Milestone 1: Project Proposal

Felipe Rodriguez

Bellevue University

DSC 680 Applied Data Science

Professor Amirfarrokh Iranitalab

April 11, 2024

Topic

Weather forecasting using ARIMA Model: Predicting Temperature Trends for Informed

Decision

Business Problem

A local company wants to understand weather data in London. Additionally, the company

wants to build a model to accurately forecast temperature. The data provided contains historical

weather data from 1979 to 2020 that will aid in the analysis and modeling. The model and

predictions will be used to plan outdoor events.

Datasets

The data is being obtained from Kaggle and is sourced from European Climate Assessment.

The data contains the following information: date, cloud_cover, sunshine, global_radiation,

max_temp, mean_temp, min_temp, precipitation, pressure, and snow_depth. The data will allow

for the analysis of climate based on the features provided. The main features that will be used for

modeling and predictions are mean_temp, precipitation, and snow_depth.

Methods

The methods used in this project will include various visualizations of various weather

features such as mean temp, precipitation, and snow depth. Additionally, a correlation matrix of

all variables to understand how the factors interact. Lastly, an Autoregressive Integrated Moving

Average (ARIMA) model will be created to forecast the weather in the future. The visualizations

will include various scenarios of the three fields above, which will include charts illustrating the

weather features. The correlation matrices that will be created will help understand the correlation between the various weather features.

Ethical Considerations

An ethical consideration is that this data should be used for the sole purpose of scheduling events for the company and should be shared for misuse to harm the environment in any way or form. Additionally, with the use of weather data, there is always uncertainty. Because of this uncertainty, it is important to communicate this to end users affected by the model.

Challenges/Issues

Using ARIMA model might not account for seasonality in the data, which can be a challenge, however a Seasonal Autoregressive Integrated Moving Average model can help with this if there is a lot of seasonality. Before creating the model, the data needs to be checked for any gaps in weather data which could impact the performance of the model. The model could contain outliers as well which could impact model performance.

References

The data used is sourced from European Climate Assessment (ECA). The ECA data gathers data from various meteorological stations to maintain this daily dataset. Using this set, a model can be created to forecast the weather. One thing to note is that the data contains historic weather data. By splitting this data into testing and training sets, we can confirm the accuracy of the model.