Akshar Shrivats

IB Computer Science HL 4B

Potter

3 March 2021

Project Description

**Link to Data Set**: <https://www.kaggle.com/usdot/flight-delays?select=flights.csv>

**Description of Problem**: Travelers around the world are faced with disappointment, anger, and confusion daily as their flights are delayed or cancelled due to adverse weather, maintenance issues, or airport logistics problems. My dataset tracks date, day of week, airline, origin and destination, departure time, departure delay, time until wheels up. Cancellation status, and the form of delay. I know that some of these factors will be more heavily weighted than others - specifically the day of week (the most busy flight days are Mondays and Thursdays) and type of delay (Weather, Late Aircraft, Airline, Security, or Air System). My KNN model aims to predict if a flight will get delayed based on factors including departure airport, airline, weather conditions, ATC status, and more.

**Motivation**: Recently, I have begun working toward my Private Pilot’s License and learned a lot about airport procedure and aviation protocol. Even for my current training flights, I have to generate full scale weather reports and often have to delay or cancel my flights due to adverse AIRMETS (AIRman's METeorological Information) or NOTAMs (Notice To Airmen). This includes information regarding everything from wind speeds and direction to thunderstorm warnings, unlit towers, and missile testing (Yes, I had to reroute a flight plan because of Fort Bragg!). Commercial flights have to follow the same briefing structure as any pilot, so it will be interesting to see how predictable a delay is and test it on my future training flights to determine if they will actually get off the ground,