PROJECT PLAN BY GROUP E

A DATA SCIENCE APPROACH TO FORECAST ENERGY CONSUMPTION IN AUSTRALIA

Abdul El-Hamawi (z5019165), Chris Strods (z5329477), David Anderson (z5343521)

Jamie Twiss (z5353394), Shubhankar Dutta (z5304573), Sonal Chawla (z5092985)

School of Mathematics and Statistics
UNSW Sydney

March 2022

Abstract

The ability to forecast energy demand is essential to ensuring the stability of Australia's electricity grid, and therefore to supporting the Australian economy. In the past, this has been done through a combination of econometric (top down) and end-use (bottom-up) modelling approaches and operator intuition. Advances in computing make it such that more sophisticated approaches using artificial intelligence can improve the accuracy of existing models. Building on research already conducted on the use of artificial intelligence (AI) in forecasting, this project will build a new demand-forecasting model, using previous temperature and demand as inputs. A variety of methodologies will be used to seek to create a new model that can generate more accurate forecasts and demand scenarios.

Contents

Introduction and Motivation

Accurate forecasts of energy demand are important for accuracy and price stability. Outages or price spikes can have a sharp negative economic effect, as seen in the 1973 oil crisis (Kettell, 2020) when oil prices surged 300% in 5 months. The problem faced in trying to forecast energy usage is an extremely difficult one, (Craig, Gadgil, & Koomey, 2002) shows that long-term forecasts in the United States were more than twice the actual rates due to underestimating uncertainties in input data. This project aims to provide an energy usage forecast analysis that could be used by a business or market body that operates within the energy sector to add value or ease their decision-making process.

In recent years forecasting models that provide a single valued output for forecast energy demand has recently come under criticism due to their inherent inaccuracy caused by the chaotic nature of most standard model inputs i.e. temperature or weather data (Hong & Fan, 2016; Segarra, Ruiz, & Fernandez, 2020). Probabilistic load forecasting (PLF) is an approach that provides a range of different forecasts with their prescribed probabilities, this has been accepted as providing better utility and benefit to the end clients over a single point forecast model.

We will aim to solve the problem of creating an accurate energy forecast by utilizing various modelling techniques such as time series, regression, neural networks, and support vector machines. If these models do not perform sufficiently, we may consider pre-processing steps such as the Monte Carlo method to achieve better results.

Brief Literature Review

A variety of models including statistical and AI techniques are currently being used in energy forecasting. Some of which are:

- Regression of energy usage along with other factors such as temperature (Ruzic, Vuckovic, & Nikolic, 2003).
- Time series methods (McDonald & Fan, 1994; Cho, Hwang, & Chen, 1955) implement autoregressive moving average with exogenous variables (ARIMAX) for energy forecasting. This is the most used time series model as it can use temperature and time of day as inputs.
- Artificial neural network model have been developed (Bakirtzis, Petridis, & Kiartzis, 1996) using historic data to forecast several days ahead. Similar models (Xu, Wang, & Tang, 2019) create a PLF output (which is preferred) with best accuracies one day ahead.
- Support vector machines (Mohandes, 2002) in short-term forecasting and are shown to achieve superior results to an autoregressive method.
- The Monte Carlo method can be used to pre-process weather data to reduce its chaotic behavior and achieve higher forecasting accuracy (Zhao, Duan, & Liu, 2018; Fan, Liao, Zhou, Zhou, & Ding, 2020).

Simple models also exist such as (Codoni, Park, & Ramani, 1985) modelling energy demand in relation to consumer income and have been shown to be just as accurate as more complex ones at the time (Armstrong, 2001; Craig, Gadgil, & Koomey, 2002). In the scope of our project the integration of such methods into more complex ones could be particularly useful. They could also serve as an evaluation metric to compare to evaluation metrics of more complicated models.

After a review of relevant literature, there is potential for a forecasting model that utilizes an ensemble of older, simpler models along with newer machine learning models (and potentially the Monte Carlo method) to perform very well in short-term forecasting of energy demand.

We will require our model to have a PLF output rather than a single point, aiming at achieving the lowest possible discrepancy between predicted and actual energy usage (accuracy) as the ideal measure of model success (as used in all analyses included in this review).

Methods, Software and Data Description

To achieve our goal of developing an effective energy demand forecasting model, we will be analysing past temperature and energy usage data in Australia (along with any other data we can find that is relevant). The data was forked from GitHub and downloaded to individual machines. Once recombined and extracted there were twelve separate csv files containing each of temperature, forecast demand, and actual demand across four states (NSW, VIC, QLD, and SA). The forecast demand data is the largest by a wide measure as it contains multiple energy usage predictions for each DATETIME object.

Anomalies in the data include:

- A workaround was required to unzip large forecast demand files.
- Datetime data is formatted differently across states (YYYY-MM-DD HH:MM vs DD/MM/YYYY HH:MM:SS).
- Datetime data is not all sampled at the same (or regular) intervals.
- SA temperature file contains an additional index column.
- Some file names misspelt.

After resolving these issues, the data is simple, clean and does not include any missing values. The difficulty in working with the data from this stage will be in dealing with the huge amount of data points.

The data analysis will be done with; Python, Tableau, and R. Python and R will be used for data cleaning and modelling using various data and Machine learning libraries, such as pandas, scikit-learn, keras and others. Tableau is ideal for creating visualisations of the data and models which will be instrumental in the analysis. Other tools may be used if required.

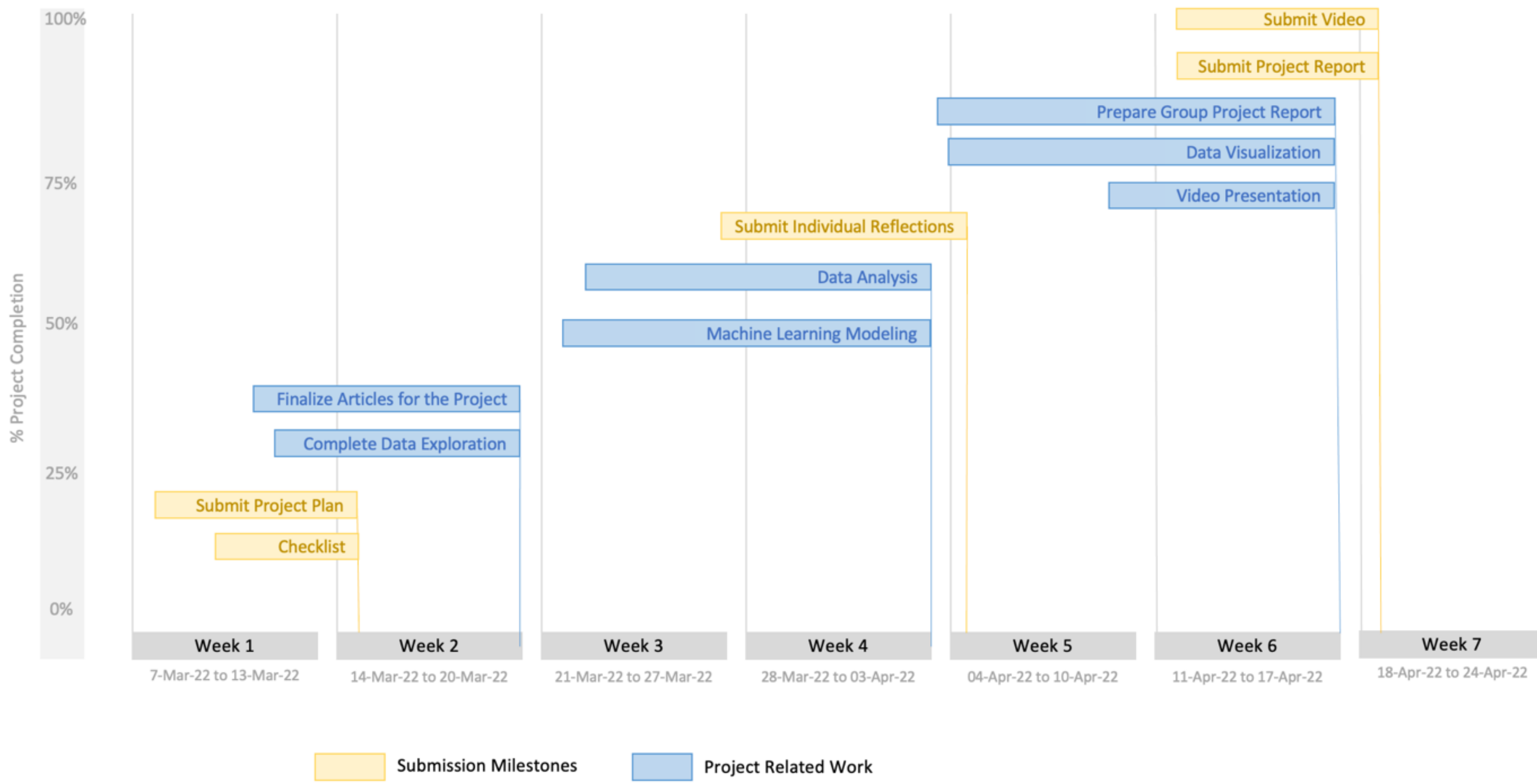
Activities and Schedule

The project will be conducted over a six-week period (of which the first week is already complete. The high-level project plan (see appendix 1) highlights the high-level work packages and the planned milestones of the project. The high-level work packages are then further subdivided into multiple activities, and a detailed plan (see appendix 2) is created using the high-level project plan as the baseline.

Note that the detailed plan will keep changing as the project moves forward; it is based on our current understanding of the necessary activities and the natural sequencing within those activities, as well as estimates of how long it will take to complete them.

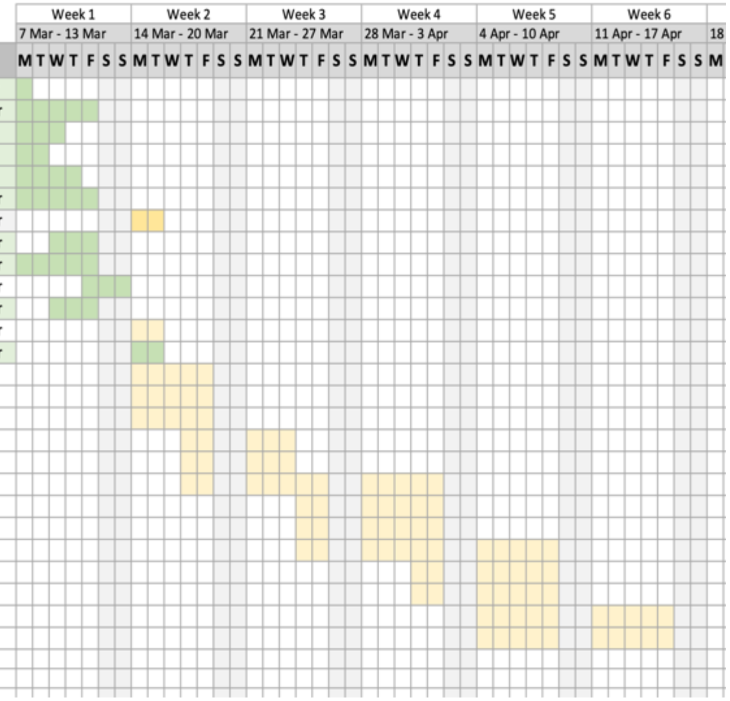
Need to include the roles of each team member.

Appendix 1: High-Level Plan



Appendix 2: Detailed Plan

TASK	ASSIGNED TO	Planned in We	PROGRESS	START	END
Team Introduction	Team	Week 1	Complete	6-Mar	6-Mar
Selection of Project Tools	Team	Week 1	Complete	7-Mar	13-Mar
Set up MS Teams	David	Week 1	Complete		
Set Up Project Planner	Shuba	Week 1	Complete		
Minutes of meeting	Team	Week 1	Complete		
Team Member Roles	Team	Week 1	Complete	7-Mar	13-Mar
Project Implementation Checklist		Week 1	In Progress	7-Mar	13-Mar
Project Gantt Chart	Shuba	Week 1	In Progress	7-Mar	13-Mar
Project Plan Report Template	Shuba, Jamie	Week 1	Complete	7-Mar	13-Mar
Project Plan Report Contents	David, Abdul,Jamie	Week 1	In Progress	9-Mar	14-Mar
Set up GitHub - Forked from the UNSW Project	Shuba, Chris	Week 1	Complete	12-Mar	13-Mar
Project Plan Submission	Jamie	Week 1	In Progress	7-Mar	13-Mar
Week 2 - Meeting Agenda	Shuba , Sonal	Week 2	Complete	13-Mar	14-Mar
Data Exploration		Week 2			
Research		Week 2			
Finalize Articles for Project					
Review Data Analytics Techniques					
Finalize Statistical Tools					
Perform Dimensionality Reduction					
Develop Machine Learning Algorithms					
Data Analysis					
Machine Learning					
Structuring of Report					
Develop Visualization					
Prepare Group Project Report					
Prepare Presentation Video					



References

Armstrong, S. (2001). *Principles of Forecasting.* Boston, MA: Springer.

Bakirtzis, A. G., Petridis, V., & Kiartzis, S. J. (1996). A Neural Network Short-Term Load Forecasting. *IEEE Transactions on Power*, 858-863.

Cho, M. Y., Hwang, J. C., & Chen, C. S. (1955). Customer Short-Term Load Forecasting by using ARIMA Transfer Function Model. *International Conference on Energy Management and Power Delivery*, (pp. 317-322). Singapore.

Codoni, R., Park, H., & Ramani, K. (1985). *Integrated Energy Planning: A Manual.* Kuala Lumpar: Asian and Pacific Development Center.

Craig, P., Gadgil, A., & Koomey, J. (2002). What Can History Teach Us? A Retrospective Examination of Long-Term Energy Forecasts for the United States. *Annual Review of Energy and the Environment*, 83-118.

Fan, C., Liao, Y., Zhou, G., Zhou, X., & Ding, Y. (2020). Improving cooling load prediction reliability for HVAC system using Monte-Carlo simulation to deal with uncertainties in input variables. *Energy and Buildings*.

Hong, T., & Fan, S. (2016). Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, 914-938.

Kettell, S. (2020). *oil crisis.* Chicago: Encyclopedia Britannica.

McDonald, J., & Fan, Y. (1994). A Real-Time Implementation. *IEEE Transactions on Power Systems*, 988-994.

Mohandes, M. (2002). Support Vector Machines for Short-Term Electrical Load Forecasting. *International Journal of Energy Research*, 335-345.

Ruzic, S., Vuckovic, A., & Nikolic, N. (2003). Weather Sensitive Method. *IEEE Transactions on Power Systems*, 1581-1586.

Segarra, E. L., Ruiz, G. R., & Fernandez, C. (2020). Probabilistic Load Forecasting for Building Energy Models. *Sensors*, 6525.

Xu, L., Wang, S., & Tang, R. (2019). Probabilistic load forecasting for buildings considering weather forecasting uncertainty and uncertain peak load. *Applied Energy*, 180-195.

Zhao, J., Duan, Y., & Liu, X. (2018). Uncertainty Analysis of Weather Forecast Data for Cooling Load Forecasting Based on the Monte Carlo Method. *Energies*, 190-199.