

# A Data Science Approach to Forecast Electricity Consumption in Australia

Christian Behan (z3376823)

Mia Jensen (z3460140)

Kim Dien Nguyen (Kevin) (z3318038)

Daniel Karp (z5338909)

13/03/2023

## Table of Contents

<b>0.1</b>	<b>Introduction and Motivation .....</b>	<b>2</b>
<b>0.2</b>	<b>Brief Literature Review .....</b>	<b>2</b>
<b>0.3</b>	<b>Methods, Software and Data Description .....</b>	<b>3</b>
	Methods.....	3
	Software .....	3
	Data Description .....	3
<b>0.4</b>	<b>Activities and Schedule .....</b>	<b>4</b>
	<b>References .....</b>	<b>5</b>
	<b>Appendix A: Dataset Variables .....</b>	<b>6</b>
	<b>Appendix B: Activities and Schedule .....</b>	<b>7</b>

## 0.1 Introduction and Motivation

Forecasting electricity demand in the short to medium term has been explored over the past few decades, with the majority of studies comparing the accuracy of different individual and ensemble model types. Although it has been shown that adding in weather and price data can improve the forecasts, it doesn't seem that any studies have shown to what extent each type of data has contributed to the model outcomes. By understanding the contribution of each feature type to the accuracy in the demand forecast predictions, we will be able to optimise the models and rationalise data collection. This will in turn mean only the required resources will be sought which may deliver ongoing savings in data collection and development effort and data storage.

With the motivation to optimise data collection, model code development and data storage whilst still delivering an accurate demand forecast solution, we have decided to explore what optimal combination of weather and pricing data will result in the most accurate electricity demand forecasts.

Our proposed solution is to develop an ensemble model utilising a combination of 3 of the most accurately reported unitary demand models from the literature to forecast electricity demand. This will be achieved using different combinations of weather and pricing data such that the effect of relying on each becomes evident. Any synergistic effects will also be noted.

## 0.2 Brief Literature Review

Electricity demand forecasting is utilised by both the electricity generation industry for capacity planning, and governments to inform energy policy. It has been discussed in the literature for at least the last 35 years with the focus changing from low-accuracy single method machine learning models to highly sophisticated multi-model ensemble approaches.

The current report concentrates on the impact of incorporating energy prices and weather forecast data on the accuracy of electricity demand forecasting. It draws on the early work of Tribble (2003) that found a strong relationship between ambient temperature and electricity demand, as well as other later contributors to the field such as Matsumoto and Misao Endo (2021) and Sgarlato & Ziel (2022) who rely on a variety of weather data - not just temperature, and Fatema, Kong & Gengfa (2021) who uses price data for demand forecasting.

We also used a selection of non-academic publications by the JWA (2023) and the US Department of Energy (2022) to better inform our understanding of the field.

There is a difference in the literature between short-term and long-term electricity demand forecasting, with the former being studied by Vilar, Cao & Aneiros (2012) who used non-parametric regression techniques to forecast next-day electricity demand, and the latter being studied by Sgarlato & Ziel (2022) and the US Department of Energy (Cox, de Silva, Jorgenson, & O'Neill, 2021), providing multiple day demand forecasts. These studies all incorporated the use of temperature and/or demand prices into their models.

Changes in electricity demand based on fluctuations in temperature and other weather phenomena have also been noted across the globe with studies by Zhang & Liao (2019) in China, Cassarino, Sharp, and Barrett (2018) in Europe and Manderson & Considine (2021) in the UK, though not all of these studies also incorporated electricity price as the independent variable.

The individual impact on the accuracy of the models when utilising variations of weather and pricing data is not readily seen in the literature. The current report aims to address this gap.

To achieve this, ensemble learning has been selected to forecast electricity demand, as it creates a more robust prediction by leveraging the strengths of more than one machine learning model (Wang, Mao, Wilamowski & Nelms, 2020).

### 0.3 Methods, Software and Data Description

#### Methods

The machine learning method we will use is ensemble learning. We aim to test a variety of forecast horizons. The Total Electricity Demand dataset contains historical datapoints of electricity demand, which needs to be cleaned and preprocessed, including handling missing values and encoding categorical variables. As we are interested in observing the effects of weather and energy prices, whilst taking into consideration historical forecasts, the Forecast Demand, Electricity Demand, Air Temperature, and Price and Demand datasets will be merged in the preprocessing phase. Feature engineering to identify key variables will then be performed. Each model in the ensemble will then be trained using the prepared data. Following training, the ensemble model will be cross-validated and tested using a subset of unseen historical data to measure model accuracy and reliability. Forecasting will be performed to produce predicted energy demand. Once we have a validated ensemble model, we will use it as a basis to compare its performance with variations of the weather and price data. Finally, we will compare the results of all models.

#### Software

We will be using Microsoft Teams for group discussions, instant messaging and conference calls. Project planning will be managed through ClickUp, allowing all members to collaboratively manage check lists, kanban boards, and Gantt charts in the cloud. We found that using cloud-based tools was optimal as the group members are in different states and countries.

Several of our group members are more confident with Python than R, hence we have opted to use Python to perform computational and visualisation tasks. The code will be stored and run in Jupyter Notebooks as it's a good resource for compiling both code and text in a markdown type environment, resulting in an aesthetically pleasing format. The Python packages we expect to use include:

- OS: for working directory navigation and management
- Pandas: for easier data handling. Tasks include cleaning, merging and feature engineering.
- NumPy: for easier handling of numeric datasets
- SciKit-Learn: for splitting data into train, test, and validation sets as well as performing various machine learning tasks i.e., regression, SVMs, decision trees, ensembles.
- Matplotlib: for building graphs for graphs/visualisation tasks
- Seaborn: for graphs/visualisation tasks

#### Data Description

Sourced from the Market Management System database, the Total Electricity Demand dataset contains total energy demand data for NSW between 2010 and 2022. The 1,323,398 rows have no missing values. See [Figure 1](#) for the dataset description.

Sourced from the Market Management System database, the Forecast Demand dataset contains forecast energy demand data for NSW between 2010 and 2022. The 11,619,503 rows have no missing values. See [Figure 2](#) for the dataset description.

The Australian Data Archive for Meteorology-sourced Air Temperature dataset contains air temperature data measured at Bankstown Airport weather station between 2010 and 2022. There are 247,646 rows and no missing values. See [Figure 3](#) for the dataset description.

Sourced from the Australian Energy Market Operator (AEMO) website, the Price and Demand dataset contains energy price and demand data in NSW between 2010 and 2022. There are 302,448 rows and no missing values. See [Figure 4](#) for the dataset description.

## 0.4 Activities and Schedule

A survey was conducted, within the team to identify the relevant skill sets and experience required for the project activities. Each team member completed a SWOT analysis of the technical and non-technical skill sets required, (see Appendix B, [Figure 5](#)) with the analysis results forming the basis for defining team member roles and responsibilities.

As can be seen in [Figure 6](#), certain team members have the knowledge and experience to be proficient in most of the required skills, and some have either a more technical or a more non-technical background. With the SWOT analysis results in mind, the team member roles and responsibilities were defined and are outlined in [Figure 7](#).

Christian was nominated as Team Leader and Project Manager for his experience with managing teams, and as his strength lies in non-technical skills like communication and business sense. Daniel's primary role is the project Data Analyst, however, as he is the most experienced team member, and is proficient in both technical and non-technical skills, he will support the Research Lead role. Given Mia's work experience, she will have similar roles and responsibilities as Daniel with primary focus as a Data Analyst and supporting the Research Lead. As Kevin does not have a technical background, but a keen interest in this area, his primary role will be Research Lead and he will be acting in a support role for the Data Analyst.

The full list of project tasks and activities held within the chosen cloud-based project management tool, ClickUp, was identified (Appendix B, [Figure 8](#)). The activities outlined cover all requirements for the successful completion of the project. A phased approach is described, where tasks are outlined in an orderly and structured manner. This is illustrated in the Gantt chart in [Figure 9](#). These activities include Data Cleaning, Exploratory Data Analysis and Modelling. When selected in the ClickUp portal, each activity contains a detailed summary, with supporting resources attached, and a checklist for the completion of that activity as can be seen in the example in [Figure 10](#).

Project activities were distributed among team members based on their assigned roles, to create efficiencies in task completions. For example, the Research Lead(s) will focus on the literature review activity whilst the Data Analysts commence the data cleaning and exploratory data analysis phases. Some activities require all team members to contribute, including the report writing phase. Each section of the report will, however, be assigned to the team member who has completed the work for that section through the designated checklist assignments.

## References

- Alea Business Software AleaSoft Energy Forecasting: AleaModel (2023) <https://aleasoft.com/aleamodel>
- Australian Energy Market Operator (2021, May). Forecasting Approach – Electricity Demand Forecasting Methodology. [https://aemo.com.au/-/media/files/stakeholder\\_consultation/consultations/nem-consultations/2020/electricity-demand-forecasting-methodology/final-stage/electricity-demand-forecasting-methodology.pdf?la=en](https://aemo.com.au/-/media/files/stakeholder_consultation/consultations/nem-consultations/2020/electricity-demand-forecasting-methodology/final-stage/electricity-demand-forecasting-methodology.pdf?la=en)
- Cassarino, T. G., Sharp, E., & Barrett, M. (2018). The impact of social and weather drivers on the historical electricity demand in Europe. *Applied energy*, 229, 176-185.
- Cox, J., De Silva, T., Jorgenson, J., & O'Neill, B. (2021). Load Forecasting for the Moroccan Electricity Sector (No. NREL/TP-5R00-77427). National Renewable Energy Lab.(NREL), Golden, CO (United States).
- Israt Fatema, Xiaoying Kong & Gengfa Fang (2021) Electricity demand and price forecasting model for sustainable smart grid using comprehensive long short term memory, *International Journal of Sustainable Engineering*, 14(6), 1714-1732, DOI: 10.1080/19397038.2021.1951882
- Japan Weather Association (2022). Electric Power Demand Forecasting <https://www.jwa.or.jp/english/service/electric-power-demand-forecasting/>
- Manderson, E., & Considine, T. (2021). The Effect of Temperature on Energy Demand and the Role of Adaptation (No. 2112). Economics, The University of Manchester.
- Matsumoto, T., & Endo, M. (2021). One-week-ahead electricity price forecasting using weather forecasts, and its application to arbitrage in the forward market: An empirical study of the Japan electric power exchange. *Journal of Energy Markets*, 14(3).
- Sgarlato, R., & Ziel, F. (2022). The role of weather predictions in electricity price forecasting beyond the day-ahead horizon. *IEEE Transactions on Power Systems*.
- Shah, I., Iftikhar, H., & Ali, S. (2022). Modeling and forecasting electricity demand and prices: A comparison of alternative approaches. *Journal of Mathematics*, 1DOI: 10.1155/2022/3581037
- Tribble, A. N. (2003). The relationship between weather variables and electricity demand to improve short-term load forecasting. The University of Oklahoma.
- Vilar, J. M., Cao, R., & Aneiros, G. (2012). Forecasting next-day electricity demand and price using nonparametric functional methods. *International Journal of Electrical Power & Energy Systems*, 39(1), 48-55.
- Wang, L., Mao, S., Wilamowski, B. M., & Nelms, R. M. (2020). Ensemble learning for load forecasting. *IEEE Transactions on Green Communications and Networking*, 4(2), 616-628.
- Zhang, C., Liao, H., & Mi, Z. (2019). Climate impacts: Temperature and electricity consumption. *Natural Hazards*, 99(3), 1259-1275.

## Appendix A: Dataset Variables

Total Electricity Demand		
Variables	Data Type	Additional details
DATETIME	Date time	Format: YYYY-MM-DD HH:MM:SS. Date range 2010-01-01 to 2022-08-01, all times 00:00:00.
REGIONID	Text	Categorical. Only one value: "NSW1". Can be removed.
TOTAL DEMAND	Float	Units: megawatts (MW)

Figure 1: Variables in the Total Electricity Demand dataset

Forecast Demand		
Variables	Data Type	Additional details
DATETIME	Date time	Format: YYYY-MM-DD HH:MM:SS. Date range 2010-01-01 to 2022-08-01, all times 00:00:00.
REGIONID	Text	Categorical. Only one value: "NSW1". Can be removed.
FORECASTDEMAND	Float	Units: megawatts (MW)
PREDISTPATCHSEQNO	Integer	Unique identifier of predispach run. 10 digits, format YYYYMMDDPP
PERIODID	Integer	Period count. Range 1 to 79.
LASTCHANGED	Date time	Format: YYYY-MM-DD HH:MM:SS. Date time range 2009-12-30 12:31:49 to 2022-07-31 23:31:52.

Figure 2: Variables in the Forecast Demand dataset

Air Temperature		
Variables	Data Type	Additional details
DATETIME	Date time	Format: YYYY-MM-DD HH:MM:SS. Date range 2010-01-01 to 2022-08-01, all times 00:00:00.
LOCATION	Float	Categorical. Only one value: "94766.". Can be removed.
TEMPERATURE	Float	Units: degrees celcius

Figure 3: Variables in the Air Temperature dataset

Price and Demand		
Variables	Data Type	Additional details
REGION	Text	Categorical. Only one value: "NSW1". Can be removed.
SETTLEMENTDATE	Date time	Format: YYYY-MM-DD HH:MM:SS
TOTALDEMAND	Float	Units: megawatts (MW)
RRP	Float	Units: Australian dollar value
PERIODTYPE	Text	Categorical. Only one value: "TRADE". Can be removed.

Figure 4: Variables in the Price and Demand dataset

## Appendix B: Activities and Schedule

Non-technical skills	Technical skills
Critical thinking	Ability to prepare data for effective analysis
Effective communication	Ability to write efficient and maintainable code
Proactive problem solving	Ability to apply math and statistics appropriately
Intellectual curiosity	Ability to leverage machine learning and artificial intelligence
Business sense	

Figure 5: Essential Skill Sets for Data Scientists



Figure 6: Team Skill Sets and Experience Venn Diagram



Team Member	Role	Responsibilities
Christian Behan	Team Lead /Project Manager	Submits work, administration, presentation. Team enabler, clears hurdles & makes actionable insights.
Daniel Karp	P: Data Analyst S: Research Lead	Data cleaning & visualising. Advises on statistical and modelling methods.
Kevin Nguyen	P: Research Lead S: Data Analyst	
Mia Jensen	P: Data Analyst S: Research Lead	Pushes the team to ask interesting questions. Knows the most about the topic.

Figure 7: Team Member Roles and Responsibilities \*P: Primary Role & S: Secondary Role\*

Tasks <span>NEW TASK</span>					
STATUS	TASKS	ASSIGNEE	START DATE	DUE DATE	PRIORITY
REVIEW	1 TASK				
	Group Project Plan <input checked="" type="checkbox"/> 2/4	MJ CB DK LN	Mar 6	Tomorrow	High
	+ New task				
IN PROGRESS	2 TASKS				
	Data Cleaning <input checked="" type="checkbox"/> 0/5	MJ	Tomorrow	Thu	Medium
	Literature Review <input checked="" type="checkbox"/> 0/5	KN DK	3 days ago	Sun	Medium
	+ New task				
OPEN	14 TASKS				
	Team Meeting <input checked="" type="checkbox"/> 0/5	CB		Today, 7pm	Medium
	DUE: A1A Group Project Plan (20%) <input checked="" type="checkbox"/> 0/2	CB	Tomorrow	Tomorrow, 5pm	High
	DUE: A1B Individual implementation checklist (5%) <input checked="" type="checkbox"/> 0/1	CB MJ DK LN	Tomorrow	Tomorrow, 5pm	High
	Lecturer Meeting <input checked="" type="checkbox"/> 0/2			Fri, 6pm	Medium
	Exploratory Data Analysis <input checked="" type="checkbox"/> 0/6	MJ CB DK LN	Fri	Mar 23	Medium
	Modelling <input checked="" type="checkbox"/> 0/12	MJ DK	Mar 23	Apr 2	Medium
	DUE: A2 Individual reflection portfolio (5%) <input checked="" type="checkbox"/> 0/1	CB MJ DK LN	Apr 4	Apr 4, 5pm	Medium
	Report writing <input checked="" type="checkbox"/> 0/9	MJ CB DK LN	Mar 31	Apr 10	Medium
	Peer marking (group) <input checked="" type="checkbox"/> 0/4	MJ CB DK LN	Apr 11	Apr 14	Medium
	DUE: A3 Group Project Report (50%) <input checked="" type="checkbox"/> 0/2	CB	Apr 15	Apr 15, 10pm	Medium
	DUE: A3 Peer marking (group) <input checked="" type="checkbox"/> 0/3	CB	Apr 15	Apr 15, 10pm	Medium
	Prepare Video Presentation <input checked="" type="checkbox"/> 0/6	CB	Apr 7	Apr 17	Medium
	DUE: A3 Video Presentation (20%) <input checked="" type="checkbox"/> 0/2	CB	Apr 18	Apr 18, 5pm	Medium
	DUE: A3 Peer Review <input checked="" type="checkbox"/> 0/4	CB MJ DK LN	Apr 21	Apr 21	Medium

Figure 8: List of Project Activities in ClickUp



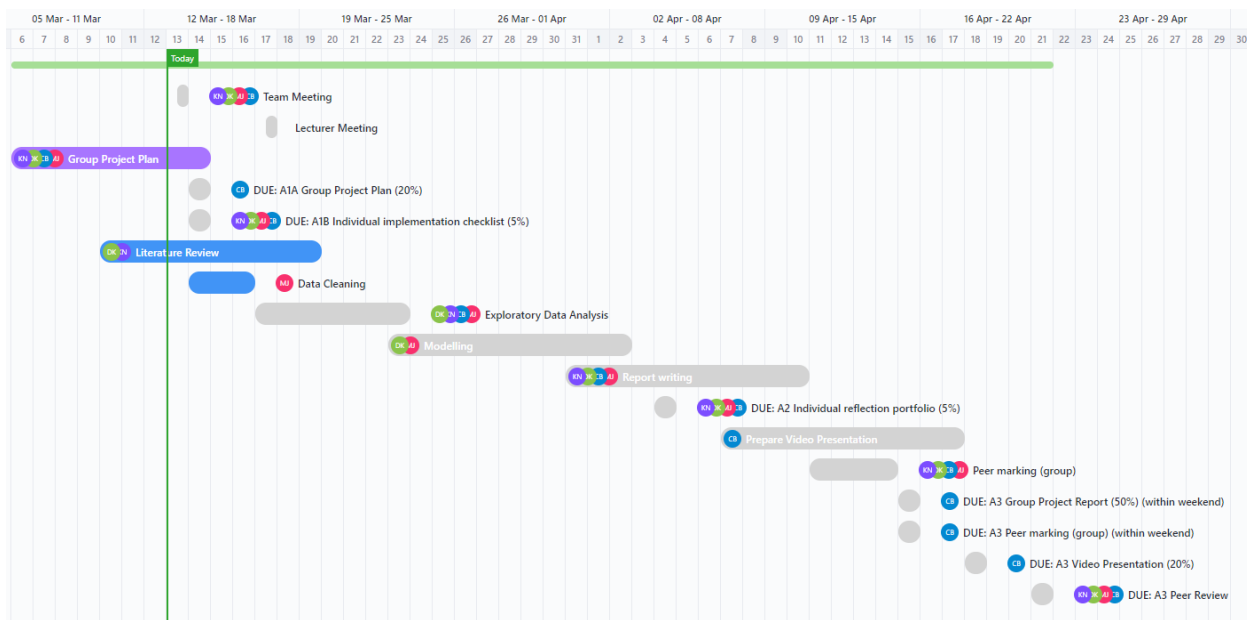


Figure 9: Project Activities Gantt Chart in ClickUp

Task 860q6fz8p

## Group Project Plan

Status

REVIEW

Assigned to

CB MJ KM DK

Dates

Mar 6 – Tomorrow

Priority

High priority

You must make contact with each member of your team in week 1 to write a 3-page plan establishing:

- An introduction and motivation to the project
- A brief preliminary literature review
- A brief initial outline of the methods, software, and description of the data you intend to work with

Show more

### Custom Fields

+

### Subtasks

+

### Checklists 2/4

Checklist (2/4)	
<input checked="" type="checkbox"/>	Introduction and motivation (20%) DK
<input type="checkbox"/>	Literature review (20%) KM
<input checked="" type="checkbox"/>	Methods, software and data description (20%) MJ
<input type="checkbox"/>	Activities and schedule (20%) CB
+ New checklist item	

Figure 10: Example Activity Resources & Checklist in ClickUp