Analysis of aircraft NAS delay using t-Distributed Stochastic Neighbor Embedding (t-SNE) techniques

Issoufou A. Fadjimata[1]
Embry-Riddle Aeronautical University (e-mail: issoufof@my.erau.edu)
Faculty Mentor: Radu Babiceanu[2] **(e-mail: babicear@erau.edu)**

INTRODUCTION

- The purpose of this study is to explore the many causes of National Airspace System aircraft delay, why it occasionally occurs, and predict the likelihood of delays using historical data and machine learning techniques.
- The hypothesis is:

1. In the event of an aircraft delay being abruptly announced by the airport flight information display, people are all set to take off for their preferred destination. This often impact and jeopardize travelers plans and can be costly to aircraft carriers.

2. The analysis and prediction of NAS delays using new machine learning algorithm is an important step towards better understanding the causes of delay and a prediction of their occurrence. The study can lead to a better planning and sustainability of the aviation and airlines industries and serve as a helpful tool for future travelers. Recently, delays attributed to NAS have negatively impacted the aviation ecosystem.

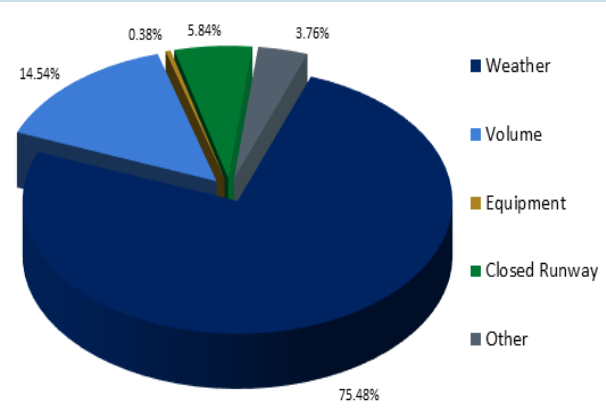

Figure 1. NAS major causes of Delay (FAA).

METHODOLOGY

t-SNE (t-distributed stochastic neighbor embedding) is a nonparametric data visualization technique used in classical machine learning. This is an unsupervised nonlinear technique mostly used for data exploration and high dimensional data visualization.

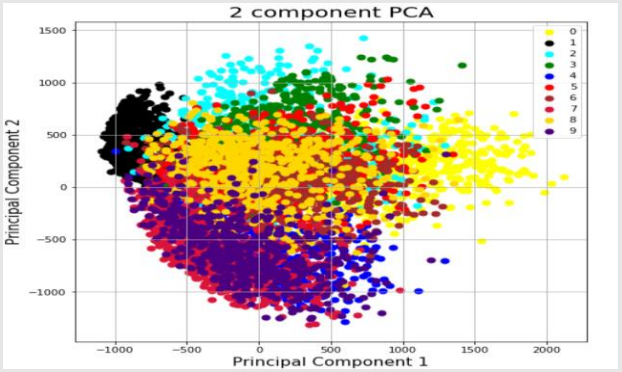

Figure 2. Example of t-distributed stochastic neighbor embedding technique by Renu Khandelwal

The study will use t-SNE algorithm to classify and analyze NAS delay parameters (non-extreme weather conditions, airport operations, heavy traffic volume, air traffic control) and predict future delays related to those parameters based on historical delay data. Results will be displayed in two-or three-dimensional space. Historical NAS delays data will be collected from the Bureau of Transportation Statistics (BTS) and FAA.

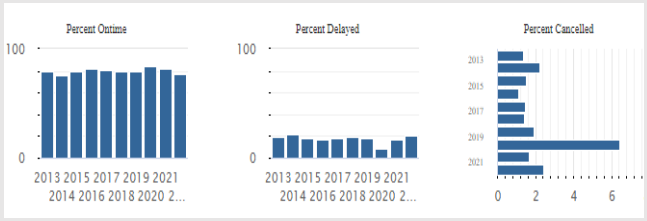

Figure 3. Reporting Operating Carrier Flight Delays at a glance (BTS)

SCOPE OF STUDY

Recent studies have described NAS as a significant contributor to aircraft delays. Weather parameters have also been attributed as the main cause of air traffic delays in the national airspace system. This study will seek new findings by analyzing all the parameters that cause delays in NAS, their frequency of occurrence as well as a prediction of future delays by causes.

Delay Cause by Year, Percent of Total Delay Minutes (%)

	Air Carrier Delay	Aircraft Arriving Late	National Aviation System Delay	Security Delay	Extreme Weather
(Jun-Dec)	26.3	30.9	36.5	0.3	6.1
2004	25.8	33.6	33.5	0.3	6.9
2005	28.0	34.2	31.4	0.2	6.2
2006	27.8	37.0	29.4	0.3	5.6
2007	28.5	37.7	27.9	0.2	5.7
2008	27.8	36.6	30.2	0.1	5.4
2009	28.0	36.2	30.6	0.1	5.0
2010	30.4	39.4	25.7	0.2	4.4
2011	30.1	40.8	24.8	0.1	4.1

Delay Cause by Year

Table 1. Delay cause by year, as a percent of total delay minutes (BTS)

CONCLUSION

Today most airlines give an estimate of departure time, but these things can be difficult to predict as many parameters such as weather, computer crash can escalate and what seems at first to be a glitch within the system can turn out to be a main cause for delay. Understanding delays and their causes can enlighten passengers and airliners. But the most important thing is to be able to have access to a tool for predicting delay with broader data and to be prepared in the event of their occurrence. This is why this study is important for the aviation system and the stakeholders involved.