**Material and Methods**

**## Software**

Python and R/RSudio software are used to Analyse the data. Libraries and packages such as pandas, matplotlib, seaborn for Python and ggplot2, dplyr, caret for R are required in this analysis. RMarkdown, knitr are also utilized for putting the analysis together.

Scikit-learn is a machine learning library in Python, widely used in this analysis. The algorithms we used for forecasting such as Linear Regression, Multi-Layer Perceptron, Random Forest and XGBoost are all available in scikit-learn library.

For project management, cloud storage, version control and code collaboration GitHub gave us pro level access as a student.

**## Description of the Data**

**## Pre-processing Steps**

Part a and b of forecastdemand\_nsw.csv.zip were unzip and then concatenated into a single file.

The data type on the columns which contain the date time are not date time, therefore the column have been cast to datetime for better analysis. A binary column is utilised to indicate if the particular day is a public holiday or not.

**## Data Cleaning**

**## Assumptions**

**## Modelling Methods**

**### Models**

**#### Linear Regression**

**#### Multi-Layer Perceptron**

**#### Random Forest**

**#### XGBoost**

**#### Facebook Prophet**

Prophet is an another forecasting procedure for timeseries data. It is an open source library published by Facebook. It is based on additive model where non-linear trends are fit with yearly, monthly, plus holiday effects (Prophet, 2023). If there is strong seasonal effect in the timeseries data then with a few historical seasonal data it can perform better. It deals with outliers very well and robust to the missing data (Prophet, 2023).

**## Measures of forecast accuracy**

To measure the accuracy, we take the difference between actual and the predicted value by the model, which is also known as forecast error. The lesser the forecast error the more accurate the model. There are several accuracy measures. Depending on the problem's nature and the model's implications, selected accuracy measures are chosen. In this analysis the following accuracy measures are considered:

### Mean Squared Error (MSE)

MSE is the average of the squared difference between the target and the predicted value by the regression model. MSE is calculated by the following formula:

Since it squares the differences, it penalizes even a small error. The intuition of squaring the error is to make the large errors appear big. It is easy to calculate but sensitive to outliers.

### Mean Absolute Error (MAE)

It is one of the simplest accuracy measures, the mean of absolute difference between target and the value predicted by the model.

MAE is more robust to the outliers as it takes only the absolute value of the error. MAE does not penalize the error as extreme as MSE. So, when there are outliers in the data, MAE is preferable to use.

### Root Mean Squared Error (RMSE)

It is simply the square root of the mean of squared errors and calculated as:

RMSE is used to compare forecasting errors of different models for a particular dataset and not between the datasets (Wikipedia, 2023). RMSE is always non-negative and a zero indicates a perfect fit. The lesser the RMSE the better the model fits the data. This accuracy measure is sensitive to outliers.

### Mean Absolute Percentage Error (MAPE)

MAPE is a measure of prediction accuracy in a forecasting model (Wikipedia, 2023). MAPE is calculated by the following formula:

MAPE makes comparison of forecasting methods easier (and more useful) because working with percentage “standardizes” the errors. The time series’ original units no longer matter. MAPE is affected by outliers.

**References:**

# Wikipedia, the free encyclopedia 2023, ‘ *Root-mean-square deviation*’, accessed 3 Oct 2023, <<https://en.wikipedia.org/wiki/Root-mean-square_deviation>>

# Wikipedia, the free encyclopedia 2023, ‘ *Mean absolute percentage error*’, accessed 3 Oct 2023, <<https://en.wikipedia.org/wiki/Mean_absolute_percentage_error>>

# Prophet, Facebook Open Source 2023, ‘*Forecasting at scale’, accessed 4 Oct 2023,* <https://facebook.github.io/prophet/>