Optimizing Business-Travel Schedule with Machine Learning

Team Members:

110013122

180009879

180008373

170026763

180011028

Executive Summary

Flight delays have been a growing issue and have reached an all-time high in recent years. We present a boosted decision tree which correctly predicts if a flight will be delayed more by than 20 minutes with 72% sensitivity. These predictions were made on hundreds of thousands of observations that had not been used to train the model. Using this model will massively improve your ability to avoid booking employees on flights which are subsequently delayed. This will save on costs, time and employee satisfaction(Gustafson, 2012).

Problems and Objectives

As a by-product of the growing of your businesses, business-travels consist of significant proportion of your operating expences.(Shargorodsky et al., 2016) Moreover, delays and reschedules of flights impose costs on your business operation and reduce satisfaction of not only your employees but also your potential client. These are especially acute for your businesses that relay heavily on cross-region transactions of goods and services. Longer flight times leave employees spending time in the air that could be used more productively or enjoyably. Rescheduling flights cause pressure and disactisfaction. Uncertainty about arrival and departure times leads to inconvenience and can leave your employees stranded on the runway or in the departure lounge, when they are expected by their clients.

We are delighted to present our solution, an accurate and easily interpretable model that correctly predicts delays with 72% sensitivity. There are always different ways to approach an answer, but we understood the importance to you of using a model that is simple and provides accurate results. Therefore, after testing many models, we present a boosted decision tree fits both of these criteria. This model will bring many benefits to your business. Firstly, it will save administration time as it is quick and easy to use and provides an immediate answer on if the flight will be delayed or not. Also, time will be saved as employees will be unlikely to experiences delays when travelling. This also has the benefit of improving employee satisfaction, as studies have shown that business travel can have a detrimental effect on job satisfaction . Therefore, it is important that this experience is as pleasant as possible. In addition to this, it has been reported that flight delays cost billions and passengers have to foot the cost for much of this .

Solutions and Findings

We conduct thorough backgroud investigation and collect data from the U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics (BTS), who tracks the on-time performance of domestic flights operated by large air carriers. We came up with several findings. In Figure 1, different carriers have significantly different popularity. Delays are most common for busy airlines, such as WN. These types of flight carriers are better to avoid if possible when arranging bussiness travals. Also, Figure 2 shows delays distribution for different departure time. Delays are more serve during 0-3 a.m. in the moring, and significantly drop at 4 a.m.

Combine with these findings, we used decision tree to select the most important factors that may contribute to flights delays . The following attributes are considered the driving factors of flights delays:

1. Hour of scheduled departure
2. Airline
3. Distance of flight
4. Origin
5. Destination

Using only these variables as predictors, different models were built and the best one were chosen. Our machine learning model, boosted decision tree, can predicts if a flight will be delayed by 20 minutes or longer. This allows you to make an accurate prediction with readily available information. More importantly, it was the best model for correctly predicting flight delays, which we understand is more important than correctly predicting on-time departures (although it does this with great accuracy as well). The model was tested against unseen data and returned accuracy ratings as shown in the table below.

Table 1: Model performance

Overall Accuracy	Sensitivity
64.0%	72.9%

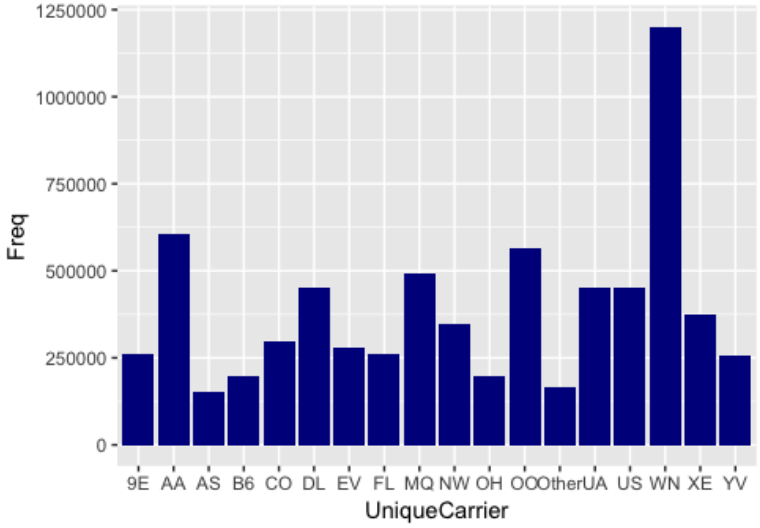

Figure 1: Flights by Unique carrier

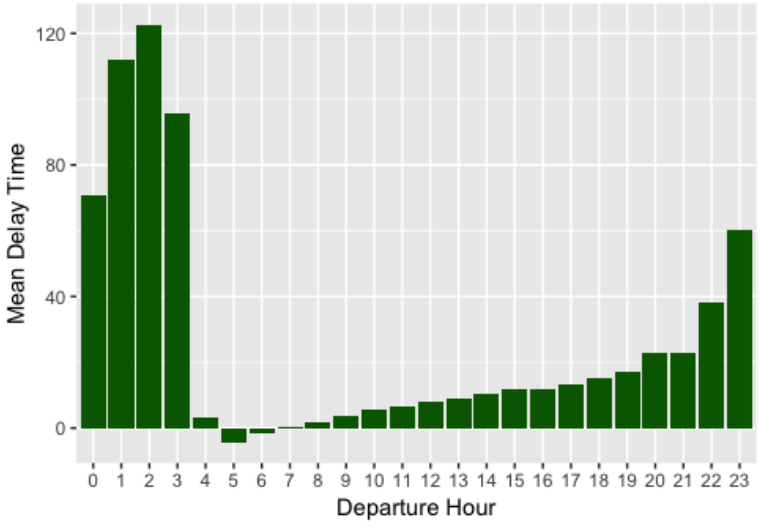

Figure 2: Average Delays by departure hour

Conclusion

With operating costs always under review in most businesses and resource planning central to an effective operation, the earlier the warning of flight delays, the better for businesses. Our model helps your company and your employees plan and respond to a change in scheduled arrival times, from managers on travel for business meetings, to finance department effectively reallocate resources and subsequent contingency planning. (Ball et al., 2010)

In figure 3, you can have a straightforward understanding of how our model can help your company reduce losses. The x axis represents the fraction of total cost induce by False Positive mistakes, y-axis represents the fraction of total benefit generate by True Positive prediction. Campare with random guesses, our model significantly reduce costs, measuring by time loss due to flight delay, and even generate profit, measuring by time save due to avoiding flight delay.

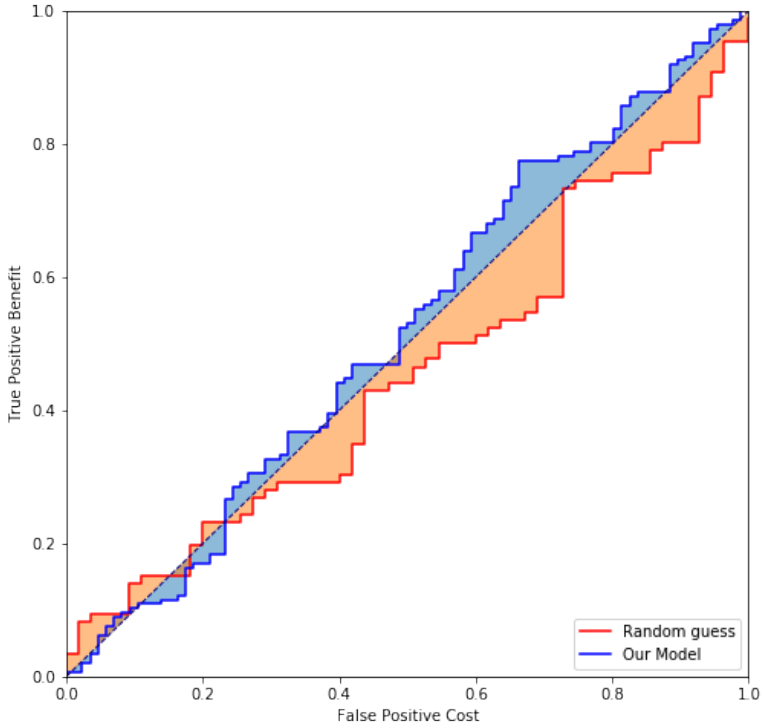

Figure 3: Cost Sensitive ROC (Fawcett, 2011)

References

Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A., Peterson, E., Sherry, L., Trani, A., Zou, B., et al. (2010). Total delay impact study. In *NEXTOR Research Symposium, Washington DC. http://www. nextor. org*, pages 1–67.

Fawcett, T. (2011). Roc graphs: Notes and practical considerations for researchers. *Machine learning*.

Gustafson, P. (2012). Managing business travel: Developments and dilemmas in corporate travel management. *Tourism Management*, 33(2):276–284.

Shargorodsky, J., Zheng, L., Stillman, F., Soong, A., Navas-Acien, A., and Reh, D. (2016). The economic cost of airline flight delay. *Journal of Transport Economics and Policy*, 6:437–444.