Li Guan:

I reviewed two papers for our literature review. One is “Airline Flight Delay Prediction Using Machine Learning Models” by Dr.Tang. Another one is “Flight Delay Prediction Based on Aviation Big Data and Machine Learning” by Dr. Gui and his team. Both of the papers use machine learning models to predict airline delays and the decision tree has the best result from these papers.

What I learned is how to choose the data from the data set. From papers, they picked time of the day, date, weather condition, travel distance, start airport and destination airport to train the model, which could be the reference for our project.

Robin Sah:

Introduction:

During my study on predicting flight delays, I came across two different research articles that offered insightful information about this field. Both studies looked at several approaches to deal with the difficulties of predicting aircraft delays utilizing data analytics and machine learning methodologies.

In the first article titled "Flight Disruption Insights with Big Data Analytics," I found that the researchers used structured data from various sources like flights, airports, and weather. Their aim was to identify key reasons causing flight disruptions. Making sure the data was accurate and merging various data sources to determine the reasons for flight cancellations was one of their biggest challenges. The study was divided into three key sections: data engineering, data exploration, and connecting flight disruption prediction. They employed logistic regression and linear regression to categorize and forecast delays brought on by bad weather. Their model had a 30% accuracy rate for predicting interruptions. They also found that delays over 45 minutes have a big effect on connecting flights.

In the second article on "Flight delay prediction based on deep learning and Levenberg-Marquart algorithm", the researchers used a deep learning model to predict flight delays. They added the Levenberg-Marquardt (LM) algorithm to improve the model's accuracy. It was more difficult to estimate delays due to the data imbalance. They attempted two solutions to this. They discovered that fewer instances of non-delayed flights were preferable to more instances of delayed flights. They were more accurate in their predictions since they used the LM method and denoising autoencoders.

Lessons for our Project:

From these articles, I've realized that having balanced data is very important for making good predictions. It also seems to me that the accuracy of a model can be significantly increased by employing the appropriate methods. In our project, where I'm also trying to use Big Data to predict flight delays, I should ensure that the data is well-balanced. I also plan to experiment

with different algorithms to find the best one for our model.

Hengchuan Zhang:

I did literature reviews including new big data methods and machine learning methods. I also learned further about our targeting dataset, which is comprehensive and tedious. I attempted to analyze the most important and current problem faced by airlines and try to learn the latest research around the globe. Hopefully, our final project would contribute to the aviation society to some extent.