Nathan Augenbraun
Harshitta Gandhi
Chris Wozny
November 3, 2023
EECE5645
Project Proposal

Background: Flight delays and cancellations are growing increasingly over time due to both a degrading climate and poor airline practices. The U.S. Bureau of Transportation Statistics has kept track of on-time statistics for flights since 1987. They provide information like date, carrier, origin airport, destination airport, scheduled departure time, actual departure time, schedule elapsed time, and actual elapsed time amongst a few other features [1]. Just looking at data for Delta Airlines for flights departing from Logan International (BOS) Airport from January 1st, 2016 to December 31st, 2022 yields 111,442 samples.

Additionally, since the climate is degrading across the world there is a focus on collecting data to help analyze and predict meteorological events. One such data set collected information from January 2016 to December 2022 [2]. This data set contains 8,627,182 entries of weather event (snow, rain, fog, etc), start/stop time, precipitation in inches, airport code, etc.

Objective: The objective of this project will be to take the airline data collected by USDOT about major airports and carriers along with scheduling information and interpolating that data against the US Weather Events data set to attempt to model and predict when flights would be delayed and by how long based upon which origin airport, destination airport, carrier, scheduled departure time, and the weather event/precipitation at the origin and destination airport at that time. Employing parallelism for predicting flight delays is crucial as it allows for the efficient processing of vast datasets, and facilitates real-time responsiveness by distributing computations across multiple processors. It effectively manages the complex modeling needed for this task, ensuring timely and precise predictions for passengers and airline operators.

Data Sets: Citations for both data sets are included in the references section. The weather data is already tabulated in to a large comma-separated value file, but the USDOT data will need to be mined and downloaded via their active server page extended interface.

References

[1] https://www.transtats.bts.gov/ONTIME/Departures.aspx

[2] https://www.kaggle.com/datasets/sobhanmoosavi/us-weather-events