Rainfall Prediction

Project Deliverable 1, MAIS 202

1. Dataset

For this project, I will be using historical climate data from the Government of Canada[1] to predict precipitations in a city given the previous weather conditions of nearby[2] cities. This dataset contains extensive weather data about many Canadian cities. Daily reports can be found containing features such as minimum/maximum temperatures, total precipitation, and wind direction.

2. Methodology

i. Data Preprocessing

Given the reliability of the dataset, I don't expect there to be much preprocessing. I will mainly be on the lookout for any extreme outliers in the data that might skew my results.

ii. Machine Learning Model

I want to create a regression model predicting whether a given city will experience precipitations using the amount of precipitations and the direction of the wind from nearby cities for my model. Additional features from the dataset may be added if they are found to be beneficial to the model.

iii. Final Conceptualization

I will integrate my model within a program that takes an input from a list of Canadian cities as well as a date and outputs whether there will be precipitations that day. The input must be limited to a list given that there is no web API that would allow me to have access to all of the data without having to download it all.

---

[1] Government of Canada, (2019, September 19). *Historical Data*. Retrieved from climat.meteo.qc.ca/historical_data/search_historic_data.html

[2] Note: Despite being "nearby", there still has to be a significant distance between the cities to make it interesting.