www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 4 April 2015, Page No. 11668-11677

Flight Delay Prediction System Using Weighted Multiple Linear Regression

Sruti Oza[1], Somya Sharma[2], Hetal Sangoi[3], Rutuja Raut[4], V.C. Kotak[5]

[1,2,3,4], *BE-I.T Student,* [5] *,I/C Principal Shah & Anchor Kutchhi Engineering College, Chembur, Department of Information Technology, Mumbai University, India*

26sruti@gmail.com, somyarsharma@gmail.com, hetal.sangoi393@gmail.com rautrutuja77@gmail.com, vinit_kotak@yahoo.com

Abstract: ***Airline delays caused by bad weather, traffic control problems and mechanical repairs are difficult to predict. If your flight is canceled, most airlines will rebook you on the earliest flight possible to your destination, at no additional charge. Unfortunately for airline travelers, however, many of these flights do not leave on-time. The issue of delay is paramount for any airlines. 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 delays along with the intensity of the delays. 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. The results of data analysis will suggest that flight delays follow certain patterns that distinguish them from on-time flights.
Our system also provides current weather details along with the weather delay probability. We have achieved much better accuracy in predicting delays. We may also discover that fairly good predictions can be made on the basis on a few attribute.

Keywords: Prediction, Classification, Flight Delay, METAR, OneR classification algorithm,Weighted Multiple Linear Regression

1. INTRODUCTION

A flight is delayed 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.

Some of the causes of flight delays are as follows: Maintenance problems with the aircraft, Fueling Extreme weather, such as tornado, hurricane, or blizzard, 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.

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.

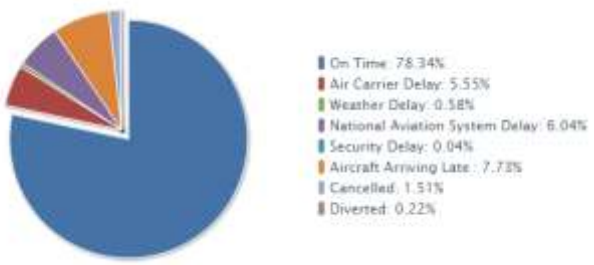

Figure 1: Delay Statistics (January-December 2013)

2. LITERATURE SURVEY

Delays can 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.

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.

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.

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.

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.

Figure 2: Heat Map

3. PROPOSED WORKING SYSTEM

The models developed in this system can be applied to predict the occurrence of flight delay at airports. Such predictive capabilities would help traffic managers and airline dispatchers to prepare mitigation strategies for reducing traffic disruptions. The models are calibrated using historical data on meteorological conditions and traffic demand.

In this paper we have chosen to calibrate the models using weather, late aircraft, NAS, carrier and security conditions. At a given instant the most recent available weather forecast can be used to generate the meteorological conditions. A prediction can be accomplished by applying the models calibrated using historical data. By setting appropriate thresholds on these probabilities one can classify the chances of a delay program under conditions as yes or no.

Figure 1 depicts the working of model for predicting the flight delay. The model developed has 14 attributes including Reason_delay as class label. Data preprocessing help to remove outlier causing noise and redundancy. With 25 combinations of delay, weights are assigned to each label and indexes are assigned to every label based on difference between the weights of every consecutive label. Apart from this, the attributes departure_delay_new and arrival_delay_new are divided into 5 classes viz; Negligible, Insignificant, Nominal, Significant and Indefinite using Discretization. The results of these attributes is further combined with the weather report from METAR to obatin the final prediction.

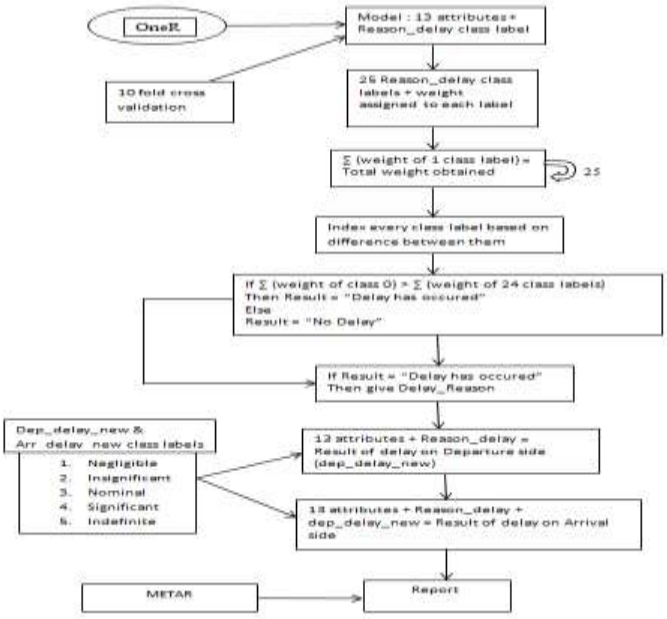

Figure 3: Model for Predicting Delay

4. METHODOLOGY

In this section, we describe the application of the algorithms in predicting, and discuss how we evaluate their performances.

Table I: Performance of Classifying Algorithms

Classifier	**Accuracy**
J48	62.88%
OneR	**64.08%**
REPTree	62.12%
BayesNet	54.81%
Naïve Bayes	54.37%
Ibk(k=37)	62.68%

A. OneR Algorithm

OneR algorithm is one-level decision tree that generates a set of rules that test one particular attribute assuming nominal attributes. The attribute with least error rate is the best attribute.

The accuracy obtained using OneR algorithm is described as follows:

Initially OneR was applied on 15 attributes with 590951 records .After removing misclassified records, 475140 records were obtained. The classifying attribute used by OneR algorithm is Arr_Time .Inverting the selection and again applying OneR 115811 records were obtained.

Accuracy obtained is 99.98%.

Removing Arr_time attribute OneR algorithm is again applied on 115811 records.

After removing misclassified records, 35020 records were obtained. The classifying attribute used by OneR algorithm is FL_NUM. Inverting the selection and again applying OneR 80791 records were obtained.

Accuracy obtained is 94.79%.

Thus OneR algorithm was applied to obtain nine models for better accuracy.

Weight is calculated based on the accuracy and RDS(Relative dataset size). RDS is calculated as follows:

1. *Total weight of 1 class label =*

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

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

Then

Result = "Delay"

Else

Result = "No Delay"

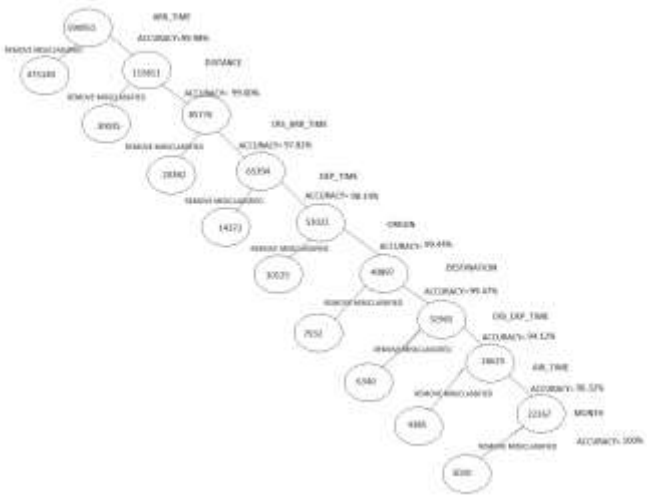

Figure 4: Model development using OneR Algorithm

2. *Relative Dataset Size (RDS) =*

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

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

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

Equation 4 gave better accuracy than equation 3.

Table II: Delay and Delay types

B. Departure and Arrival Delay using Discretization

Initially the range of dep_delay_new and arr_delay_new was 0-1925 minutes which was discretized to give 5 classes.

Discretization is done to use classifier to handle only nominal data.

Attributes like dep_delay_new and arr_delay_new are converted from numeric to nominal by applying discretization. This means we 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.

The WEKA discretization filter, can divide the ranges, or used various statistical techniques to automatically determine the best way of partitioning the data. In this case, we will perform simple binning. First we will load our filtered data set into WEKA. Now, we activate the Filter dialog box and select "weka.filters.unsupervised.attribute.Discretize".

No of records	After Filter	On Attribute	Accuracy	Weight
590951	475140	Arr_time	99.98%	80.39
115811	30035	Distance	99.00%	25.68
85776	20382	CRS_Arr_time	97.82%	23.25
65394	14372	Dep_time	98.14%	21.57
51022	10125	Origin	99.44%	19.73
40897	7932	Dest	99.47%	19.29
32965	6340	CRS_Dep_time	94.12%	18.1
26625	4385	Air_time	96.32%	15.77
22267	3020	Month	100%	13.56

WEKA has assigned its own labels to each of the value ranges for the discretized attribute. For example, the lower range in the "dep_delay_new" and "arr_delay_new" attribute is labeled "(-inf-34.333333]" while the middle range is labeled "(34.333333-50.666667]", and so on. These labels now also appear in the data records where the original dep_delay_new and arr_delay_new value was in the corresponding range.

The range in the above the table are obtained by discretization in WEKA. Initially the range of dep_delay_new and arr_delay_new was 0-1925 minutes which was discretized to give 5 classes.

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 we 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.

The WEKA discretization filter, can divide the ranges, or used various statistical techniques to automatically determine the best way of partitioning the data. In this case, we will perform simple binning. First we will load our filtered data set into WEKA. Now, we activate the Filter dialog box and select "weka.filters.unsupervised.attribute.Discretize".

Next click on the box immediately to the right of the "Choose" button to open the Discretize Filter dialog box. We enter the index for the the attributes to be discretized. Since we are doing simple binning all other options are set to false. On clicking "Apply" the result in a new working relation with the selected attribute partitioned into 5 bins.

Table III: Discretization of Departure delay and arrival delay

	Departure	Arrival
Negligible	0	0
Insignificant	1-15	1-16
Nominal	16 -49	16-49
Significant	50-109	50-109
Indefinite	>109	>109

WEKA has assigned its own labels to each of the value ranges for the discretized attribute. For example, the lower range in the "dep_delay_new" and "arr_delay_new" attribute is labeled "(-inf-34.333333]" while the middle range is labeled "(34.333333-50.666667]", and so on. These labels now also appear in the data records where the original dep_delay_new and arr_delay_new value was in the corresponding range. The instances of the old patterns with the new one can be replaced. Furthermore, the outliers were identified and removed and hence the dataset for departure and arrival delay prediction model had 590889 records.

The instances of the old patterns with the new one can be replaced. Furthermore, the outliers were identified and removed and hence the dataset for departure and arrival delay prediction model had 590889 records.

The Departure and Arrival Delay model are shown in Figure 5 and Figure 6.

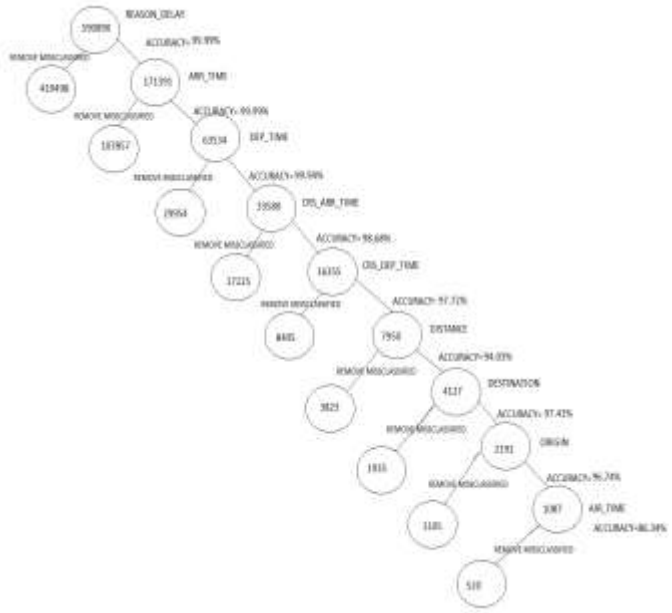

Figure 5: Departure Delay Model

Table IV: Departure Delay

No of records	After Filter	On Attribute	Accuracy	Weight
590890	419498	Reason_delay	99.99%	70.99
171391	107857	Arr_time	99.99%	62.93
63534	29954	Dep_time	99.94%	47.12
33580	17225	CRS_Arr_time	98.68%	80.62
16355	8405	CRS_Dep_time	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

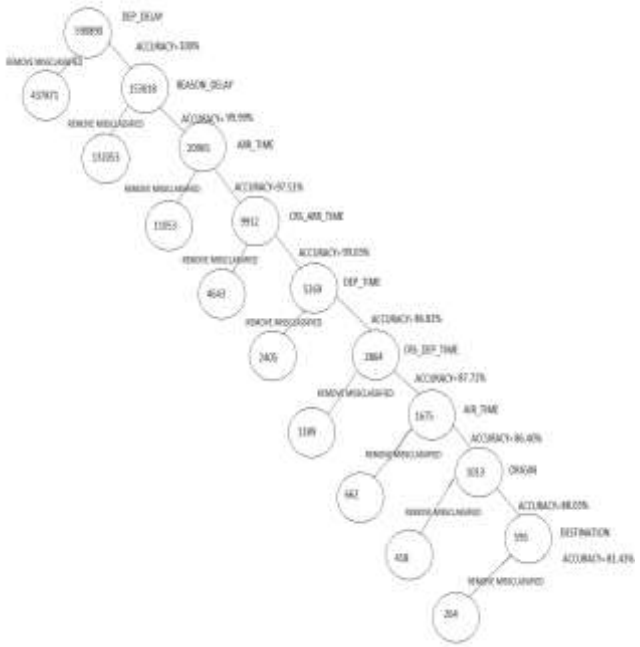

Figure 6: Arrival Delay Model

Table V: Arrival Delay

No of records	After Filter	On Attribute	Accuracy	Weight
590890	437871	Dep_DeLAY	100%	74.1
153018	132053	Reason_delay	99.99%	86.3
20965	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

C. Delay Reasons

Reason delay consisting of 25 class labels has weights assigned to each label.

Table VI: 25 combinations of delay

C	W	N	S	L	Reason Delay
0	0	0	0	0	0
0	0	0	0	1	L
0	0	0	1	0	S
0	0	0	1	1	SL
0	0	1	0	0	N
0	0	1	0	1	NL
0	0	1	1	0	NS
0	0	1	1	1	NSL
0	1	0	0	0	W
0	1	0	0	1	WL
0	1	1	0	0	WN
0	1	1	0	1	WNL
0	1	1	1	0	WNS
1	0	0	0	0	C
1	0	0	0	1	CL
1	0	0	1	0	CS
1	0	0	1	1	CSL
1	0	1	0	0	CN
1	0	1	0	1	CNL
1	0	1	1	0	CNS
1	0	1	1	1	CNSL
1	1	0	0	0	CW
1	1	0	0	1	CWL
1	1	1	0	0	CWN
1	1	1	0	1	CWNL

Types of delay can be categorized as:

1. CarrierDelay (abbreviated as 'C'): The cause of the cancellation or delay was due to circumstances within the airline's control (e.g. maintenance or crew problems, aircraft cleaning, baggage loading, fueling, etc.).
2. WeatherDelay (abbreviated as 'W'): Significant meteorological conditions (actual or forecasted) that, in the judgment of the carrier, delays or prevents the operation of a flight such as tornado, blizzard or hurricane.
3. NASDelay (abbreviated as 'N'): Delays and cancellations attributable to the national aviation system that refer to a broad set of conditions, such as non-extreme weather conditions, airport operations, heavy traffic volume, and air traffic control.
4. SecurityDelay(abbreviated as 'S'): Delays or cancellations caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines in excess of 29 minutes at screening areas.
5. LateAircraftDelay (abbreviated as 'L') **:** A previous flight with same aircraft arrived late, causing the present flight to depart late.

D. Equations

∑weight (0) =Total weight assigned to class '0'

∑weight (CL) =Total weight assigned to class 'CL'

This gives total 25 weights for class labels.

Based on the difference between the weights assigned, each class label is indexed to \check whether delay has occurred or not.

If ∑weight (0) > (∑weight (CL) +100) then Delay has occurred else if

∑weight (0) < (∑weight (CL) +100) then Delay did not take place.

If scheduled time and actual time of departure is available and if the delay has occurred then result is given along with reason for Delay.

Accuracy for the proposed algorithm is: 82.72%

Table VII: Confusion Matrix for the Algorithm

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

E. METAR

Weather is a major contributor to large delays. 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 which is a four character alphanumeric code.

For example, the ICAO code KLAX stands for Los Angeles International Airport, United States. The METAR Reader takes ICAO code as input and provides string which contains various parameters like report time, winds, visibility, clouds, temperature, pressure and other remarks about the snow/rain precipitation.

A METAR report contains the following sequence of elements in the following order:

1. Type of report.
2. ICAO Station Identifier.
3. Date and time of report.
4. Modifier (as required).
5. Wind.
6. Visibility.
7. Runway Visual Range (RVR).
8. Weather phenomena.
9. Sky conditions.
10. Temperature/dew point group.
11. Altimeter.
12. Remarks (RMK).

ICAO Station Identifier. The METAR code uses ICAO 4-letter station identifiers. In the contiguous 48 States, the 3-letter domestic station identifier is prefixed with a "K;" i.e., the domestic identifier for Seattle is SEA while the ICAO identifier is KSEA. Elsewhere, the first two letters of the ICAO identifier indicate what region of the world and country (or state) the station is in. For Alaska, all station identifiers start with "PA;" for Hawaii, all station identifiers start with "PH." Canadian station identifiers start with "CU," "CW," "CY," and "CZ." Mexican station identifiers start with "MM." The identifier for the western Caribbean is "M" followed by the individual country's letter; i.e., Cuba is "MU" Dominican Republic "MD;" the Bahamas "MY." The identifier for the eastern Caribbean is "T" followed by the individual country's letter; i.e., Puerto Rico is "TJ." For a complete worldwide listing see ICAO Document 7910, Location Indicators.

A mapping of ICAO codes and IATA (International Air Transport Association) is prepared to obtain ICAO code for corresponding airport like LAX i.e. Los Angeles International Airport has KLAX ICAO code. 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.

To determine the weather conditions at airport the string is generated according the threshold values given below.

1. Wind: speed given in Knots (KT)
   Wind_speed <10KT is considered as moderate wind in air
   Wind_speed >25KT is considered as high wind in air.

2. Visibility: given in Statute Miles(SM).
   Visibility <4SM is considered to be low visibility

3. Sky/Cloud conditions: given as (OVC/FEW/CLR/SCT/CB/TCU)

   OVC: Overcast
   FEW: Few
   CLR: No clouds below 12,000 ft. (3,700 m)
   SCT: Scattered
   CB: cumulonimbus cloud
   TCU: towering cumulus
4. Weather information follows the format:
   Intensity... Description... Precipitation... Obscuration... Other
   (-/+)... (PR/SH/TS/FZ/DR)... (DZ/RA/SN)... (BR/FG/FU/SA)... (SS/DS/FC/SQ)
   PR: partial
   SH: showers
   TS: thunderstorm
   FZ: freezing
   DR: drift
   DZ: drizzle
   RA: rain

SN: snow
BR: mist
FG: fog
FU: smoke
SA: sand

Example of METAR report and explanation:
METAR KSFO 041453Z AUTO VRB02KT 3SM BR CLR 15/12 A3012 RMK AO2
METAR aviation routine weather report
KSFO San Francisco, CA
041453Z date 4th, time 1453 UTC
AUTO fully automated; no human intervention
VRB02KT wind variable at two
3SM visibility three
BR visibility obscured by mist
CLR no clouds below one two thousand
15/12 temperature one five, dew point one two
A3012 altimeter three zero one two
RMK remarks
AO2 this automated station has a weather discriminator (for precipitation)

Descriptor			
BC - Patches	**BL** - Blowing	**DR** - Drifting	**FZ** - Freezing
MI - Shallow	**PR** - Partial	**SH** - Showers	**TS** - Thunderstorm
Weather Phenomena Precipitation			
DZ - Drizzle	**GR** - Hail	**GS** - Small Hail/Snow Pellets	
IC - Ice Crystals	**PL** - Ice Pellets	**RA** - Rain	**SG** - Snow Grains
SN - Snow	**UP** - Unknown Precipitation in automated observations		
Obscuration			
BR - Mist (≥5/8SM)	**DU** - Widespread Dust	**FG** - Fog (<5/8SM)	**FU** - Smoke
HZ - Haze	**PY** - Spray	**SA** - Sand	**VA** - Volcanic Ash
Other			
DS - Dust Storm	**FC** - Funnel Cloud	**+FC** - Tornado or Waterspout	
PO - Well developed dust or sand whirls		**SQ** - Squall	**SS** - Sandstorm

Figure 7: Weather information description

4. RESULTS AND DISCUSSION

From this system, the results achieved are feasible and accurate enough to predict delay. At the beginning the dataset is preprocessed to identify the outliers. The preprocessed 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.

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.

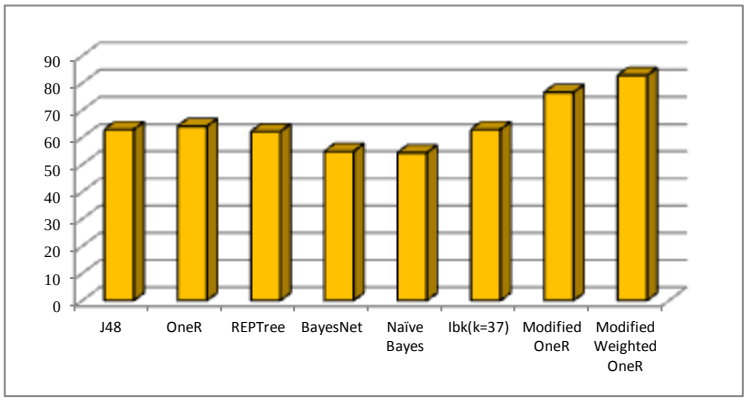

Figure 8: Performance of Algorithms

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).

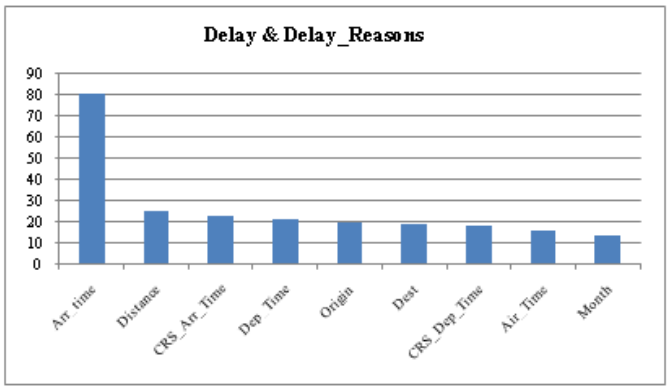

Figure 9: Delay and Delay Reasons

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).

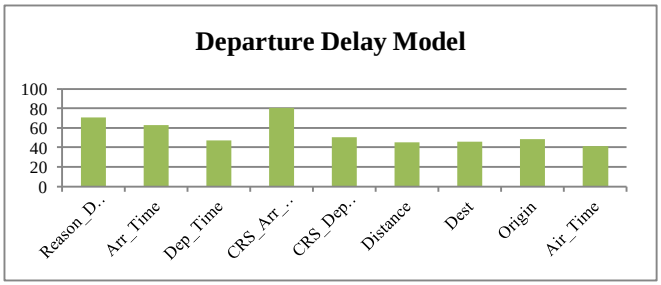

Figure 10: Departure Delay Model

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).

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

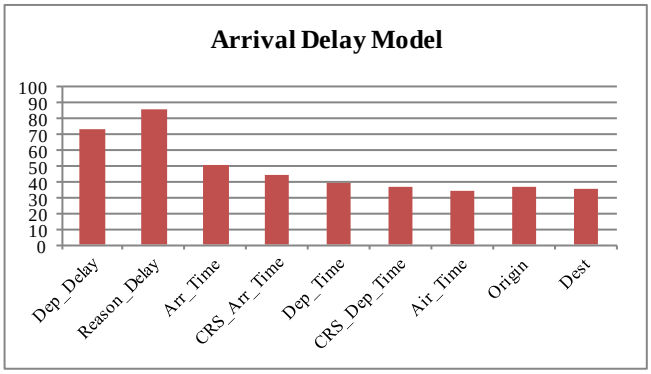

Figure 11: Arrival Delay Model

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.

The accuracy of the models after preprocessing was 64%. 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 confusion matrix was obtained giving the number of instances correctly classified and instances which are misclassified.

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.

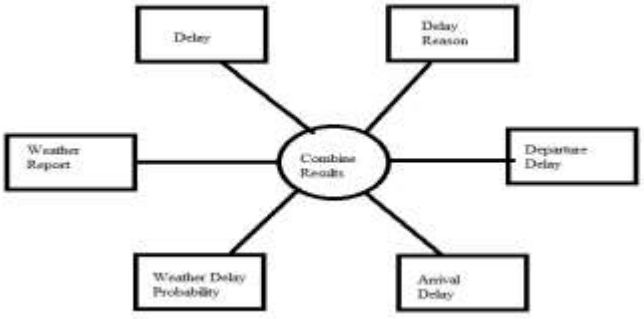

Figure 12: Output of the system

Results provided by the system are as follows:

1. Delay: If delay is present then 'Yes' , else 'No'
2. Delay Reason: The delay reason can be out of any possible combinations of Carrier delay, Weather delay, NAS delay, Security delay and Late Aircraft delay
3. Departure delay: It is given as: Negligible(0 min), Insignificant(1-16 min), Nominal(16-49 min), Significant(50-109) and Indefinite(>109 min)
4. Arrival delay: It is given as: Negligible(0 min), Insignificant(1-16 min), Nominal(16-49 min), Significant(50-109) and Indefinite(>109 min)
5. Weather Delay Indicator: It calculates the probability of delay based on the METAR report.
6. Weather Report: It displays the decoded METAR report of the origin and destination airports

Table VIII: Confusion matrix for algorithm

The weather conditions are found to be the most significant factors that influence the arrival delay due to the destination airport. Thereby 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.

5. 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.

We hope this paper will help to develop a better flight delay prediction system which can minimize the inconvenience caused to the flight passengers due to delays.

REFERENCES

[1] [Online]. Available: https://www.ivao.aero/training/tutorials/metar/metar.htm. [Accessed Dec 2014].

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

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

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

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

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

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

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

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

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

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

[12] 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].

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

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

[15] G. B. a. D. Derry, "A Simple Enhancement to One Rule Classification," Melbourne.

[16] "METAR Reader," [Online]. Available: http://www.metarreader.com/. [Accessed Dec 2014].

[17] "Airport List," [Online]. Available: http://en.wikipedia.org/wiki/List_of_airports. [Accessed Dec 2014].

[18] "METAR Tutorial," [Online]. Available: http://www.wunderground.com/metarFAQ.asp. [Accessed Dec 2014].