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 (AI) can improve the accuracy of existing models. Building on research already conducted on the use of AI in forecasting, this project will create 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 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. Forecasting energy usage has proven to be challenging, (Craig, et al., 2002) show that long-term forecasts in the United States were more than twice the actual rates due to underestimating uncertainties in input data.

In recent years, forecasting models that provide a single valued output for forecasted energy demand have come under criticism due to their inherent inaccuracy, resulting from the chaotic nature of most standard model inputs, such as temperature and weather data (Hong & Fan, 2016; Segarra, et al., 2020). Probabilistic Load Forecasting (PLF) is an approach that provides a range of different forecasts with their prescribed probabilities, and has been accepted as providing greater utility than a single point forecast model.

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

Literature Review

- A variety of models including statistical and artificial intelligence (AI) techniques are currently being used in energy forecasting, including: Time series methods (McDonald & Fan, 1994; Cho, et al., 1955) which 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 models (Bakirtzis, et al., 1996) using historic data to forecast several days ahead. These models (Xu, et al., 2019) create a PLF output (which is preferred) with best accuracies one day ahead.
- Support vector machines (Mohandes, 2002) which have been used in short-term forecasting and can achieve superior results to an autoregressive method.
- The Monte Carlo method which can be used to pre-process weather data to reduce its chaotic behavior and achieve higher forecasting accuracy (Zhao, et al., 2018; Fan, et al., 2020).

Simple models also exist such as (Codoni, et al., 1985) modelling energy demand in relation to consumer income, and have been shown to be similarly accurate as more complex contemporaneous models (Armstrong, 2001; Craig, et al., 2002). In the scope of our project the integration of simple models into more complex ones could be particularly useful. They could also serve as an evaluation metric to compare to more sophisticated models.

After a review of relevant literature, there is potential for a PLF forecasting model that utilises 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 energy demand forecasting. Our model will feature a PLF output rather than a single point, aiming at achieving the lowest possible discrepancy between predicted and actual energy usage (highest accuracy) as the ideal measure of model success - replicating the measures used in the previously reviewed analyses.

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 relevant data). The data was forked from GitHub and downloaded to individual machines. Once recombined and extracted, there were twelve separate csv files containing temperature, forecast demand, and actual demand across four states (NSW, VIC, QLD, and SA). The forecast demand data files are the largest, containing multiple energy usage predictions for each DATETIME object.

The following difficulties were encountered in data preparation:

- A workaround was required to unzip large forecast demand files
- Some file names were misspelt
- Some formatting issues within the data itself were observed. This can be discussed later as part of data exploration.

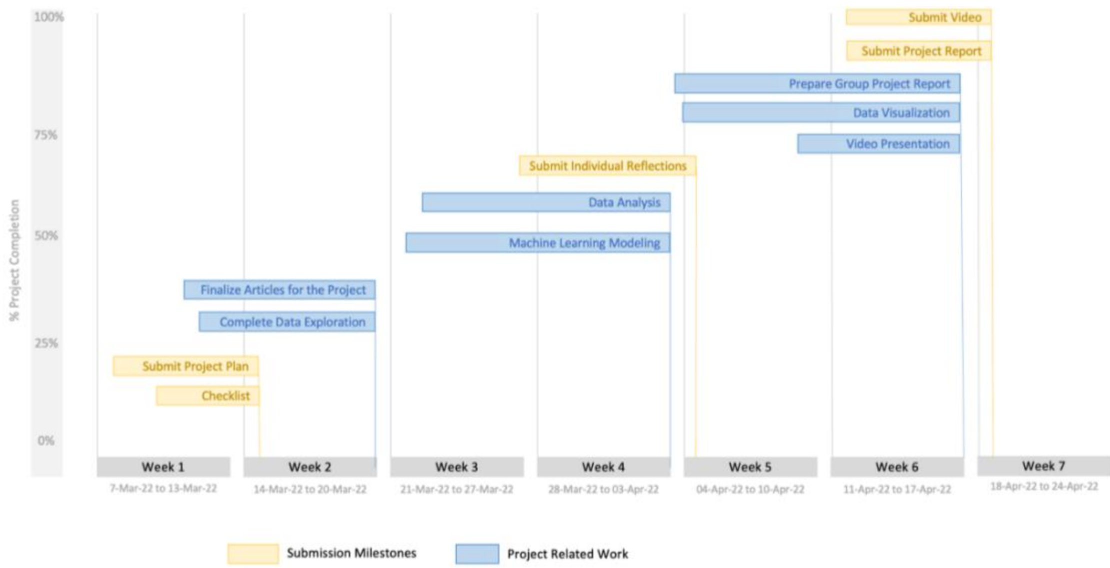
After resolving these issues, the data appears to be simple, clean and does not include any missing values. At this stage, the only challenge with the data seems to be dealing with the large amount of data points.

Data analysis will be performed through Python, Tableau, and R. Python and R will be used for data cleaning and modelling, implementing various data and machine learning libraries, such as pandas, scikit-learn, keras and others. Tableau will be used for developing visualisations for the purpose of interpreting the data and our model outputs. Other tools may be used if required.

Fundamentally, this is a windowed time-series analysis; the group will focus on short-term predictions based on recent data. We will test a range of models, with a particular focus on univariate vs. multivariate models.

Activities and Schedule

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



Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
7 Mar - 13 Mar	14 Mar - 20 Mar	21 Mar - 27 Mar	28 Mar - 3 Apr	4 Apr - 10 Apr	11 Apr - 17 Apr

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					

Table 1 states the roles and responsibilities. Through group discussions it was decided that each team member will have a primary role and act as a support for other roles.

Role	Primary Responsibility	Team Members
Project Lead	• Steer the project and coordinate with the team • Highlight red flags and keep track of tasks and activities • Create and allocate any missing tasks to team members • Update minutes of meeting and track action items • Serve as point of contact for any queries regarding team progress	Shubhankar Dutta (z5304573)
Data Scientists	• Search literature used as a reference for the project • Analyse the energy demand data • Develop and evaluate the forecast models to be used • Execute the models and compare if forecast matches with the demand	Abdul El-Hamawi (z5019165) David Anderson (z5343521)
Data Visualisation Specialists	• Determine the visualisation tool to be used. • Understand the data and develop visualisations required for the project.	Chris Strods (z5329477) Sonal Chawla (z5092985)
Communications Specialists	• Draft and finalise reports, templates and videos required. • Synchronise final documents in GitHub twice a week. • Take stewardship of the documents/report for submissions. • Ensure documents are correctly updated.	Jamie Twiss (z5353394) Shubhankar Dutta (z5304573) Sonal Chawla (z5092985)

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,* pp. 858-863.

Cho, M. Y., Hwang, J. C. & Chen, C. S., 1955. *Customer Short-Term Load Forecasting by using ARIMA Transfer Function Model.* Singapore, s.n., pp. 317-322.

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,* pp. 83-118.

Fan, C. et al., 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,* pp. 914-938.

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

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

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

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

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

Xu, L., Wang, S. & Tang, R., 2019. Probabilistic load forecasting for buildings considering weather forecasting uncertainty and uncertain peak load. *Applied Energy,* pp. 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,* pp. 190-199.