A Data Science Approach to Forecast Electricity Consumption in Australia

Eugene Ho (z5497345) Data and Technical Specialist,

Majuwana Kariyawasam (z5398970) Data Specialist and AI Engineer,

Tariq Khan (z5414837) Data Specialist and ML Engineer,

Tom Woodley (z5450185) Group Leader.

03/09/2024

Table of Contents

[0.1 Introduction and Motivation 1](#_Toc176259158)

[0.2 Brief Literature Review 2](#_Toc176259159)

[0.3 Methods, Software and Data Description 3](#_Toc176259160)

[0.4 Activities and Schedule 3](#_Toc176259161)

[0.5 References 4](#_Toc176259162)

[0.6 Appendices 5](#_Toc176259163)

## 0.1 Introduction and Motivation

Forecasting energy demand in the Australian market is crucial to ensuring adequate supply for consumers and businesses. The Australian Energy Market Operator (AEMO) produces forecasts [1] over various timescales that allows energy producers and grid users to manage their outputs and bids for electrical supply. Energy demand is primarily influenced by temperature and other weather events. AEMO also tracks energy demand for “Shock Events” that create an unexpected amount of demand for energy in the grid. With the advent of more reliable forecasting of the El Nino-Southern Oscillation (ENSO) which governs much of the climate in the Eastern states during summer, it is proposed that incorporating an additional variable, along with a selection of weather metrics will improve the medium-term forecasts of energy demand.

**Could incorporating the phases of the ENSO improve electricity demand forecasts produced from weather observations and forecasts using General Linear Models, Convolutional Neural Networks, Extreme Gradient Boosting or Random Forest Regressors in New South Wales?**

The project investigates whether including ENSO phases in models using General Linear Models, Convolutional Neural Networks, Extreme Gradient Boosting, or Random Forest Regressors improves electricity demand forecasting in New South Wales. This focus on NSW is due to its population size and manageable data volume. The methodology could be applied to other states or the nation. The assumption that weather is the main factor won’t be examined, only whether adding the ENSO metric enhances model performance compared to existing literature.

The client for this project would be fossil fuel-using energy producers as an El Nino or La Nina event occurs over the course of a season and those producers (as opposed to battery or renewable producers) require a longer time period to ramp up production and procure fuels. For this reason, the forecasting scale will be monthly while data granularity will be hourly which represents two dispatch cycles in the grid.

## 0.2 Brief Literature Review

In recent years, numerous studies have explored the relationship between weather conditions and energy demand. For instance, studies have shown that temperature is a key driver of energy consumption, particularly in regions with extreme climates like NSW. Research has also highlighted the importance of accounting for seasonal variations and extraordinary events in demand forecasting models.

Many models rely on general linear models (GLMs) [2], [3], [4] which, while accurate, often struggle with higher-level interactions and non-linear effects. Machine learning models like Random Forest [5], [6] address non-linearities and seasonal influences effectively. More complex models such as Convolutional Neural Networks [7] and XGBoost [5] have also been successful, handling seasonality, trends, and non-linearities well. However, none of these studies integrate macro-seasonal factors like ENSO, presenting a gap in current models.

This work expands on previous research by incorporating ENSO projections and population growth into models. Linear models and advanced machine learning techniques will be used to address this knowledge gap.

## 0.3 Methods, Software and Data Description

### Data requirements -

Historical electricity usage data of different regions (provided by course staff)

Corresponding historical weather data that includes temperature, precipitation and humidity (some provided by course staff, some procured from the Bureau of Meteorology)

Population growth data from New South Wales (produced by ABS)

ENSO cycle data such as Southern Oscillation Index (SOI), sea surface temperature (still to source from US National Centers for Environmental Information and/or BoM)

### Technologies –

|  |  |  |
| --- | --- | --- |
| **Domain** | **Tool** | **Justification** |
| Communication | Discord, Microsoft Teams | Discord is useful for asynchronous messaging, Teams is useful for video calls |
| Collaboration | Github, Microsoft Teams, Microsoft Planner | Industry standard version control. Microsoft’s suite of software is useful for real time collaboration. |
| Analysis | Python, Pandas | Familiarity of language |
| Models | GLM, Random Forest, XGBoost, CNN | See literature review |
| Report writing | Jupyter notebooks, libtools library | Industry standard tools for Python reporting and referencing |
| Visualisations | Matplotlib and Seaborn | Industry standard Python visualisation libraries. |
| Presentations | Microsoft PowerPoint, Screencast-O-Matic, OBS Studio | Familiarity with PowerPoint. Recording software equally effective, OBS is FOSS. |

## 0.4 Activities and Schedule

Role names can be found at the top of this document.

Gantt Chart –

| **Activity** | **Week 1** | **Week 2** | **Week 3** | **Week 4** | **Week 5** | **Week 6** |
| --- | --- | --- | --- | --- | --- | --- |
| 1. Literature Review | X | X |  |  |  |  |
| 1. Data Assessment | X |  |  |  |  |  |
| 1. Algorithm Research | X | X |  |  |  |  |
| 1. Data Cleaning & Enrichment |  | X | X |  |  |  |
| 1. Data Summary & Integration |  | X | X |  |  |  |
| 1. Model Development |  |  | X | X |  |  |
| 1. Output Analysis & Recommendations |  |  |  | X | X |  |
| 1. Visualisation |  |  |  | X | X |  |
| 1. Report Writing |  |  | X | X | X | X |
| 1. Video Presentation |  |  |  |  | X | X |

Assignment of personnel (numbers in Assignments column correspond to the task in the Gantt chart)-

|  |  |  |
| --- | --- | --- |
| **Person** | **Assignments** | **Justifications** |
| Eugene | 1, 2, 4, 5, 6, 8, 10 | Excellent coding skills, works with high level visualisations |
| Majuwana | 2, 3, 4, 5, 6, 7, 10 | Repo owner, extremely experienced, familiar with CNNs as a model |
| Tariq | 3, 5, 6, 7, 8, 9,10 | Detail oriented, has experience with a range of ML models |
| Tom | 1, 3, 5, 6, 7, 9, 10 | Research background, project management |

## 

## 0.5 References

[1] Australian Energy Market Operator, “Forecasting Approach-Electricity Demand Forecasting Methodology,” 2022.

[2] R. Porteiro, L. Hernández-Callejo , and S. Nesmachnow, “Electricity demand forecasting in industrial and residential facilities using ensemble machine learning,” *Revista Facultad de Ingeniería, Universidad de Antioquia*, 2022, doi: https://www.doi. org/10.17533/udea. redin.20200584.

[3] H. Fan, I. F. MacGill, and A. B. Sproul, “Statistical analysis of drivers of residential peak electricity demand,” *Energy Build*, vol. 141, pp. 205–217, Apr. 2017, doi: 10.1016/j.enbuild.2017.02.030.

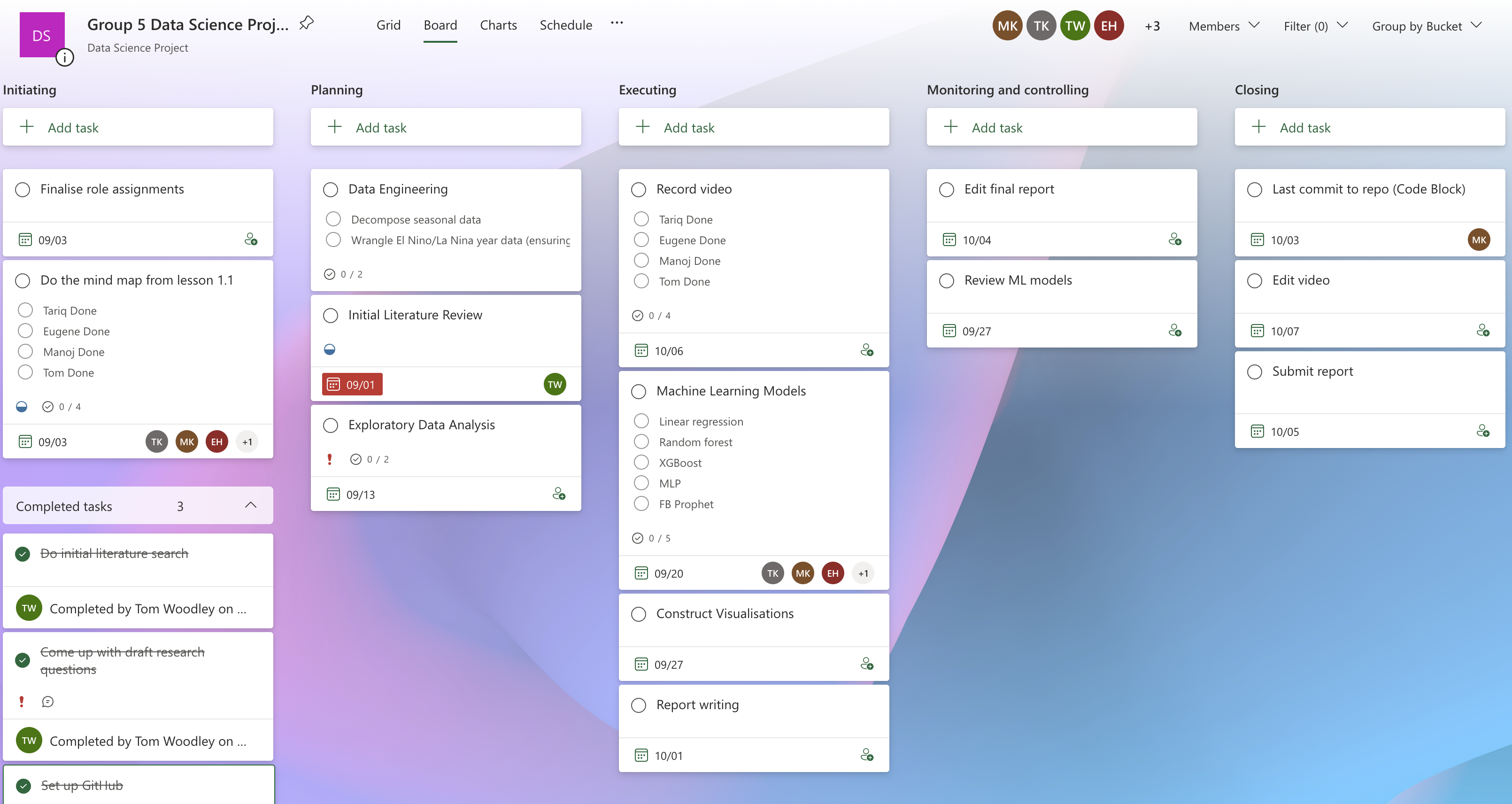
[4] A. Leung, “A Multivariate Model for Electricity Demand using Facebook Prophet,” 2022.

[5] G. Vijendar Reddy, L. J. Aitha, C. Poojitha, A. Naga Shreya, D. Krithika Reddy, and G. Sai Meghana, “Electricity Consumption Prediction Using Machine Learning,” in *E3S Web of Conferences*, EDP Sciences, Jun. 2023. doi: 10.1051/e3sconf/202339101048.

[6] J. Bedi and D. Toshniwal, “Deep learning framework to forecast electricity demand,” *Appl Energy*, vol. 238, pp. 1312–1326, Mar. 2019, doi: 10.1016/j.apenergy.2019.01.113.

[7] I. Koprinska, D. Wu, and Z. Wang, “Convolutional Neural Networks for Energy Time Series Forecasting,” in *2018 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Jul. 2018, pp. 1–8. doi: 10.1109/IJCNN.2018.8489399.

## 0.6 Appendices

*Appendix 1. Screenshot of scheduling software being used*