CIS 662: Course Project

Instructions: This is a group project. You will work in teams of 4 students that will be assigned at random. There are multiple due dates corresponding to various activities. As a final document you need to upload a project report as a PDF file, your final predictions as a CSV file, as well as a PDF version of the python code you used including the outputs generated. As in the assignments, make sure the PDF documents are readable.

- **Team formation**: By Tuesday March 26 11:59pm decide on a team leader and that person sends an email with the names of all students in their team to me ngautam@syr.edu, Cc the TAs (Sanup: ssarabal@syr.edu and Shivakumar: shsuresh@syr.edu), as well as ALL the team members. Only the team leader needs to send an email but they MUST copy the others. Teams should have 4 students each..

- **Problem Description**: A few airports such as Chicago (ORD), New York (JFK), and Orlando (MCO) have multiple airlines that fly into Syracuse (SYR). In particular, for ORD consider United Airlines UA 1400 (departs 18:22) and American Airlines AA 3324 (departs 18:39); for JFK consider Jet Blue B6 2516 (departs 22:05) and Delta Airlines DL 5371 (departs 23:00); for MCO consider Jet Blue B6 656 (departs 08:55) and Southwest Airlines WN 1305 (departs 11:30). The goal is to predict 1-4 days in advance if the latter flight's arrival time into SYR would be early, on-time, or delayed conditioned on the status of the earlier flight. Also predict the status of the earlier flight.

- **Data**: We are not imposing any restrictions on the data you can use. One source of data is from the Bureau of Transportation Statistics `https://www.transtats.bts.gov/ontime/` but there could be others.

- **Additional Data**: You are welcome to read about the factors that cause airline delays. One factor is weather and there are many sites that provide weather predictions on an hourly basis for the next few days at both arrival and departure airports.

- **Final Predictions**: We will provide a CSV file with an empty column for predicting the arrival time (early, on-time, or delayed) of the latter flights AA 3324, DL 5371, and WN 1305. The other columns would be date (April 19-22 to be predicted on April 18), origin airport, and flight number. In addition, we will require predictions based on all 3 flight statuses of the earlier flights UA 1400, B6 2516, and B6 656. However, for the prediction accuracy we will only use the values corresponding to the actual statuses. In addition, the teams would have to provide predictions for the earlier flights.

- **Ground Truth**: This will be based what is posted on the website `https://www.flightstats.com/v2/flight-tracker/search` for all the flights. If the flight is more than 5 minutes early, we will call it *early*; if it is within 5 minutes (plus or minus) of the scheduled time, we will call it *on-time*; if it is more than 5 minutes late, we will call is *late*. There could be a discrepancy between the above BTS website and the flight-stats site, so if you are using BTS, you may want to calibrate. Also, note that the BTS site will use the operating carrier, so check flight-stats and it will mention the carrier name as well.

- **Methodology**: While we do not want to prescribe a method, our only requirement is that it **not** be based on neural networks or deep learning. Any approach using regression or classification that we did in class, as well as any statistical analysis including time series would be reasonable choices.

- **Deadlines and Expectations**: All are 11:59pm on

  1. Tuesday March 26, 2024: Email ngautam@syr.edu (with Cc to team and TAs) team information (names of team members and mention leader)

  2. Tuesday April 2, 2024: Email ngautam@syr.edu (with Cc to team and TAs) project status information (in one paragraph explain what you have done so far and what you have observed)

  3. Tuesday April 9, 2024: Email ngautam@syr.edu (with Cc to team and TAs) initial prediction as a CSV file for April 10-13

  4. Thursday April 18, 2024: Email ngautam@syr.edu (with Cc to team and TA) final prediction as a CSV file for April 19-22

  5. Thursday May 2, 2024: Upload on blackboard the same final prediction as a CSV file by adding the actual values observed, PDF file of report (describing what was done in the project), and PDF file of the code with output. **Every student must submit all three files on Blackboard so we can assign a grade. It is not enough if one member of a team uploads it.**

  Out of the 25 points for this project, each of the deadlines met are worth 1 point (0.5 points for late submission until the next deadline, and 0 points after that). Mark your calendars and instructions to be sure expectations and deadlines are not missed.

- **Report**: Your report must be well written so that you can use it to show what you did for the class, say to a prospective employer. You must explain in some detail so the graders can understand what you did without looking into the code. Also, you must cite sources of data and make references to any information you used (such as journal articles). Even though each team member submits a separate report, the reports for students in a team can be identical.

- **Grading**: Grading will be based on the creativity used in the approach taken and the justification provided in the report. Individual grades will depend on level of participation in the team. Any obvious errors in terms of the code will be penalized. While it is important to get accurate predictions, that will not be the main focus. However, as an incentive, *we will take the winning team to lunch* on Tuesday 4/23, Wednesday 4/24, or Friday 4/26.