Team I1.2 - Predicting On-time Performance in the Airline Industry Using Statistical Modeling

Nydia Huynh, Carlos Alvarez, Anthony Lako

TEXAS STATE UNIVERSITY
The rising STAR of Texas

Sabre

Background Information

- Occasional flight delays cause disruptions in the worldwide transportation system.
- Airlines suffer penalties and fines which drive up operation costs.
- Passengers plan earlier appointments, increasing travel cost, in order to ensure on-time arrival.
- Accurate on-time performance predictions are crucial for decision making by all stakeholders in commercial aviation.

Problem Statement

The application of existing prediction models has fallen short of expectations in terms of their accuracy.

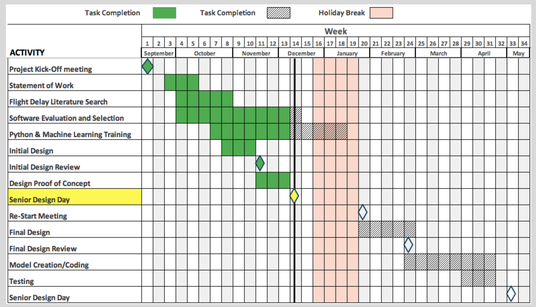
Project Purpose

This project will attempt to improve the precision of on-time flight performance prediction by applying machine learning and statistical methods to improve existing prediction models. Improved accuracy will enable all stakeholders to make better decisions, plans, and schedules.

Objectives

- Create code that interprets expected variables and their effect on on-time performance
- Implement machine learning to improve existing flight on-time performance models
- Create tabular and graphical output of on-time performance data that is user-friendly

Project Schedule



Design Approach & Deliverables

Software Applications

An improved on-time performance model will be developed using Python language and machine learning applications in Spyder.

Machine Learning — python™ — SPYDER

Design Model Structure

The model was constructed in two phases. Phase One is structured to predict whether a flight will be on-time or late. Phase Two will be structured to predict the number of minutes a flight will deviate from its scheduled departure and arrival time based on various input variables.

Phase One: Fall Semester

Inputs:

Departure Time

MONTH	AIRLINE	DEPARTURE _TIME	DEPARTURE _DELAY
1	AS	2354	-11
1	AA	2	-8
1	US	18	-2
1	AA	15	-5
1	AS	24	-1
1	DL	20	-5

Arrival Time

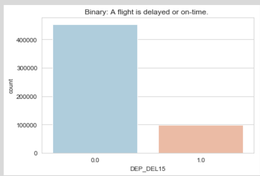
MONTH	AIRLINE	ARRIVAL_ TIME	ARRIVAL_ DELAY
1	AS	408	-22
1	AA	741	-9
1	US	811	5
1	AA	756	-9
1	AS	259	-21
1	DL	610	8

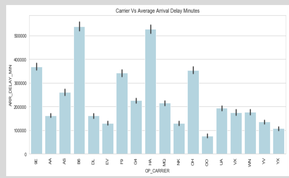
Model to Predict On-Time Flight Performance

Outputs:

Binary Solution: On-Time or Late

Month	Day of Month	Day of Week	Carrier	Scheduled Departure Time	Scheduled Arrival Time	Predicted Departure Time	Predicted Arrival Time	Departure Delay	Arrival Delay
3	7	5	HA	900	1007	900	1010	1	1
3	7	5	HA	835	1145	855	1200	0	1
3	7	5	HA	1205	703	1203	703	1	1
3	7	5	HA	1503	1704	1512	1710	1	0
3	7	5	HA	1705	1900	1730	1910	0	1
3	7	5	HA	800	1007	818	1023	0	0
3	7	5	HA	948	1203	945	1201	1	1






```
Code Snippet:
print "Flight from JFK to " + destination + " , Distance" + str(Distance)
if p==0
   print "Your flight does not have a delay, Accuracy = %1.2f %%" % (Accuracy*100)
else:
   print "Your flight will be delayed, Accuracy = %1.2f %%" % (Accuracy*100)

Output Screen:
Flight from JFK to OAK, Distance: 2576
Your flight does not have a delay, Accuracy = 82.17%
```

Phase Two: Spring Semester

Inputs:

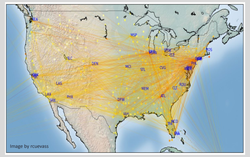
Existing Variables:

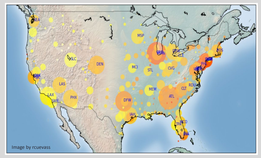
Traffic, Weather, Maintenance, Crew Schedule, National Aviation System, Airport Parking Congestion, Security, Baggage Loading & Unloading, Fueling, Customer Conflicts

Outputs:

Month	Day of Month	Day of Week	Carrier	Scheduled Departure Time	Scheduled Arrival Time	Predicted Departure Time	Predicted Arrival Time	Departure Delay	Arrival Delay
3	7	5	HA	900	1007	900	1010	0	3
3	7	5	HA	835	1145	855	1200	20	15
3	7	5	HA	1205	703	1203	703	-2	0
3	7	5	HA	1503	1704	1519	1710	9	16
3	7	5	HA	1705	1900	1730	1910	25	10
3	7	5	HA	800	1007	818	1023	18	16
3	7	5	HA	948	1203	945	1201	-3	-2

Improved Graphical Outputs:

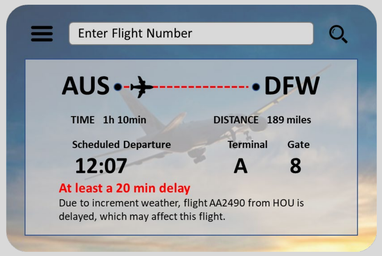




```
Output Screen:
Flight from JFK to OAK, Distance: 2576
Weather on given date will be 82 degrees.
Maintenance is scheduled.
The flight will be delayed 17 minutes, Accuracy = 89.31%
```

User Interface:

How To: The user will enter their flight number in the search bar. The app will output the flight's current status and reasoning.



Literature Review

- Prediction models have been in existence for quite some time.

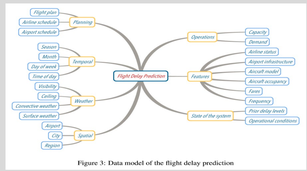
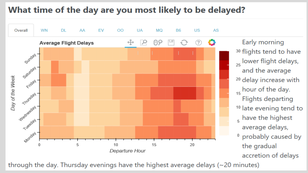


Figure 3: Data model of the flight delay prediction

- Models have been developed using statistical modelling, such as linear regression.
- Now, machine learning tools are being integrated in prediction models.
- These tools are capable of the most accurate prediction by seeking the correct methodology to use for the data.
- Example using machine learning tools:



- Google Flights is one of the few in the market that has incorporated machine learning tools to predict on-time flight performance. Their model has a prediction accuracy of 89%.

Evaluation Criteria

Phase One:

Metric: Binary

- 1: Flight is considered on-time if it arrives within 15 minutes or less
- 0: Flight is considered late if it is delayed more than 15 minutes

Scale:

Prediction Accuracy	Score
95% - 100%	10
89% - 94.99%	9
83% - 88.99%	8
77% - 82.99%	7
71% - 76.99%	6
65% - 70.99%	5
59% - 64.99%	4
53% - 58.99%	3
47% - 52.99%	2
41% - 46.99%	1
< 41%	0

Phase Two:

Metric:

- Early by 1 or more mins
- On-Time (delayed by 0-14 mins)
- Delayed 15-29 mins
- Delayed 45-59 mins
- Delayed 60 or more mins

Scale:

Prediction Accuracy	Score
90% - 100%	10
89% - 89.99%	9
88% - 88.99%	8
87% - 87.99%	7
86% - 86.99%	6
85% - 85.99%	5
84% - 84.99%	4
83% - 83.99%	3
82% - 82.99%	2
81% - 81.99%	1
< 81%	0

Acknowledgements

We would like to thank the following individuals for their help in our project:
Dr. Jiadong Wang - *Sabre Corporation*
Dr. Patrick Thomas - *Texas State University*
Dr. Zhijie Dong - *Texas State University*