COMP 496 PROJECT PROPOSAL

Concept:

Throughout my experience of living in Canada, one factor that has always influenced my life externally is snowfall. Snow looks beautiful for a while until it starts affecting the life of everyone. Roads get dangerous to drive on, it causes health hazards as it can get really chilly at times. Which is why, it is important to know when and how much it can snow, which will help in making more informed decisions about when to go outside and when to stay indoors.
It can influence people's vacation plans, work plans, or even personal trips.

How awesome it will be to just be able to predict and know how much it will snow beforehand. Similarly, we can be more informed about the level of sunshine that we will be receiving or the amount of rainfall, etc. If we could just know the days which are more prone to receiving precipitation, we can avoid making outdoor plans or even just help us keep safe by avoiding going out on the days when the weather conditions are extreme.

In this project, we will be performing a research-oriented analysis for the weather conditions of the City of Calgary on a given day, week or month by the help of Machine Learning.
Since I currently reside in the City of Calgary and I am personally related to this project, I will be collecting the historical data about Calgary's weather and will include different factors such as precipitation, sky conditions, temperature, rain, etc. and then I will model the data into a Machine Learning algorithm which will help me predict the outcomes of the weather in the future.

Focus:

The focus of this project will be to use the existing knowledge that I have learnt using Python programming language. I will be using majority of Python libraries such as Numpy, Pandas, Seaborn, and Matplotlib. Along with this, I will be using Machine Learning algorithms that I have learnt in the previous courses that I have taken. I will be using the library Scikit learn to perform ML modelling of data into the algorithm.

Content:

The content of this project will be predicting the weather conditions for the city of Calgary for a given day or a week. The factors that I will be focusing on while building my model and its prediction will include the Wind, Water, Rain, Precipitation, cloud condition and sky condition.
I will be importing the historical data of over 80 years and use that in my ML model. I will be using the weather data around the City of Calgary from the year 1960 to present.

Preliminary Design:

The design would follow a basic structure of a Jupyter notebook in Python. Cells would be used to write the code and then displaying the findings for instance- in the form of a table showing the top 5 entries and so on. After displaying the table, I will also be displaying the visualizations in the cells below by typing the code in python libraries- Matplotlib, and Seaborn. For Machine Learning modelling, I will be using Sci-kit learn and fit the data into the model by storing them in Pandas and Numpy arrays.

Planned Look and Feel:

The final report will have a weather poster in the beginning, which will give the feel of a weather page. Below that, the design structure that was discussed in the Preliminary Design section will be followed by showing the findings in the form of tabular structure, followed by visualizations.

I will also be working with visualizations comparing the actual data as well as the predicted data and showcase that in the form of a graph to display the difference between the values of prediction and real life which will also show the accuracy of the model.

Milestones and Timelines:

1-4 weeks: Project planning, documentation, and research about the data source and data cleaning

2-8 weeks: Learn about different Machine Learning Model and work with trying and choosing the best model that will fit the data and for the type of prediction building.
Learning about Numpy, Pandas, Matplotlib and Seaborn. Implement Numpy and Pandas to demonstrate Data Analysis on the data collected and show them in the tabular form in the Jupyter Notebook.

9-12 weeks: Fitting the data into the model and then outputting the values for our desired work. Perform an analysis and comparison between the actual data values that we used as input and comparing it with the actual prediction values. Work with Perform Data Visualization by using Matplotlib and Seaborn. Along with this, working on the documentation of the project and the visuals.

12-15 week: Preparing to present the project. Prepare a report which would present the final findings of the research and present it to the supervisor. Showcase why I chose the model that I chose and going over the steps that were taken to complete the project.

Suggested Evaluation Criteria:

Right Machine Learning practices used and the choice of algorithm suiting the data and the desired output to make predictions. Correct implementation of the Python language and its library. Valid representation of different factors such as Batsmen, Bowlers, and their stats in the form of table and visualizations.
Proper information being provided at each step in the form of Markdown cells which would present the fact about the finding and conclude the result, which would make the decision making easier.

The model should be able to predict the weather condition of a given day or a week series based on the historical data that we feed it. If the model does a good job of doing that, the project can be considered a success.

Importance of this research-

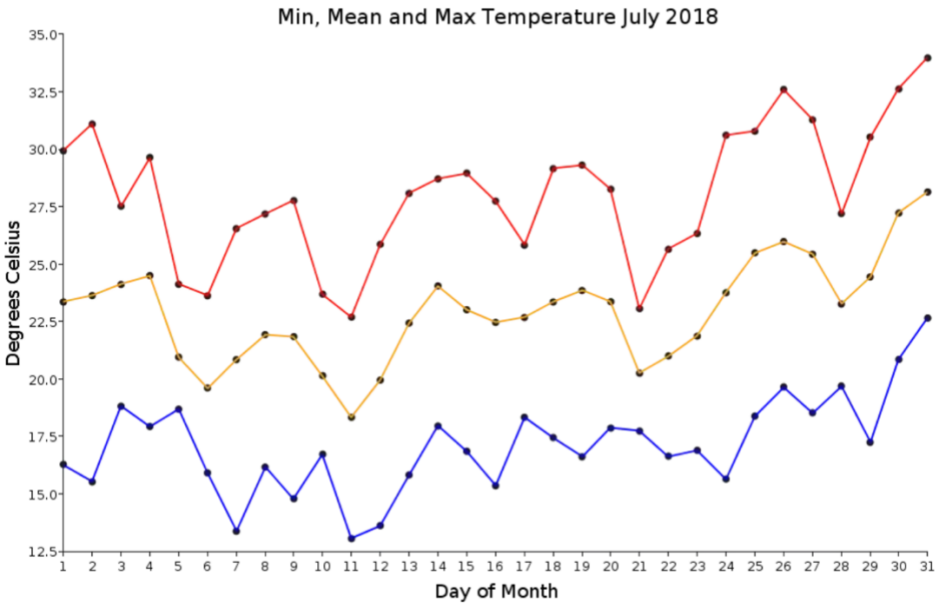
Weather conditions impact the lives of so many people and entities. For example, a sports league investing tons of money in conducting a league and then later the tournament gets disrupted due to the weather conditions such as snow or rain. This wastes a lot of money and also impacts the emotions of the fans.
With this research project using Machine Learning, informed decisions can be taken as this research model will tell the weather condition of a given day and then everyone can make more informed decisions by using those predictions.

Millions of dollars are being invested in making models that are more and more accurate with their weather prediction as it is such an important area. This project also performs research in a similar area.

Data Source:

The data that I will be collecting will be from an open-source platform which contains historical data for a given region, ranging back to decades. The source of the data will be: National Centres for Environmental Information

Data Visualization:



This is a sample data visualization after the model has predicted the weather values. The project that I will be working will have similar kind of interface and will show visuals of the predicted values in the form of graphs as well as tabular form.