Application of Machine Learning Algorithm to Predict Flight Delays

Nathalie Kühn and Navaneeth Jamadagni {nk1105, njamadag}@stanford.edu

STANFORD BIO-X

$19 Billion: ***Annual Economic Impact of US domestic flight delays to the airlines***

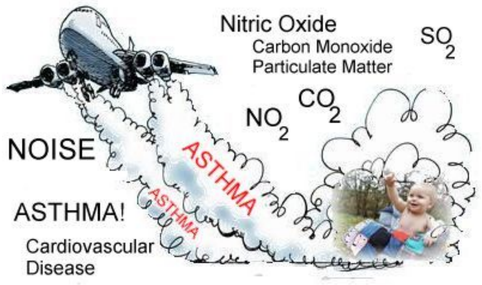
$41 Billion: ***Annual Economic Impact of US domestic flight delays to the national Economy***

The FAA considers a flight to be delayed when it is 15 minutes later than its scheduled time

MOTIVATION

While the recent air traffic growth has been beneficial to airline growth and airport network expansion, it has also gone in hand with massive level of aircraft delays on the ground and in the air. In response to growing concerns of fuel emissions and their negative impact on people's health, actual research is concerned with finding relevant techniques to predict flight paths and flight delays. Given the stochastic and volatile nature of these factors, predicting their impact on a given flight is not an easy task.

In this project, we compare two Supervised Learning algorithms to predict arrival delays: The **Decision Tree** algorithm and the **Neural Network** algorithm.



Main factors responsible for flight delays:

- weather
- Airport congestion
- Aircraft technical issues
- Air traffic management

DATA & FEATURES

The data comes from a publicly available Kaggle dataset, originally from Bureau of Transportation Statistics for the year 2015. The dataset consists of over 5 Million samples. We used the following features to training, validating and testing our models:

{Month, Day, Day of the week, Flight Number, Origin airport, Destination Airport, Scheduled departure, departure delay, taxi-out, distance, Scheduled Arrival}

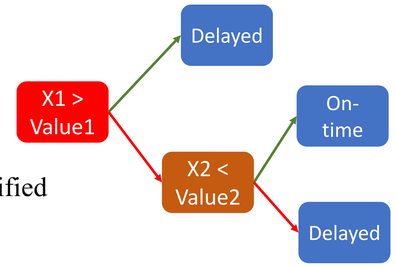
MODELS

Model 1: Decision Tree

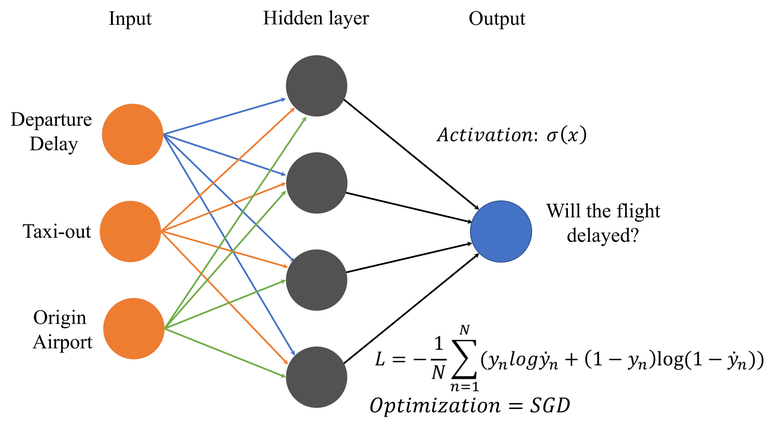
- Divide and conquer algorithm
- Measures the purity of subset, s
- Uses Shannon Entropy

$H(s) = \sum_{x \in X} -p(x) log_2 p(x)$

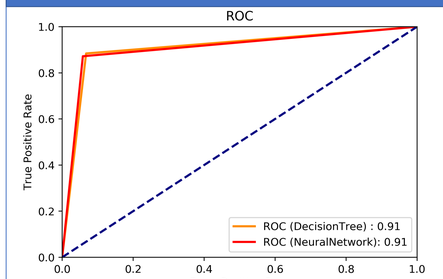
- *H(s) = 0, then s* is perfectly classified



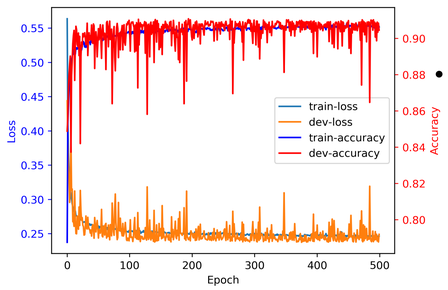
Model 2. Neural Network



RESULTS & DISCUSSION



ROC curves for DecisionTree and NeuralNetwork



Plots showing training and dev-set loss and accuracy during training

- Our final Decision Tree model was 7-level deep and contained 250 leaf nodes for training 70,000 examples.
- We find it very interesting that a simple single-layer Neural network with sigmoid activation function was able to achieve the same level of accuracy as the Decision Tree model. This suggests that a simple neural network can detect many possible interactions between predictor variables and also to detect complex non-linear relationships between dependent and independent variables.

RESULTS

Samples = 30,000		PREDICTED			
		Class 0 (On-time)		Class 1 (Delayed)	
		DescTree	NeuralNet	DescTree	NeuralNet
TRUE	Class 0 (On-time)	13,971	14,098	1,028	901
	Class 1 (Delayed)	1,718	1,945	13,283	13,056

Top 4 features by DescTree	DEPARTURE _DELAY	TAXI_OUT	ORIGIN_ AIRPORT	DISTANCE
Importance Score	0.8478	0.1438	0.0031	0.0028

- Both Decision Tree and Neural Network achieved training and test accuracy of approximately 91%

FUTURE WORK

Predict other flight delay types such as taxi-delays while considering airport runway and taxiway configurations, where very little work has been done. We can formulate predicting taxi-delays either as a multi-class classification or regression problem and use Decision Tree algorithm.

REFERENCES

[1] Cetek, C., Cinar, E., Aybek, F., & Cavcar, A. *Capacity and delay analysis for airport maneuvering areas using simulation*. Aircraft Engineering and Aerospace Technology, Volume 86,2014, Pages 43-55.
[2] Clewlow, R., Simaiakis, I., & Balakrishnan, H. *Impact of arrivals on departure taxi operations at airports.*
[3] Gopalakrishnan, K., & Balakrishnan, H. *A Comparative Analysis of Models for Predicting Delays in Air Traffic Networks.*
[4] Hamsa Balakrishnan, *Control and optimization algorithms for air transportation systems*, In Annual Reviews in Control, Volume 41, 2016, Pages 39-46, ISSN 1367-5788.
[5] https://www.transtats.bts.gov/ONTIME/Departures.aspx
[6] Khadilkar, H., & Balakrishnan, H. *Network congestion control of airport surface operations.* Journal of Guidance, Control, and Dynamics, Volume 37, 2014, Pages 933-940.
[7] Nogueira, K. B., Paulo H C Aguiar, & Weigang, L. *Using ant algorithm to arrange taxiway sequencing in airport.* International Journal of Computer Theory and Engineering, Volume 6, 2014, Page 357.
[8] Pyrgiotis, N., Malone, K. M., & Odoni, A. *Modelling delay propagation within an airport network.* Transportation Research Part C: Emerging Technologies, Volume 27, 2011, Page 60.
[9] Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.