, Chembur, Mumbai-400088, India.
sakec.swati@gmail.com

Abstract: ***The purpose of this system is to find out if a flight is getting delayed during departure and arrival then what are the reasons for the delay. Therefore we intend to aid the airlines by predicting the delays by using certain data patterns from the previous information. This system explores what factors influence the occurrence of flight delay. Classification algorithm is applied to classify flights into delay categories. Using OneR, a classification algorithm, models are developed to predict delay on both arrival and departure side. Discretization was applied using Weka, a data mining tool, to divide the delays on departure and arrival side into five categories viz; Negligible, Insignificant, Nominal, Significant, Indefinite. The result thus obtained from these categories was further analyzed to predict overall reason for delay. The delay predicted can be due to Weather, Security, Carrier, National Aviation System (NAS) and Late Arrival. The models further combine this result with Meteorological Terminal Aviation Routine Weather Report (METAR) to give the report of weather conditions at origin and destination airport. The results of data analysis will suggest that flight delays follow certain patterns that distinguish them from on-time flights. We may also discover that fairly good predictions can be made on the basis on a few attributes. Classification can be used for analyzing future data trends. It is important that the classification is appropriate so that the data prediction is accurate.***

Keywords: Data Mining, Multiple Regression, Classification of nominal attributes**,** METAR, FAA**.**

1. Introduction

A flight delay is a when an airline flight takes off and/or lands later than its scheduled time. The **Federal Aviation Administration** (FAA) considers a flight to be delayed when it is 15 minutes later than its scheduled time. A cancellation occurs when the airline does not operate the flight at all for a certain reason. Some of the causes of flight delays are as follows:

- Maintenance problems with the aircraft.
- Fueling.
- Extreme weather, such as tornado, hurricane, or blizzard.
- Airline glitches.
- Congestion in air traffic.
- Late arrival of the aircraft to be used for the flight from a previous flight.
- Security issues.

Flight delays are an inconvenience to passengers. A delayed flight can be costly to passengers by making them late to their personal scheduled events. A passenger who is delayed on a multi-plane trip could miss a connecting flight. Anger and frustration can occur in delayed passengers.

To refine proactive scheduling, proposal system is based on Improving Classification Accuracy Using Weighted Multiple Regressions term for flights into delay categories. Our method is based on archived data at major airports in current flight information systems. Classification in this scenario is hindered by the large number of attributes, which might occlude the dominant patterns of flight delays. Therefore introduced a technique which identifies locally relevant attributes for the classification into flight delay categories. An algorithm that efficiently identifies relevant attributes. Our experimental evaluation demonstrates that our technique is capable of detection relevant patterns useful for flight delay classification.

The results of data analysis will suggest that flight delays follow certain patterns that distinguish them from on-time flights. The system may also discover that fairly good predictions can be made on the basis on a few attributes.

Classification and prediction can be used for analyzing future data trends. It is important that the classification is appropriate so that the data prediction is accurate. The regression model will estimate the probability of delay and the classification model classifies whether delay is likely to occur based on the input variables. Results of both the models perform prediction of delay.

2. Review of Literature

Flight delay is a complex phenomenon, because it can be due to problems at the origin airport, at the destination airport, or during airborne. A combination of these factors often occurs. Delays can sometimes also be attributable to airlines. Some flights are affected by reactionary delays, due to late arrival of previous flights. These reactionary delays can be aggravated by the schedule operation. Flight schedules are often subjected to irregularity. Due to the tight connection among airlines resources, delays could dramatically propagate over time and space unless the proper recovery actions are taken. Even if complex, flight delays are nowadays measurable. And there exist some pattern of flight delay due to the schedule performance and airline itself. The Bureau of Transportation Statistics (BTS) compiles delay data for the benefit of passengers. They define a delayed flight when the aircraft fails to release its parking brake less than 15 minutes after the scheduled departure time.

The FAA is more interested in delays indicating surface movement inefficiencies and will record a delay when an aircraft requires 15 minutes or longer over the standard taxi-out or taxi-in time. Generally, flight delays are the responsibility of the airline. Each airline has certain number of hourly arrivals and departures allotted per airport. If the airline is not able to get all of its scheduled flights in or out each hour, then representatives of the airline will determine which flights to delay and which flights to cancel. These delays take one of three forms, ground delay programs, ground stops, and general airport delays. When the arrival demand of an airport is greater than the determined capacity of the airport, then a ground delay program may occur. Generally, ground delay programs are issued when inclement weather is expected to last for a significant period of time.

Second, ground stops are issued when inclement weather is expected for a short period of time or the weather at the airport is unacceptable for landing. Ground stops mean that traffic destined to the affected airport is not allowed to leave for a certain period of time. Lastly, there are general arrival and departure delays. This usually indicates that arrival traffic is doing airborne holding or departing traffic is experiencing longer than normal taxi times or holding at the gate. These could be due to a number of reasons, including thunderstorms in the area, a high departure demand, or a runway change. Our research finds that arrival and departure delays are highly correlated. Correlation between arrival and departure delays is extremely high (around 0.9). This finding is useful to prove that congestion at destination airport is to a great extent originated at the departure airport.

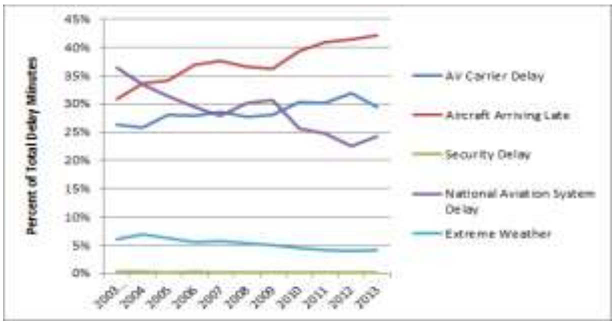

Figure 1: Graph for delay

Popular Implementation

2.1 KnowDelay.com predicts flight problems 3 days in advance

KnowDelay.com [10] crunches weather, airline and airport data to predict weather-related flight delays up to three days in advance. Weather delays are the most stubborn and intractable delays because oftentimes it can take travelers a day or more to get out. "But if you can say three days in advance, don't connect over airport XYZ, change your flight and connect over airport ABC, you may be able to avoid the problem."

KnowDelay users can view a map with colored dots — red, yellow and green — and a slider that lets them see forecasted delays over the next 72 hours. Red means there's 60-percent chance of a 60-minute weather delay; yellow equals a 40-percent chance of a 30-minute delay, and green means a 6-percent chance of delay. During two years of beta testing, the site has accurately predicted 90 percent of weather-related delays. It now covers 37 U.S. airports, including major hubs and other destinations that are typically affected by bad weather.

2.2 How FlightCaster Squeezes Predictions from Flight Data

FlightCaster [1] predicts flight delays using an advanced algorithm that scours data on every domestic flight for the past 10-years and matches it to real-time conditions. They help you evaluate alternative options and help connect you to the right person to make the change.

FlightCaster uses data from:

- Bureau of Transportation Statistics
- FAA Air Traffic Control System Command Center
- FlightStats
- National Weather Service

2.3 Characterization and Prediction of Air Traffic Delays

This system presents a new class of models for predicting air traffic delays [3]. The proposed models consider both temporal and spatial (that is, network) delay states as explanatory variables, and use Random Forest algorithms to predict departure delays 2-24 hours in the future. The System analyzes the performance of the proposed prediction models in both classifying delays as above or below a certain threshold, as well as predicting delay values. Delay prediction has been the topic of several previous efforts.

Jetzki(2009) studied the propagation of delays in Europe, with the goal of identifying the main delay sources. Tu et al. (2008) developed a model for estimating flight departure delay distributions, and used the estimated delay information in a strategic departure delay prediction model. By contrast, Bratu and Barnhart (2005) focused on the impact of delays on passengers.

Other prediction models (Klein et al. 2007, 2010, Sridhar and Chen 2009) have focused on weather-related delays, and the development of a Weather Impacted Traffic Index (WITI). By contrast, the goal of this system is to evaluate the potential of network-scale delay dependencies in developing delay prediction models. The models presented in this system therefore attempt to predict future departure delays on a particular origin-destination (OD) pair by considering current and/or past delays in the network.

The main objective of this study is to predict the departure delay on a particular link or at a particular airport, sometime in the future. The departure delay of a link at time t is an estimate of the departure delay of any flight(s) taking off at time t, and flying on that link. For example, if the BOS-MCO departure delay state two hours from now is estimated to be 30 minutes, it means that the estimated departure delay for BOS-MCO flights taking off two hours from now is 30 minutes. Two types of prediction mechanisms are considered: classification, where the output is a binary prediction of whether the departure delay is more or less than a predefined threshold, and regression, where the continuous output is an estimate of the departure delay along the link.

Ten training sets (3,000 points each) and ten test sets (1,000 points each) were sampled from the 2007-2008 dataset. The prediction models were fit and tested for each of the 10 training and test set pairs, respectively, providing measures of the variability and test error. The training and test sets were over-sampled from the 2007-2008 data. Over-sampling is the over selection of samples of the minority class in order to achieve balanced training and test datasets with sufficient representatives of both the majority and minority classes (Upton and Cook 2008). Since the majority of links do not experience delays of more than 60 minutes, a naive classification algorithm that predicts no delays of more than an hour would be correct most of the time. For this reason, the true evaluation of a classifier's performance is its ability to correctly predict delays in a balanced data set in which half the points have delays of less than 60 minutes (the so-called "majority class"), and half the points have delays of more than 60 minutes (the "minority Class").Different classification and regression models (logistic regression, single classification trees, bagging, boosting, linear regression, neural nets and random forests) were tested, and Random Forests were chosen due to their superior performance.

2.4 Decision Support Tool for Predicting Aircraft Arrival Rates, Ground Delay Programs, and Airport Delays from Weather Forecasts

The system shows the possibilities of a Weather Delay Prediction Tool and what it can do to help NAS stakeholders **[5]**. The algorithm is capable of classifying weather forecasts into three sets, where each set represents a specific AAR. The principle bottlenecks of the air traffic control system are major airports Atlanta, Detroit, St. Louis, Minneapolis, Newark, Philadelphia, and LaGuardia all expect to be at least 98% capacity by 2012. The general procedure used to determine a connection between weather forecast and airport capacity was:

- Collect data from the various available data sources,
- Using assorted tools, format the data into a usable layout,
- use a classification tool to connect the two sets, and
- test the data to ensure there is a correlation.

Weather forecast products are uncertain and the uncertainty increases with lead-time. Useful applications of weather forecasts requires either refinement, consultation, and application of the weather forecast to estimate air traffic capacity or decision support tools that take forecasts and make predictions based on past forecasts and those forecasts connections to NAS capacity. This system describes a methodology used to create one such decision support tool known as the Weather Delay Prediction Tool.

2.5 DelayCast.com

DelayCast [2] is a website that helps you obtain reasonable estimates of the flight delays you may experience. These estimates are based on factors such as the airline, flight origin and destination, developing trends, holidays, date and time of the flight and so on. Apart from estimated predictions you can also use it to check more general overview of the best days, times and airlines to fly.

Features:

- Check out flight delay estimates before booking a flight.
- Airports: Departure and arrival predictions made for the top 60 airports in the United States.
- Currently supported airlines: Southwest, Northwest, JetBlue, America, American, Eagle, Continental, America, West, Delta, US Airways, Alaska Airlines, United and AirTran.
- Free. No-registration required.

2.6 METAR Reader(metarreader.com)

METAR Reader is a website that helps the user to retrieve weather information in METAR string format by just giving the four-letter ICAO Code as the input. This fetches the current weather information of that region code. The information in the METAR string contains time, date, temperature, wind, visibility, sky and cloud conditions, weather behavior and additional remarks. Apart from this user can also convert the METAR string into a simple English language format.

Features:

- The site has an option for viewing the map called as Heat Map, where user can have a topographical view of map along with the regions ICAO Code link. On opening any link user retrieves the current weather information for that region and weather predictions for coming week.

Statistical models and simulation method are used to analyze flight delay. But the analysis on delay is carried on data with only a few days. That is because of the huge data of flights every day. So here the flight delay is categorized into several levels, and the logistic regression models are used here to better identify the delay pattern. Studies on airport delay and delay influence on individual flight are carried out using regression model and classifiers.

3. Report on present Investigation

3.1 Problem Definition

The purpose of Flight Delay Prediction System is to find out if a flight is getting delayed during departure and arrival then what are the reasons for the delay. The system intends to aid the airlines by predicting the delays by using certain data patterns from the previous information. This system explores what factors influence the occurrence of flight delays. The delay is predicted for domestic flights in United States of America. The dataset for the flights obtained from Transtats - Bureau of Transportation Statistics includes data about 306 airports. Using OneR, a classification algorithm, models are developed to predict delay on both arrival and departure side. Discretization was applied using Weka, a data mining tool, to divide the delays on departure and arrival side into five categories viz; Negligible, Insignificant, Nominal, Significant, Indefinite. The result thus obtained from these categories was further analyzed to predict overall reason for delay. The delay predicted can be due to Weather, Security, Carrier, National Aviation System (NAS) and Late Arrival. The models further combine this result with Meteorological Terminal Aviation Routine Weather Report (METAR) to give the report of weather conditions at origin and destination airport. METAR is a format for reporting weather information used by pilots providing a pre-flight weather briefing. The current weather report provides the weather conditions at origin and destination. Experimental evaluation demonstrates that our technique is capable of detecting relevant patterns useful for flight delay by improving classification accuracy using multiple weighted regressions.

The results of data analysis will suggest that flight delays follow certain patterns that distinguish them from on-time flights. System may also discover that fairly good predictions can be made on the basis on a few attributes.

3.2 Methodology

The use cases of FDPS are explained:

- Choose Details: The user can choose the fields from the dropdown box available in the GUI for analyzing the delay pattern.
- Analyze delays: The user can view the various graphs and charts generated as the output.
- Generate Reports: The user can generate reports, save a copy and print it as well.
- Submit Details: The user is required to submit the flight details for which the delay is to be predicted. Details like Origin, Destination, Carrier, Date, Time, etc of the flight needs to be selected.
- Predict Delays: These flight details are submitted to model and is classified as OTP or delayed. If delayed then the numeric probability prediction is made as late, very late, etc.
- Update Flight Schedule: The new flight schedule can be updated.

3.3 Implementation Plan

The proposed system will use data originally is from the Bureau of Transportation Statistics (BTS) to analyze and predict flight departure delays for commercial flights in the United States.

The system has identified following factors:

- Factors which cause flight delay.
- Predict whether the flight will be delayed.
- By how many minutes the delay has caused.

3.3.1Data preprocessing

Data pre-processing includes data cleaning, data transformation and data reduction. The original dataset contains information for all commercial flights in the United States in the year 2014. A reasonable number of records are extracted from the original dataset reducing the size of dataset from 60 lacs to 6 lacs records approximately.

3.3.2Data description

Originally the dataset had 38 attributes with approximately 60 lakh records. The records have been taken for the year of 2014. Roughly 5 lacs records about the airlines have been considered from each month.

The January-December 2014 records have been aggregated to give 60 lac records approximately with 38 attributes. Filtering of the entries was done by reducing instances using Stratified Remove Folds in WEKA. Stratified Remove Folds filter is applied to remove the attributes affecting the departure delay. This filter takes a dataset and outputs a specified fold for cross validation. This reduced the number of records to 6 lacs in the dataset. Also records with missing value labelled as “?” were removed using Remove Duplicates which is formatting to find, highlight and remove duplicate values and retain only unique values.

For simplification purpose, the flights with uncommon attributes were removed. The original dataset also contained a large number of attributes and many of these were discarded because they were irrelevant or repeated information that could be found in other attributes. Attributes like diversion, cancellation have been discarded. As FDPS mainly focuses on departure delay, irrelevant attributes were removed and a new data set was formed with 20 attributes. After resampling, dataset contains around 590951 records, which is used for analysis. On removal of outliers the final dataset contains 590889 records.

The various types of delay have been converted to format of 0s and 1s. The attributes which do not cause delay have been labelled “0” and ones causing delay have been labelled “1”. The original format consisted delay in minutes now replaced by “1”. A combination of these types formed a new attribute “Reason_delay”. An abbreviated form of delay has been taken for Reason_delay attribute.

The delays have been labelled as follows:

C: CARRIER DELAY

W: WEATHER DELAY

N: NAS DELAY

S: SECURITY DELAY

L: LATE AIRCRAFT DELAY

At the end of data pre-processing our dataset has the following features:

1. Flight Features: These include the departure and arrival time, source and destination airport and distance covered by the flight, types of delay viz. Carrier, Late Aircraft, NAS, Weather, and Security.

2. Weather Features: METAR is use for reporting weather information used by pilots and meteorologists to assist in weather forecasting.

3.3.3METAR Weather Reporting

METAR (Meteorological Terminal Aviation Routine Weather Report) is a format for reporting weather information. A METAR weather report is used by pilots for a pre-flight weather briefing. METAR Reader provides a string by entering the ICAO (International Civil Aviation Organization) code of the airport. The ICAO code contains the airport name.

The METAR reader provides the option to convert the string to raw format as well as translated format which can be understood by airport officials.

IATA codes are usually derived from the name of the airport or the city it serves, while ICAO codes are distributed by region and country. ICAO codes are also used to identify other aviation facilities such as weather stations or Area Control Centers, whether or not they are located at airports. The first letter is allocated by continent and represents a country or group of countries within that continent. The second letter generally represents a country within that region, and the remaining two are used to identify each airport.

METAR reports are also provided by Automated Weather Observing System (AWOS), Automated Surface Observing System (ASOS), and Automated Weather Sensor System (AWSS).

Automated stations report “CLR” when clouds may exist above 12,000 feet and “SKC” when sky is completely clear overhead. “OVX” indicates the sky is obscured which is the case when METAR reports vertical visibility and no cloud formation.

METAR information which is recorded on hourly basis is updated and replaced every 20 minutes from the previous observations if any major changes are being observed in weather during that particular hour.

The ICAO code entered in METAR Reader retrieves the weather information of that airport. A METARreport generated is combined with data mining algorithm. The combination output tells whether Delay will occur or not and if yes then reason for delay and by how many minutes has flight been delayed is displayed.

3.3.4 Data Mining Algorithm

The model consists of 13 attributes and Reason_delay attribute as the class label. The various types of delay have been converted to format of 0s and 1s. The attributes which do not cause delay have been labelled “0” and ones causing delay have been labelled “1”. A combination of these types formed a new attribute “Reason_delay”. An abbreviated form of delay has been taken for Reason_delay attribute.

OneR algorithm is one-level decision tree that generates a set of rules that test one particular attribute assuming nominal attributes. It is a classification algorithm that generates one rule for each predictor in the data and then selects the rule with the smallest total error as its "one rule". The attribute with least error rate is the best attribute. Since the accuracy is highest and time to build the model is least i.e. the performance of OneR algorithm is better as compared to other algorithms, the models were developed using this algorithm.

Discretization is done to use classifier to handle only nominal data. Discretization can be done with filter weka.filters.unsupervised.attribute.Discretize which uses simple binning. Attributes like dep_delay_new and arr_delay_new are converted from numeric to nominal by applying discretization.This means system can simply discretize by removing the keyword "numeric" as the type for the " dep_delay_new” and “arr_delay_new " attributes in the ARFF file, and replacing it with the set of discrete values values.

4. Results and Discussion

From this system, the results achieved are feasible and accurate enough to predict delay. At the beginning the dataset is pre-processed to identify the outliers. The pre-processed dataset is then given to the model. Models for predicting flight delay are developed using the data from The Bureau of Transportation Statistics (BTS). The model comprises of 13 attributes and class label “Reason_Delay”. These models are then integrated in the form of a system for delay assessment.

Classifier is used to detect the pattern of delay. This enables us to investigate the delay at the flight level, and the effect of a delay on the immediate flight is considered.

Figure 2 1: Confusion Matrix

The accuracy of the models after pre-processing was 71%. The accuracy was further improved by identifying the outliers and the pattern of outliers affecting those which were correctly classified. Discretization of these outliers helped in improving the accuracy. Improvement gave the accuracy of 82.72%. The following confusion matrix was obtained giving the number of instances correctly classified and instances which are misclassified.

Actual\Predicted	Delay	No Delay
Delay	25978	90109
No Delay	12029	462835

The result of the delay is displayed in form of charts or graphs. The weather conditions are found to be the most significant factors that influence the arrival and departure delay. Hence current weather information is taken from website METAR Reader giving the current weather information and conditions at the airport. The algorithm and models generate stable predictions of flight plans that have small amounts of delay.

Experiment 1:

The accuracy of Naïve Bayes and Bayes obtained is 54.37% and 54.81 respectively. IBk and OneR were approximately giving the same accuracy but the model building time was more for IBk algorithm. Hence, OneR was selected to build the model.

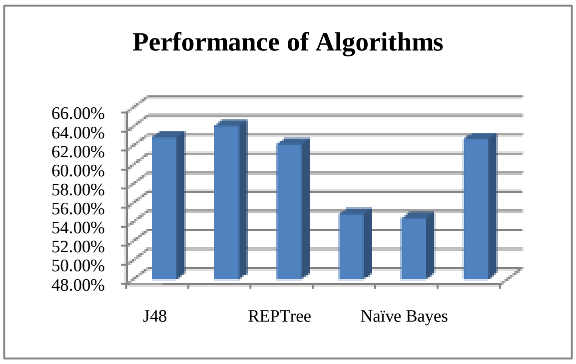

Figure 3 : Graph for Performance of Algorithms

Experiment 2:

The predictor attribute were in the order as shown in the graph below. The graph below shows the weight assigned to every attribute which was determined with the help of accuracy and Relative Dataset Size (RDS).

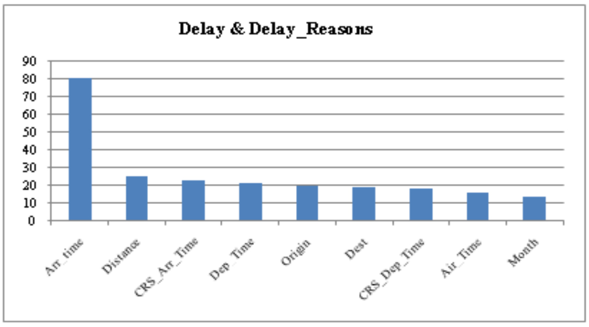

Figure 4 : Delay and Delay Reasons

Experiment 3:

The graph below shows different weights assigned to every attribute in the Departure Delay Model. The predictor attribute were in the order as shown in the graph below. Every attribute weight was determined with the help of accuracy and Relative Dataset Size (RDS).

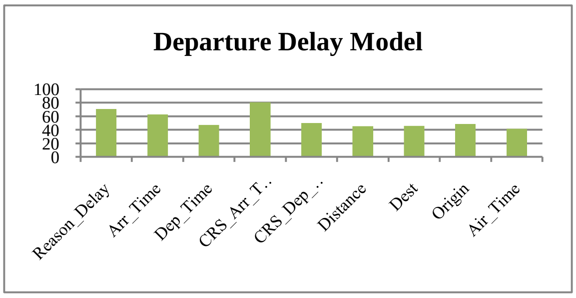

Figure 5 : Departure Delay Model

Experiment 4:

The graph below shows different weights assigned to every attribute in the Arrival Delay Model. The predictor attribute were in the order as shown in the graph below. Every attribute weight was determined with the help of accuracy and Relative Dataset Size (RDS).

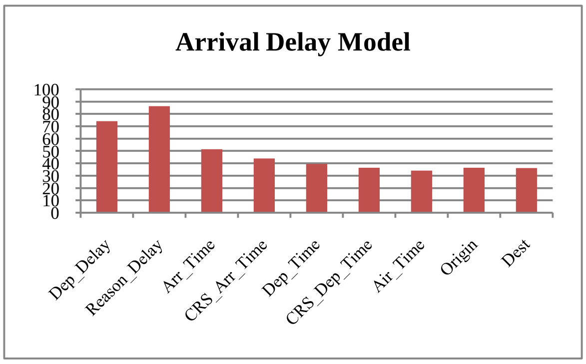

Figure 6 : Arrival Delay Model

Experiment 5:

The accuracy of Naïve Bayes and Bayes obtained is 54.37% and 54.81 respectively. IBk and OneR was approximately giving the same accuracy but the model building time was more for IBk algorithm. Hence, OneR was selected to build the model.

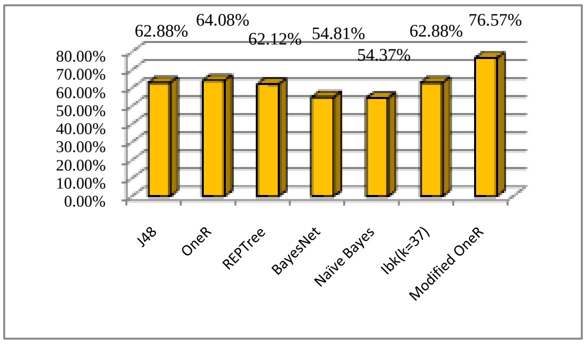

Figure 7 : Comparison of Algorithm Performance

By experimenting with the votes for each model we get the maximum accuracy of 76.57% for the modified OneR model by using the following weights:

$$\frac{9^8}{10}\ \frac{8^8}{10}\ \frac{7^8}{10}\ \frac{6^8}{10}\ \frac{5^8}{10}\ \frac{4^8}{10}\ \frac{3^8}{10}\ \frac{2^8}{10}\ \frac{1^8}{10}$$

The above weights were calculated using the following formulas.

1. *Total weight of 1 class label =*

$$\sum_{n=1}^{25}(\textit{weight assigned to a class label})$$

$$if\left(\sum_{n=1}^{25} \textit{weight assigned to class label "0"}\right) > \left(\sum_{n=2}^{25} \textit{weight assigned to 24 class labels} + 100\right)$$

Then

Result = "Delay"

Else

Result = "No Delay"

2. *Relative Dataset Size (RDS) =*

$$\frac{Dataset\ size\ after\ filtering}{Orignal\ Dataset\ Size\ before\ filtering}$$

3.

4. $Weight = \frac{Accuracy+RDS}{2}$

5. $Weight = \frac{Acuuracy*RDS}{2}$

Table 1 : Delay and Delay Types

No of records	*After Filter*	*On Attribute*	*Accuracy*	*Weight*
5909	47514	Arr_time	99.98	80.39
1158	30035	Distance	99.00	25.68
8577	20382	CRS_Arr_ti	97.82	23.25
6539	14372	Dep_time	98.14	21.57
5102	10125	Origin	99.44	19.73
4089	7932	Dest	99.47	19.29
3296	6340	CRS_Dep_ti	94.12	18.1
2662	4385	Air_time	96.32	15.77
2226	3020	Month	100%	13.56

Table 2: Departure Delay

No of records	*After Filter*	*On Attribute*	*Accuracy*	*Weight*
5908	41949	Reason_dela	99.99	70.99
1713	10785	Arr_time	99.99	62.93
6353	29954	Dep_time	99.94	47.12
3358	17225	CRS_Arr_ti	98.68	80.62
1635	8405	CRS_Dep_ti	97.72	50.22
7950	3823	Distance	94.03	45.22
4127	1935	Dest	97.41	45.67
2192	1105	Origin	96.74	48.77
1087	520	Air_time	86.34	41.37

Table 3 : Arrival Delay

No of records	*After Filter*	*On Attribute*	*Accuracy*	*Weight*
5908	43787	Dep_DeLA	100%	74.1
1530	13205	Reason_dela	99.99	86.3
2096	11053	Arr_time	97.51	51.41
9912	4643	CRS_Arr	99.03	44.05
5269	2405	Dep_time	86.81	39.63
2864	1189	CRS_dep	87.72	36.42
1675	662	Airtime	86.40	34.15
1013	418	Origin	88.03	36.33
595	264	destination	81.43	36.30

The tables above give the weights assigned to different models. The graph for the comparison of experiments is given as shown below.

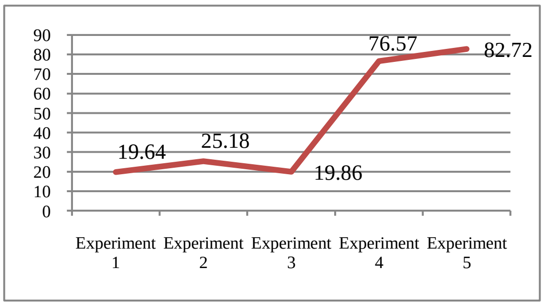

Figure 8 : Experiments Vs Accuracy

The accuracy of 82.72% is highest in Experiment 5 after applying the Modified OneR algorithm.

6. Conclusion

After the development of modules we have come to the conclusion that the models developed can be used in predicting the delay accurately at the airports. The delay distribution of an airport can make it easier to understand the airport delay. The results of the research show that the delay is highly related to the originate delay. In response to single flight delay predictions and reason for these delays that are generated by the model, which can give indications for the appropriate recovery actions to recover/avoid these delays.

The models developed can be applied to predict occurrence of delay at airports. Such predictive capabilities can help the managers and airline dispatchers to prepare mitigation strategies for reducing traffic disruptions. The models are calibrated using historical data. Including weather forecasts as input variables is a direction of future research.

A lot of factors go into predicting a delay in a flight departure. Delays in flight departure can be subjected to various reasons. The results of the data analysis suggest that flight delays follow certain patterns that distinguish them from on-time flights. We discovered that it is possible to make fairly good predictions on the basis of a few key attributes, such as departure time, date and carrier.

By including weather information, we should be able to improve our results even further, and thus get a better picture which largely determines where and when flight delays occur. This will help to save the airport time and hassle. Several factors can be identified and data related to those can be collected and can be used to build various models to better predict the delay in a flight across all airports. A wide variety and a rich collection of data would definitely be useful in building a better model to predict the delay.

The results of the data analysis suggest that flight delays follow certain patterns that distinguish them from on-time flights. From our models and analysis, we discovered that it is possible to make fairly good predictions on the basis of a few key attributes, such as carrier, departure time, arrival time, origin, and destination. A predictive trend within our data from the models that we developed was discovered.

Regression of the data can be done to predict the delay in minutes. Further, the system can be made into an android application.

.

References

[1] J. J. R. a. H. Balakrishnan, "Characterization and Prediction of Air Traffic Delays," Massachusetts Institute of Technology Cambridge, USA, Mar 2014.

[2] S. G. a. B. S. Avijit Mukherjee, "Predicting Ground Delay Program At An Airport Based On Meterological Conditions," AIAA Aviation, Atlanta, June 2014.

[3] D. A. Smith, "Decision Support Tool for predicting Aircraft arrival Rates from Weather forecasts," George Mason University, 2008.

[4] "RITA|BTS|Transtats," [Online]. Available: http://www.transtats.bts.gov/. [Accessed Aug 2014].

[5] "Directorate General of Civil Aviation," [Online]. Available: http://dgca.nic.in/. [Accessed Sep 2014].

[6] "FlightRadar24," [Online]. Available: http://www.flightradar24.com/. [Accessed Sep 2014].

[7] "Yahoo Weather," [Online]. Available: https://in.weather.yahoo.com/india/. [Accessed Sep 2014].

[8] "KnowDelay.com," [Online]. Available: https://www.nbcnews.com/business/travel/knowdelay-com-predicts-flight-problems-3-days-advance-f1C9870958. [Accessed Sep 2014].

[9] "Data Wrangling," [Online]. Available: https://www.datawrangling.com/how-flightcaster-squeezes-predictions-from-flight-data. [Accessed Sep 2014].

[10] "DelayCast," [Online]. Available: http://www.delaycast.com/. [Accessed Oct 2014].

[11] H. B. Juan Jose Rebollo, "A Network-Based Model for Predicting Air Traffic Delays," [Online]. Available: http://www.mit.edu/~hamsa/pubs/RebolloBalakrishnanICRAT2012.pdf. [Accessed Sept 2013].

[12] "Flugzeug," [Online]. Available: http://www.flugzeuginfo.net/table_airportcodes_country-location_en.php#U. [Accessed Nov 2014].

[13] "Aeronautical Information," [Online]. Available: http://www.faa.gov/air_traffic/publications/atpubs/aim/. [Accessed Jan 2015].

Author Profile

Sonali Salunke received the B.E. degrees in Computer Science & Engineering from Naganathhappa Halge College of Engineering, Parli-Vaijyanath, Dist-Beed in 2011.