CS 4824/ECE 4424: Final Project Proposal

Sam Lally, Inhan Park, Charlie Pullen, and John Shamory

Topic:

For our final project we plan to predict the weather condition from a set of categories using four different algorithms and comparing the results. The features we will use are latitude/longitude, date, time, humidity, pressure, temperature, and wind speed/direction.

Data Set Source:

Our data source is from Kaggle and it includes hourly weather data for 36 cities from 2012 to 2017. For preprocessing the data we plan to narrow the data down to the 30 cities within the US & Canada to prevent latitude/longitude outliers since the other 6 cities are in Israel. Also there are some data cells that are empty so those values will have to be removed before processing. The data can be found here (https://www.kaggle.com/datasets/selfishgene/historical-hourly-weather-data).

Machine Learning Algorithms:

Our problem requires the use of classification analysis to predict the categorically-valued weather condition.

1. K-Nearest Neighbors
2. Support Vector Machine
3. Random Forest
4. Multi-Layer Perceptron

Model Evaluation:

We will evaluate the models by tuning the four of them and then comparing the error rates between them to see which performs the best.

Team Member Roles:

Sam Lally – K-Nearest Neighbors development/tuning.
Inhan Park – Support Vector Machine development/tuning.
Charlie Pullen – Multi-Layer Perceptron development/tuning.
John Shamory – Random Forest development/tuning.