Neet Patel

Melbourne City - Sub-Team C

Documentation of the activities completed in Trimester 1 2021 for Sub-Team C

**D2I Melbourne City**

The project goal for the squad is to aid the client, i.e. City of Melbourne council in fostering a clean, green, and healthier community within the City of Melbourne. The client wishes to investigate ways to reduce emissions to progress towards Net-Zero Emission. The squad was divided into 3 sub-teams, to allow work to be completed at a granular level. This project was conducted in Trimester 1, 2021.

**Sub-Team C**

The members of Sub-Team C include:

* Neet Patel (Junior Student)
* Sivaram Krishnan (Senior Student)
* Nikita Wadekar (Senior Student and Sub-Team Leader)

The primary goal of Sub-Team C involved predicting the energy consumption for the City of Melbourne Council. Sub-Team C was responsible for completing the following project deliverables:

1. Obtaining the dataset from Kaggle
2. Update the dataset with the appropriate parameters till February 2021
3. Performing Exploratory Data Analysis on the dataset
4. Determining stationarity of the dataset
5. Develop and run multiple multivariate-time series models
6. Forecast values from March 2021 to August 2021
7. Develop a dashboard to present the predicted electricity demand
8. Incorporate the dashboard in the webpage

All aspects of the project deliverables were carried out successfully by Sub-Team C. However, there were some minor changes made to the project deliverables which are further discussed under the relevant iteration section.

The project was completed over 3 iterations:

* Iteration 1
* Iteration 2
* Iteration 3

The work completed under each iteration will be discussed under the relevant section.

**Iteration 0**

The first task in Iteration 0 involved obtaining the ‘Daily Electricity Price and Demand Data’ dataset from Kaggle (<https://www.kaggle.com/aramacus/electricity-demand-in-victoria-australia>). The Kaggle dataset consisted of 14 attributes and 2,016 observations. An observation was recorded at the end of each day. Subsequently, the next step involved extracting the known data from the Bureau of Meteorology (BOM) and updating the dataset for the attributes. This was completed as the Kaggle dataset was limited from 1 January 2015 – 6 October 2020. Hence, BOM was utilised to update the dataset values from 7 October 2021 to 28 February 2021. After extracting the known data and updating the dataset, important attributes with the dataset were analysed. To manage these tasks, each team-member was assigned specific tasks on the Trello board. It was unanimously decided with the sub-team to leave comments under the relevant card in Trello, detailing the work that was completed.   
After each team-member had successfully carried out their task, the next step involved dropping certain attributes such as ‘demand\_pos\_RRP’ and ‘RRP\_positive’ from consideration as they displayed weak correlation with other variables and were not important factors in being able to accurately predict electricity demand.

**Iteration 1**

At the commencement of Iteration 1, team-members were responsible for updating the ‘RRP’ and ‘Demand’ attribute from October 2020 to February 2021. These values were obtained from the Australian Energy Market Operator (AEMO). The trends with the electricity demand were visualised through multiple plots. Moreover, the dataset was also updated in terms of allocating values for the attributes ‘Public Holiday’ and ‘School Day’ from October 2020 – August 2021.   
Subsequently, the next task involved researching Exploratory Data Analysis (EDA) to obtain a better understanding of the dataset and determine the underlying trends within the dataset. Other work being done simultaneously included obtaining the summation of the electricity demand from verified data sources and predicting the RRP values using multivariate time-series models such as the Facebook Prophet model and VAR model. The trends in the RRP were also visualised by plotting multiple graphs. The final work in Iteration 1 involved determining the stationarity of the dataset through visualisations and by conducting the Augmented Dickey Fuller test. It was noticed that there were extreme outliers in the RRP dataset. As a result, these outliers were removed and replaced with the median RRP value. Discussions were held about researching several multivariate time-series models to predict the energy consumption. It was decided by the team to use values from January 20215 – December 2020 as the ‘Training data’ and January 2021 – February 2021 as the ‘Testing data’. The team specifically focused on tuning the parameters to obtain accurate performance metrics in terms of Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and R2 score.

**Iteration 2**

In Iteration 2, each team-member was allocated multivariate time-series models to research and implement on the dataset. Some of these models included: Prophet model, Vector Auto Regression (VAR), VAR with Moving average, Simple RNN Model, LSTNet Model, Long Short-term Model (LSTM), VAR moving average with Exogenous regressors. Out of these models, some models could not be implemented on the dataset due to their complexity/lack of resources, weak performance metrics and inaccurate predictions. As a result, these models were dropped from consideration. Out of these models, the SimpleRNN, Prophet and LSTM model were used in predicting the energy consumption from March 2021 – August 2021. Each team-member was allocated a model, and they were implemented to forecast the energy consumption. The model which showed the best performance in terms of the forecasted graph and performance metrics was selected to be the model to add to the webpage, and ultimately present to the client. Prior to iteration 2, it was decided that the Sub-team C would utilise ‘Tableau’ for visualisation. However, due to suggestions from Sub-team C members and other squad members, it was then decided that ‘Highcharts’ would be utilised to visualise the forecasts. Subsequently, the code for the forecasted graph was prepared using JavaScript. The forecasted graph was then integrated on the energy prediction webpage (<http://13.250.31.141:8080/Energy_forecast>).