Web Derived Application for Flight Delay Simulation

Siddesh Pillai
srp4698@rit.edu

Pranay Shashank Adavi
pa8546@rit.edu

Aradhya Mehta
adm9401@rit.edu

ABSTRACT

This application will give us a detailed outlook on the analytical structuring of the commercial flight dynamics viz. traffic time, schedulers, gateways, passengers travelling, traffic delays and cost of operation.

1. INTRODUCTION

The aviation industry is the one of the most operational sensitive industry. There are many complexities involved in handling data which are dynamically unsteady in patterns. And thus it becomes even more difficult to predict with scattered data. Today, the practice of prediction is done via live monitoring. Airlines today are looking for ways to make flights more efficient. From gate conflicts to operational challenges to air management, the dynamics of flights can change quickly and could lead to costly delays. The motive of this application is to give a rough idea the behavioral of working model how airport organize the traffic of flights as well as flight scheduling and also the cost of operation. This application will aid to classify the happenings easily and based on that we can practice more economical mode of operation which are efficient and consistent. Imagine if we could augment this data to the airports and pilot in the cockpit that would allow them to make adjustments to their flight patterns.

2. MOTIVATION

Today, airports invests a lot capital for their analytics modelling yet sometimes deceived in the optimality of the delivery. This application is solely designed and devised for a statistical analytical understanding on the mode of operation of the commercial aviation segment, everything under the hood and make it easy for easy access.

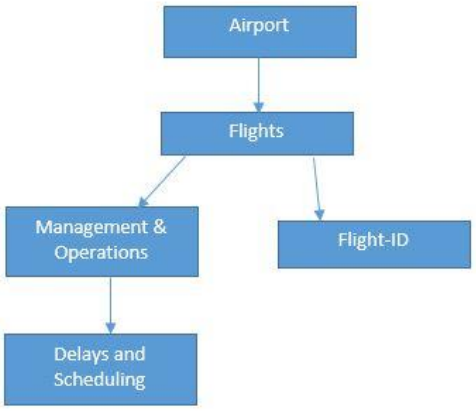
3. DATA DESCRIPTION

The data set that we used is from the gequest repository powered by kaggle. It is a whole bunch of data related to the flight dynamics, but we are concerned with the five prominent data such as Airports, flight plans, Operation with respect to each day, Flight history events and Flight history. The data set is of a recent date from January 2014 to February 2014. The files are around 146 MB in size. These reading were actually taken with respect to each airport. We are of the idea to create schema on the database with the data. With this data multiple patterns can be achieved based on the nature of the runs we make on the data-base system.

Various taxonomies involved in the data which are as follows:
Airport: 4 letter airport code.
Id: Unique identifier of a flight
Departure airport: airport identifies of where the flight departed.
Arrival airport: airport identifies of where the flights destination.
Standard Passenger: count of passenger in coach.
Premium Passenger: count of passengers in 1st class.
Initial fuel: fuel pounds the plane had at departure.
Consumed fuel: fuel pound the plane had up till cutoff time.
Schedule gate departure: hours in relation to the cutoff time.
Schedule date arrival: hours in relation to the cutoff time.
Fuel cost: USD per gallon of fuel.
Crew delay cost: cost per hour accrued by a delay.
Other delay cost: cost per hour due to delay by other factors.
Non arrival penalty: cost of failing to reach the destination.
Delay cost penalty: with respect to each passenger classified into 30 minutes to 2 hours and beyond.
Max standard delay cost: Max USD delay cost for each coach passenger.
Max premium delay cost: Max USD delay cost for each 1st class passenger.
These will help you if you wish to look into the datasets and find out what all these field really mean.

4. DESIGN

The onset of the data flow is that we get inputs in the form of flight stats in constant time with respect time, where the flight is bound to, the cutoff time, schedule time, flight id, operation cost, fuel consumption and flight dynamics with respect to altitude. We extract this data using the regression and classify the data related with which we can foresee the scope and trend. We are using static data for this model. If we are able to meet the goals this can be practiced for a much higher scale. The below picture gives us a better idea how are we going to classify our data.



Here, when we look at the flow we have Airport and its attributes, flights is class which again has its attributes which possess the functionalities. We are concerned with minimizing the delays and scheduling of the flights. This is just a rough illustration to understand how the data is classified to get a better picture.

5. IMPLEMENTATION

The dataset are of *.csv format. As discussed earlier we are concerned with tables such as Airports, Flight Plans, flight with respect to date, Flight history and Flight history events. We derive data from the table as shown below.

id	Depart Airport	Arrival Airport	Schedule Gate Depart	Actual Gate Depart	Schedule Gate Arrival	Actual Gate Arrival
3240845	KSFO	KSEA	-0.8149	-0.8315	1.068425	1.268425
3240710	KMSP	KATL	-0.8149	-0.9149	1.635091	1.635091
3240332	KBUR	KDEN	-0.4815	-0.4149	1.718425	1.768425
3240866	KSLC	KDTW	-1.3149	-1.4482	2.018425	2.218425
3240334	KBWI	KMKE	-0.8149	-0.5649	1.368425	1.185091

This is a sample format of the schema. As it is difficult to represent it all the fields on the paper. However, we have attached a sample_raw_input which can be referred to get better insight of the data sets.

.

6. CURRENT STATUS

Flight Simulator is a proposed open source web based representation that delivers a real-time flight profile to airport which would allow airports to manage them efficiently. Upon success we will go ahead to deliver even for mobile application by porting. As of now we have refined our data from a large of data set.

Timeline of Implementation:

Stage 1: Data Collection
We have all the necessary files that we could get all the data. Since this results in entries over 250,000. We have arrived at a state where we are needed to prune irrelevant data or columns.

Stage 2: Data Cleaning
The tool that we used to clean the data is Microsoft Excel. Implementation was much simpler than expected.

Stage 3: Querying
Now, all that we have is raw data. We perform queries in various files to merge all the table into a single csv file. So that we performing the mining and information retrieval becomes easy. I have attached one of the original files and then some query commands resulting into a merge file. MySql is used for implementation.

Stage 4: Mining
We have a training data set. We have performed a behavioral analysis using R which has so far been implemented in batches. We are yet to perform on all the set of records. We are set to achieve this in coming weeks.

Stage 5: Interface
We are developing an interface, where in a user looks up to schedule a travel giving the Source terminal and Destination terminal. Based on our mapping function which is purely predictive based mining from our training data set will return the percentage of delays with respect to the airports mentioned.

Stage 6: Web Application
With PHP we can groove the raw data from the MySQL. Performance and deliverance are taken into account to illustrate the model for the simulation. We have achieved 5 major functionalities such as:
1. Flight inbound/output with respect to airport.
2. Flight Delays.

We look to map into webpage when the user queries with respect to an airport, flight or day. The analytics is a structured and by automating the date we wish to achieve it even more specific. The visualization component will be R Reporting Service.

7. FUTURE WORK

In our implementation, we have entered only a limited number of airport names in the drop down list as the list was very vast. But if, for the want of completeness, we can make use of all the airport names and expand our work. Similarly, our implementation focuses on just analyzing flight delays, similarly there are many attributes in our file which can be worked on and also analyzed such as calculating fuel consumption for a certain distance, or calculating the average cutoff time for every flight in a certain range of time etc. We team-mates do plan to devote some time in the coming months to improve our application even further so that it can be of wide scale use.

8. CONCLUSIONS FROM DATA ANALYSIS

We can conclude from the snapshots provided by us in the demo project that the naïve bayes classifier used and the decision tree algorithm used to mine the data has been implemented successfully. The results or an example can be obtained by looking at the demo project. The probability calculated gives us the likeness of a flight delay occurring on that sector on a certain data. Similarly, the verdict gives us a yes/no verdict based on whether the flight has been delayed or not.