Forecasting Electricity Demand in New South Wales, Australia Using Neural Networks and Machine Learning

Abdul El-Hamawi[1][z5019165], Chris Strods[1][z5329477], David Anderson[1][z5343521], Jamie Twiss[1][z5353394], Shubhankar Dutta[1][z5304573] and Sonal Chawla[1][z5092985]

[1] University of New South Wales, Sydney NSW 2052, Australia

Abstract. Electricity prices are highly sensitive to the available production and transmission capabilities. The ability to forecast energy demand is essential to ensure the stability of Australia's electricity grid, and therefore support the Australian economy. In the past, demand forecasting has been done through a combination of econometric (top down) and end-use (bottom-up) modelling approaches, as well as operator intuition. Recent advances in computing and technology have made it possible to integrate more sophisticated approaches using Artificial Intelligence (AI), particularly Neural Networks and Machine Learning which can improve the accuracy of existing models. Building on research already conducted on the use of these techniques in forecasting, this project will aim to create a new demand forecasting model, using past weather data (temperature) and past demand as inputs. The model will be benchmarked against an existing demand forecast dataset.

Keywords: Demand Forecasting, Neural Networks, Random Forest, Machine Learning, Time Series Prediction.

1   Introduction

1.1 Introduction

Energy is critical to the economic and social development of any country. A reliable mix of energy sources, available at stable prices, can significantly improve a country's economic outlook and even social stability (Knez, Šimić, Milovanović 2022).

Electricity is a key pillar of modern energy supply, as it can be delivered instantly over a wide area, and a wide range of items at all sizes and costs can be configured to use it. It can also be generated using a wide range of fuels, including renewable sources, making it an important element in the fight against climate change. (Sims 2004).

However, unlike many other forms of energy, electricity generally must be consumed the instant it is generated. It cannot currently be stored at large scales in any economic way. Therefore, a mismatch between the supply of electricity (which often takes hours to increase or decrease, by adding or subtracting generation capacity) and demand for electricity, which can change very quickly, is extremely costly. Indeed, electrical grids can suffer from something as simple as a widely watched television program ending, after which people at home start using electrical appliances at the same time (BBC, 2007).

Therefore, accurate short- and medium-term forecasts of electricity demand are critical for system stability as well as maintaining reasonably constant prices. These forecasts can be used by system operators such as the Australian Energy Market Operator, who can then direct energy generators or large industrial customers to act differently, as well as the generators themselves, who can add or remove capacity, or place capacity on standby (Wonhas 2022).

Given the importance of accurate forecasts, a huge number of studies on forecasting methods have been conducted since early 1970s. However, forecasting energy usage has proven to be challenging; long-term forecasts in the United States were calculated to be more than twice the actual rates due to underestimating uncertainties in input data (Armstrong 2001). Recently, 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 (AEMO 2020). Artificial Intelligence (AI) energy forecasting models have started to gain popularity because of their ability to process time series data and capability to deal with noise data. 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.

This report aims to create an accurate energy demand forecasting model based on AI techniques. In particular, the report will highlight two techniques – machine learning and neural networks. We will be looking at past electricity demand data and past weather data from New South Wales as potential inputs of the model. A more

concentrated dataset like the one in this report will enable us to determine the appropriate modelling technique without having to take other factors like location into consideration.

1.2 Intended clients and desired outputs

The predicted results from a demand forecast model are intended to benefit several clients. To identify these potential clients and their needs, we interviewed a former senior official at Australian Energy Market Operator (AEMO). Based on those findings, this report aims to target three classes of clients who might be interested in the model:

1. Market operators such as AEMO—forecasts ranging from one day to one week. They will use these forecasts to anticipate and avoid mismatches between supply and demand, for example by deferring power-plant maintenance or asking a large customer to go offline.
2. Generators—forecasts ranging from 0.5 to 8 hours. Coal generators can ramp up output when required, given notice of several hours; gas generators require a time frame that is generally shorter (and can be as short as 5 minutes).
3. Traders—forecasts of any time frame. Several parties trade electricity on the wholesale market, seeking to anticipate price movements. There are a variety of electricity contracts that make it possible to trade across many different time scales; the key factor is less the timescale and more the accuracy of the model, as trading is profitable only when the trader has better information than the market.

2   Literature Review

As electricity demand forecasting has attracted a widespread interest from several practitioners around the globe, there is a variety of approaches including statistical and artificial intelligence (AI) being used, including:

- Time series methods which implement autoregressive integrated moving average with exogenous variables (ARIMAX) for energy forecasting. (Bennet, Stewart & Lu (2014). This is the most used time series model as it can use temperature and time of day as inputs.
- Artificial neural network (ANN) models (Bakirtizis, Petridis & Kiartzis 1996) using historic data to forecast several days ahead. These models [6] create a PLF output (which is preferred) with best accuracies for one day ahead predictions.
- Support vector machines (SVM) which have been used in short-term forecasting and can achieve superior results to an autoregressive method (Mohandes 2002).
- A PLF forecasting model that uses an ensemble of older, simpler models along with newer machine learning models perform well in short-term demand forecasting (Mahdavi 2020).

Additionally, the Monte Carlo method has proven to be useful in pre-processing weather data to reduce its chaotic behaviour. The resulting model usually has higher forecasting accuracy (Hong & Fan 2016).

Simple models also exist such as modelling energy demand in relation to consumer income and have been shown to be similarly accurate as more complex contemporaneous models (Ghalehkhondabil et al. 2015).

To understand how the modelling accuracy varies across different methods, the approach will be to carry out two modelling methods on the same dataset after pre-processing – Machine Learning and Neural Networks. Moreover, our model will feature a confidence interval output rather than a single point with the aim to achieve the lowest possible mean average error between observed and predicted values in a test data set as the measure of model success - replicating the measures used in the previously reviewed analyses.

3   Material and Methods

3.1   Description of Data and Pre-processing

To achieve our goal of developing an effective energy demand forecasting model, we will be analysing past temperature and energy usage data in New South Wales. A larger dataset was available which contained information on four Australian states (NSW, VIC, QLD and SA), however we decided to concentrate on just one state (NSW) for the purpose of avoiding complexities due to large datasets and different locations. This will enable us to provide a sufficiently in-depth analysis and value for our identified clients of the project. Given trends shown in this report, we think it is reasonable to assume that the models outlined would be suitable to use on other state's data with minor changes and would provide similarly valuable outcomes to a wider range of clients.

The data was forked from the main GitHub repository provided and contained csv files with information on temperature, forecast demand, and actual demand. The forecast demand data files are the largest, containing multiple energy usage predictions (up to 79) for each DATETIME object.

The three datasets for NSW contain data from January 2010 to March 2021. In order to create a more relevant and meaningful dataset for data modelling, the team decided to use only data from 2016 to 2020.

There were two reasons for this. First, we excluded the 2021 data because it only had three months of data; therefore, any model looking for year-on-year trends would find that 2021 behaved very differently for previous years. This would essentially rule out using the year as a potential feature. Second, we excluded the 2010 to 2015 data because it showed a significant downwards trend that did not continue in subsequent years. We conducted research as to why this was the case, an academic paper stipulated that it was because of three factors: efficiency mandates from the regulator (primarily in household appliances), a structural change in the economy, and a response to a recent increase in electricity prices (Saddler, 2013). We concluded that these factors were not likely to occur again in the future in the same way. Thus, data from 2016 to 2020 strikes an optimal balance for us in having a large amount of data available to better capture trends with our models, and not including data that may negatively influence our model with historic trends that are no longer relevant.

The electricity demand and temperature datasets are merged to create the main dataset to begin modelling. The electricity demand dataset has data for every 30 minutes, while temperature data is more erratic with inconsistent DATETIME. We validated that the demand data was complete for our needs, used a left join to merge the datasets which populated the demand DATETIME with temperature data and filled any missing temperature values appropriately (if missing, filled with previous temperature value).

The forecast demand data for NSW was used as a benchmark for our models to ensure that the model we produce works at par, or better than the forecasting models that exist in the market today. The only pre-processing required was to extract the appropriate forecasts e.g. forecasts with PERIODID1 correspond to all forecasts done 30 minutes out from the observed demand and this forecast would be what we use to validate our models that predict on a 30-minute interval.

3.2   Assumptions

The following assumptions were made during the data preparation, data modelling and reporting stages of the project:

- While we have checked for obvious errors, we have otherwise assumed that the provided data is correct. This includes leaving any 'outliers' in the data as there are likely extreme demand values for some days in the real world.
- Temperature measured at a single point in Bankstown is predictive of demand for the entire state of New South Wales.
- The forecasting methods used to create the forecast demand dataset are unknown. We will assume that they are suitable for the data and use it as an evaluation criterion for our models.

3.3   Software

The software used for data pre-processing and data analysis were Tableau Desktop, Tableau Prep Builder, Python and R as they offer a wide range of functionalities to perform data exploration.

While we experimented with other approaches, we ended up deciding to build two separate models, a Neural Network (NN) model and a classic Machine Learning (ML) model.

To build our neural network model, we used Pytorch because it is flexible, faster, and provides capabilities to optimise our neural network model. It also supports GPU and CPU, enabling us to execute the model on Google Colab and our personal computers. In addition, we were able to use LSTM, GRU and other additional libraries of PyTorch to design our neural network.

Our machine learning models (random forest and SVM) were created in python using the scikit-learn package as it is user-friendly, lightweight and afforded the most functionality for our needs.

3.4 Modelling Methods

The team trialled several different approaches in order to identify the most promising models to build. The criteria we evaluated the different approaches on were:

- Predictive power. Based on what we know about different model types, and some preliminary modelling attempts, which approaches do we think are most likely to lead to effective models for delivering commercial outcomes (as defined in 1.2 above)?
- Suitability for the problem. Given that the project is a windowed time-series analysis, some approaches may not be feasible or realistic, such as the spline-fitting approach.
- Familiarity. Given the tight time constraints, we prioritized modelling approaches that the team was already familiar with, such as those taught in Machine Learning and in Neural Networks.

We divided the modelling team into two sub-teams, the first of which set out to build a neural network, the second to explore a random forest regressor and a support vector machine. We agreed that we would use mean absolute error (MAE) as our loss function consistently across all models to ensure that we were comparing them consistently. In a commercial context, we agreed that this enables us to manage our risk and to have a backup in case one model did not perform sufficiently or was not ready in time for when the client expected to see the results.

3.4.1 Neural Network

As shown in our literature review, neural networks have historically been used to accurately forecast energy demand, with best results being shown in forecasting up to a day ahead.

Neural networks are particularly useful when dealing with huge amounts of data that do not show clear relationships. They can train themselves to provide new, improved understandings of the data and provide high quality and accurate results. Neural Networks have also been very successful in forecasting models that use time series such as our data.

3.4.2 Machine Learning

In addition to the neural network approach, the team experimented with a classical machine-learning approach. Two models were developed – a random forest-based model, and an SVM model. Both random forests and SVMs are both known to be robust algorithms and have been known to produce good results in similar situations (Ghalehkhondabi 2015).

A random forest was chosen as a potentially suitable model in our situation for the following reasons:

- Good performance with data that has a non-linear trend.
- The range of values that the variables take are unlikely to extend far beyond the existing values in the training set.
- Good performance in situations when there is a departure from the normality assumption of errors.

Traditionally random forests are not suitable for time series as they are unable to identify or formulate growing or decreasing trends however, this restriction is overcome in our situation as the growing/decreasing trends in our data are not the prominent trends captured. The time aspect of our data still provides useful splitting conditions (on vs off-peak energy consumption).

An SVM approach was identified as appropriate in the literature review stage of this project. SVM's are suitable in situations where we have high-dimensional data and when the number of dimensions is greater than the number of samples. This is not quite the case here. They are also slow to train and extremely difficult to achieve a range of predicted values from. Nevertheless, we thought it would be beneficial to attempt modelling to see how well we could train an SVM model due to how well sited they are in literature.

The goal of our Machine Learning models was to predict energy demand 30 minutes out with associated confidence intervals. A benchmark MAE of 62 was calculated as being achieved by the forecast demands provided.

4   Exploratory Data Analysis (EDA)

The purpose of data exploration is two-fold. First, it enables understanding the data at a high level, including the variability in both the target and input variables. Second, it serves as a first step towards feature selection, identifying how the shape of the target variable changes under different values for each of the input variables.

As a first step, the data was reviewed for completeness and consistency. As noted above, the data is relatively clean, with few anomalies. Some data files had index columns while some did not, but there was minimal missing data.

The next step was to understand the overall variability of the data and the importance of various attributes with the help of data visualisations. This, in turn, was a critical step for feature selection.

As the two main variables we had were temperature and time (leaving aside the forecast data, which as discussed in 3.1 above, we will use as an evaluation metric), we looked at scatterplots and boxplots linking actual demand to those variables. We expected some signs of cyclicality of demand across individual days, weeks, and years, and so the data was analysed for different aspects of time. Boxplots were used rather than numerical summaries as they capture the same data in a way that is more visually appealing and easier to understand/compared to other variables.

The first check was to see the relationship between temperature and demand. The team started by exploring the entire dataset (data from all 4 states – NSW, QLD, SA and VIC). Figure 1 shows the total demand for electricity against temperature for all years (note that demand is not adjusted here for population).
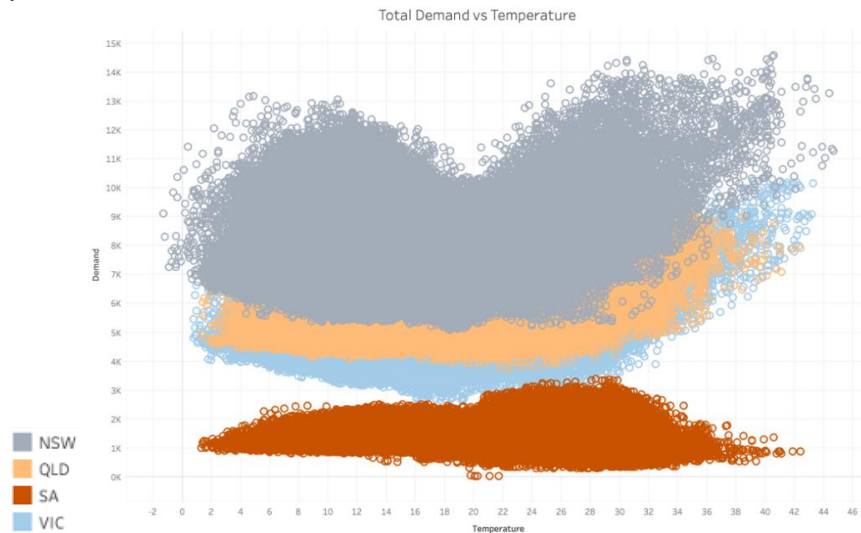

Figure 1: Total demand vs. temperature

From Figure 1 we can see that SA shows a less significant pattern while NSW and VIC are displaying the pattern anticipated from the project brief (discussed above). QLD appears to have lower peaks at extreme temperatures. Based on our external research, we believe that the SA pattern is weaker due to higher industrial use of electricity, which is not weather-dependent, while QLD shows fewer peaks due to, first, better passive cooling of homes (e.g. porches, shading, stilts) and therefore partially reduced reliance on air-conditioning, and, second, lower variability in humidity, which is also a significant driver of air-conditioning use.

Focussing on the behaviour of the data over time, Figure 2 shows the fluctuations over the years in total energy demand and temperature across the 4 states.
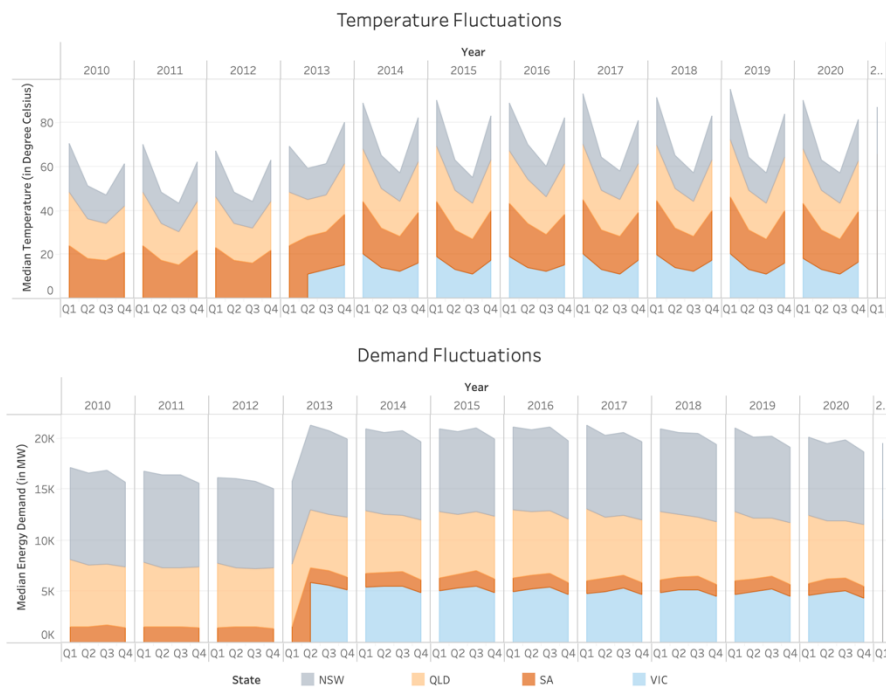
Figure 2: Temperature and demand across the different states

It appears that data for VIC over 2010-2013 is missing for both variables. It can also be noted that the data has a clear cyclical pattern in each state, with clear peaks and troughs.

We will now look at specifically NSW demand across different time scales. We started at the smallest possible scale, demand across time within a single day in Figure 3:
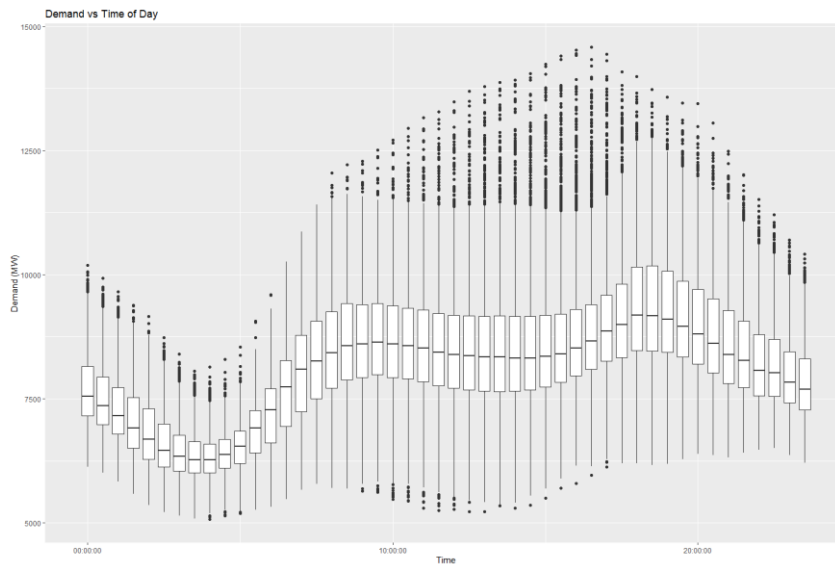
Figure 3: Relationship between energy demand in NSW and time of day

As we can see, demand is lowest overnight, building up to a first peak in the morning, then dipping slightly before building to a second peak in the evening, and then falling again.

We then stepped back to days of the week. Figure 4 shows relatively consistent demand on weekdays, but lower demand on weekends:
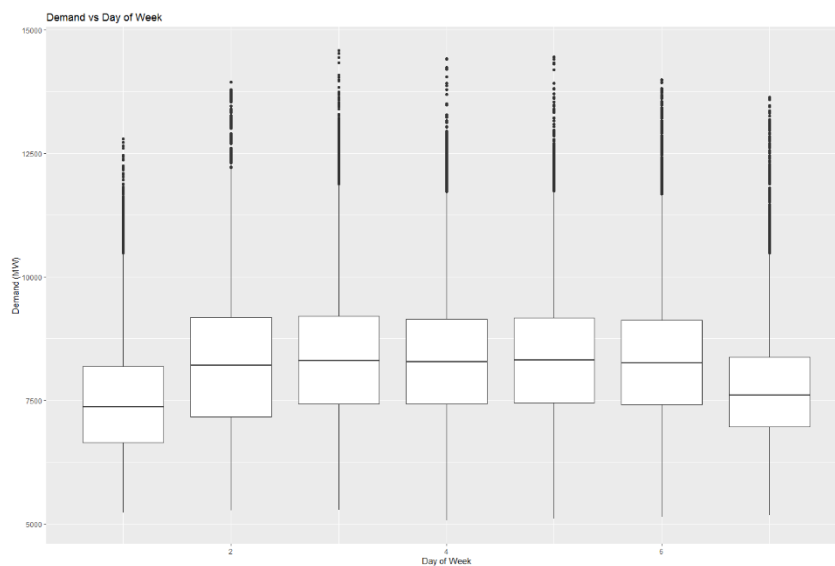
Figure 4: Relationship between energy demand in NSW and day of the week

Furthermore, we looked at demand over the course of the year. We did this both by week (Figure 5) and month (Figure 6).
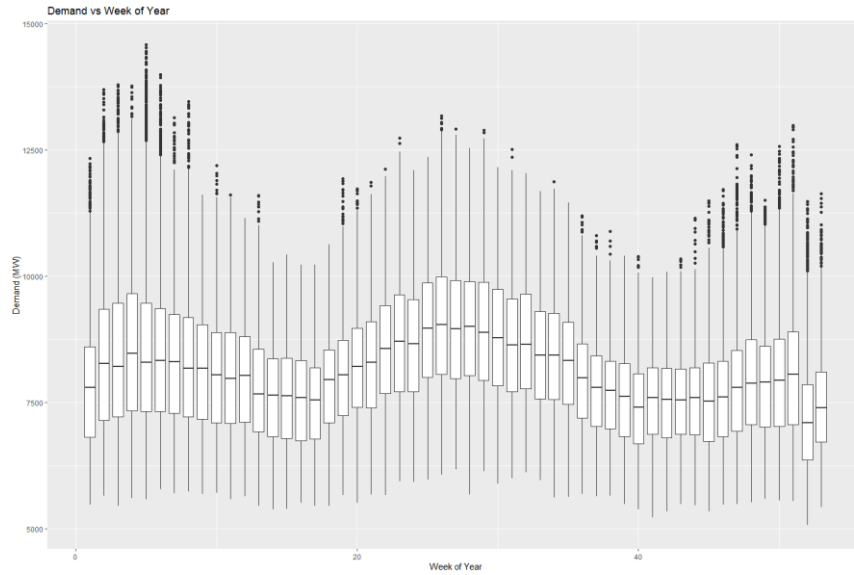
Figure 5: Relationship between energy demand in NSW and week of the year
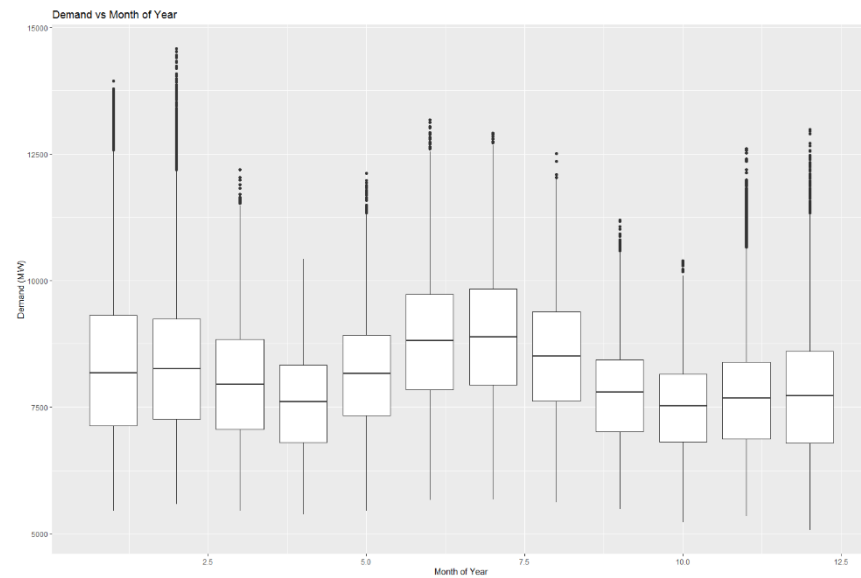

Figure 6: Relationship between energy demand in NSW and month of the year

The weekly view in Figure 5 results in sharper seasonal trends. As expected, due to the correlation between temperature and demand shown in Figure 6, demand was highest in the summer, followed by the winter, with spring and fall noticeably lower. Interestingly, winter months had the most volatile demand peaks.

The importance of all of this is that it will help us understand which features are potentially important to include in our models, as well as what data to use. Based on these charts, the group agreed that temperature, time of day, day of week, and month of year showed promise to be potential features.

Figure 7 shows the correlation between these variables and demand. The correlation between variables shows us how useful each variable is in any models that rely on linear dependencies.
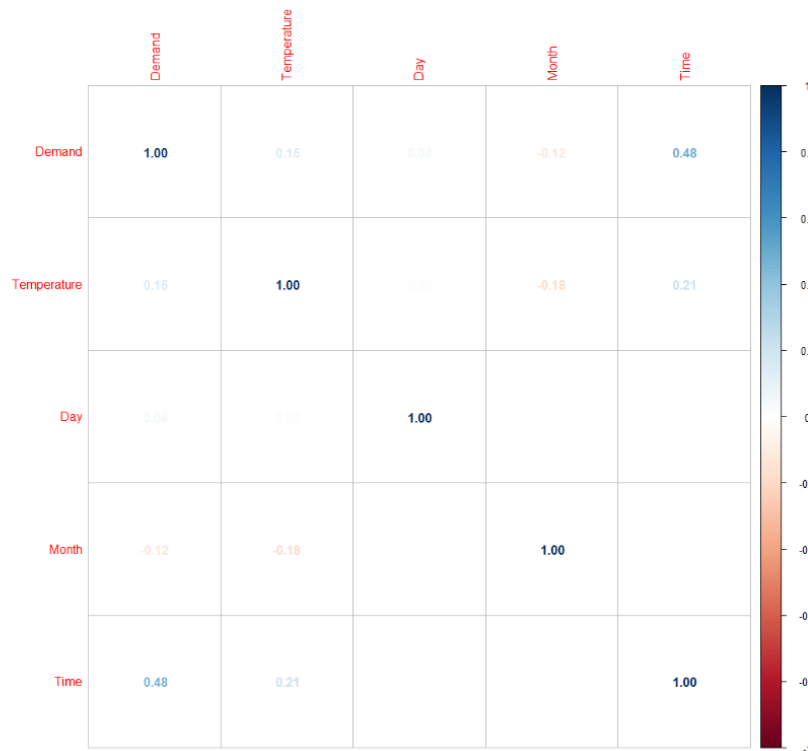

Figure 7: Correlation between demand, temperature, day, month and time of day

4.1   Feature Selection

4.1.1   Neural Network Feature Selection

Features for the neural network model were selected to be temperature, date (split into day and month), and time of day. These inputs were chosen as during our data analysis it was found that there was a correlation between the electricity demand and date/time (Figure 7).

The demand changed differently depending on the month in year, intuitively because different seasons result in different usage. The time of day also affected the electricity demand, as for example the night hours generally saw less demand than peak work hours in the middle of the day. The day was also considered as the electricity demand varied depending on the day, for example a weekday to a weekend.

4.1.2   Machine Learning Feature Selection

Feature selection was a very important part of creating a good Machine Learning model. The variables identified above as potential features were:

- Temperature
- Time of day
- Day of week
- Month of year

To overcome complexities and uncertainties with forecasting temperature and using it as an input, we decided to use historical temperature data from 30- and 60-minutes. This simplifies our model since we do not need to include a temperature forecasting feature and leads to negligible (if any) loss in model performance as temperature does not fluctuate greatly across a 30-minute interval and temperature predictions are unreliable in nature anyway.

We thought it would also be useful to look at including historical demand 30- and 60-minutes prior as an input.

Since a random forest regressor chooses ideal splitting conditions based on information gain, we determined the best way to determine initial features would be to look at mutual information. Table 1 shows the features and their dependencies.

Feature	Dependency Between Feature & Demand
Time of day	0.40
Demand 30 mins prior	1.85
Demand 60 mins prior	1.23
Temperature 30 mins prior	0.16
Temperature 60 mins prior	0.16
Weekday (0 for no, 1 for yes)	0.04
Day of the week	0.00
Month of the year	0.12

Table 1: Potential features' dependency on energy demand

From table 1 we can eliminate weekday, day of the week, and month from our features. We also believe that temperature does not change significantly from 30 minutes to 60 minutes from prediction and its identical dependency indicates extremely high correlation between the two variables, meaning that we can eliminate temperature 60 mins from prediction as an input without changing the model performance.

This gives us the following list of features:
- Time of day
- Demand 30- minutes prior
- Demand 60-minutes prior
- Temperature 30-minutes prior

5   Analysis of Results

5.1   Neural Networks

Starting with a simple neural network with two layers, each containing 200 nodes, the team confirmed that this approach had merit, with a Mean Absolute Error (MAE) of 764. This model simply predicted the temperature using time and date data, attempting a simple point forecast. This was promising enough to then consider more complex neural networks, like **Recurrent Neural Networks (RNN)**, which are commonly used for sequence prediction models and time-series problems.

First, a simple model was built which had an RNN, followed by a single output with 16 hidden nodes (diagram below). This model did not prove very powerful at predicting electricity demand using the above-mentioned inputs, with MAE errors higher than the simple 2-layer NN first attempted. However, it proved as a worthy starting point to delve deeper into more complex RNN architectures. This was promising enough for the team to then implement a structured approach to experimenting with different RNN architectures, building a model that could easily be configured, as shown in Figure 8.
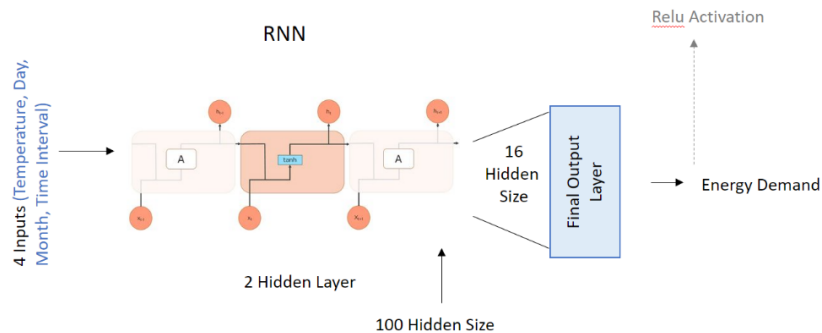

Figure 8: RNN architecture

Further, a more structured python script was written that allowed the configuration, testing and tuning of 4 RNN networks. With this script, by simply changing a few variables different architectures could be tested and tuned with their hyperparameters. Below are the four RNN architectures designed and tuned to determine the best Neural Network Model.

- Neural Network with one **Long Short-Term Memory** Network (LSTM Network) one fully connected linear layer and one output Layer
- Neural Network with one LSTM Network, two fully connected linear layers and one output Layer
- Neural Network with one **Gated Recurrent Unit** network (GRU Network), one fully connected linear layer and one output Layer
- Neural Network with one GRU Network, two fully connected linear layers and one output layer

The below diagram shows the LSTM Neural Network Architecture with two fully connected network. The Neural network model was developed in such a way the above 4 architectures could be controlled by using global parameters.
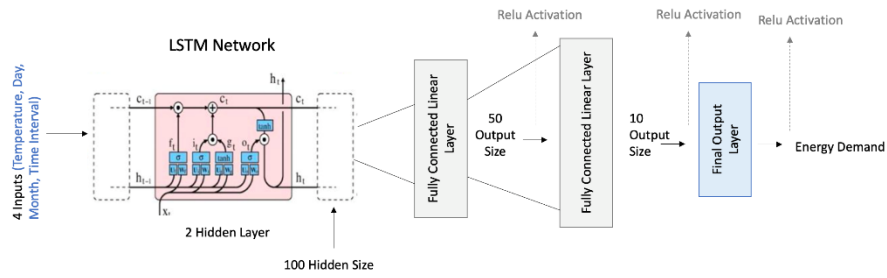
Figure 9: LSTM architecture

The team found that the model with one GRU network, two fully connected linear layers and one output layer could predict better than the models with other architectures.
The below Diagram shows the additional hyperparameters used in the GRU Neural Network model.
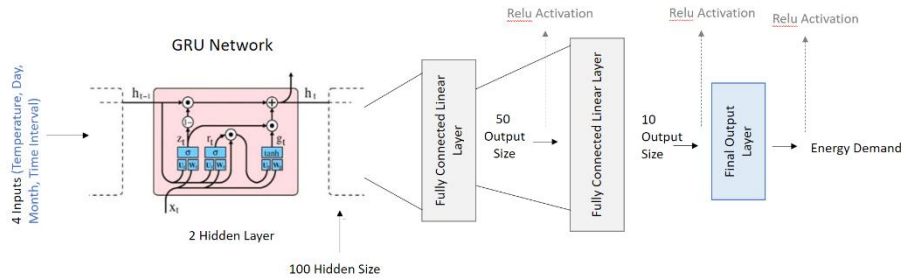

Figure 10: GRU architecture with additional hyperparameters

18

Figure 11 below compares the gradual decrease in loss of two neural networks with different architectures. This graph shows how a Neural Network Model that uses **a Gated recurrent** unit with two FC layers and one output layer learns better than a regular **Recurrent Neural Network** with one output layer. The loss during training clearly drops much lower and much quicker in the more complex GRU NN than the simple RNN.
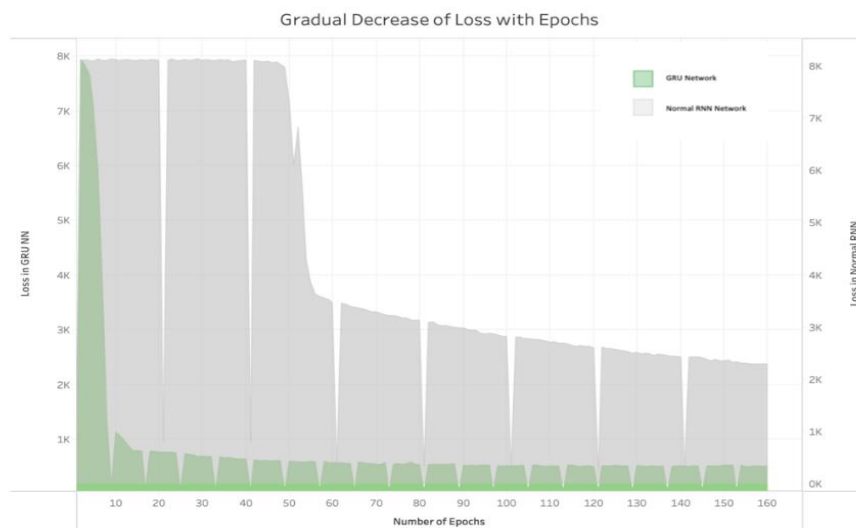

Figure 11: Loss decrease in GRU vs. RNN

In the end, this proved to be the most accurate approach, with a validated MAE loss as low as 510, compared to the RNN which achieved a MAE loss of over 2000.

Below graph shows the target vs prediction comparison for the GRU neural network developed.
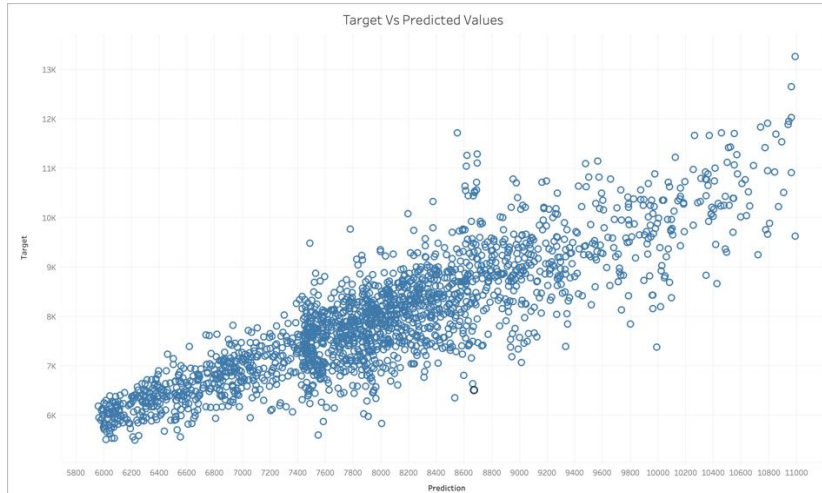
Figure 12: GRU Neural Network target vs prediction

Our reflections as a team on this model were that it was successful in predicting demand reasonably accurately. It would be viable as a commercial product for generators or other market participants seeking short- or medium-term forecasts. However, because of time limitations, feature selection was not explored very deeply with these NN models. The models could have been improved further with additional feature selection. The most notable feature we would have added next would have been previous demand data (e.g., demand at T minus 30 minutes), which proved helpful in the other models we built (see below).

5.2   Machine Learning

5.2.1   Random Forest

After creating a baseline, untuned random forest model taken from the scikit-learn package, we performed hyperparameter tuning to determine whether performance could be improved further. We developed a randomised search grid using cross validation and the following parameters were chosen for our final model (achieving the lowest MAE):

- Number of estimators – 80
- Maximum features – 'auto'
- Maximum tree depth – 15
- Minimum sample split – 4
- Min samples per leaf – 1
- Bootstrap – 'True'

We were also able to train our random forest to achieve a 90% confidence interval around predictions which achieves much better practical applications then a single point forecast.

5.2.2 Support Vector Machine

Adopting the same approach as for a random forest, we tested an untuned SVM model from the scikit-learn package, and then attempted to improve performance via exhaustive grid search of a range of values for the following hyperparameters:

- C – 1
- Epsilon – 0
- Kernel – Gaussian Radial Basis Function (RBF)

5.2.3 Machine Leaning Results

Of the two machine learning models the random forest was the best performer with a lower MAE by a wide margin – as low as 57 for the highest performing set of hyperparameters. In contrast, the best performing SVM (after hyperparameter tuning) could only achieve a MAE of 810. Table 2 shows the final performance of the machine learning models.

Model	MAE
Benchmark – 30-minute demand	62
Random Forest	57
SVM	810

Table 2: Machine Learning Results

It is important to note that both Machine Learning models were very computationally expensive, due mostly to the fact that we were working with such a large set of data. The random forest regressor was computationally much less expensive than the SVM and took much less time to train. Overall, the random forest approach is a superior solution to this forecasting challenge.

As the scatterplots on the following page show, the random forest model very credibly predicts demand to an accuracy of within 2-3%--when the model expects demand to be 8000 MW, the range of actual values is almost always between 7800 and 8200 MW. Conversely, the SVM model tends to underestimate demand in some scenarios when demand is between 8000 and 9000 MW. The benchmark has a much wider spread of estimates at all levels of demand, reinforcing that both models are an improvement over our benchmark.

The following series of charts show the visualized performance of each model, in terms of their predicted demands versus actual demand:
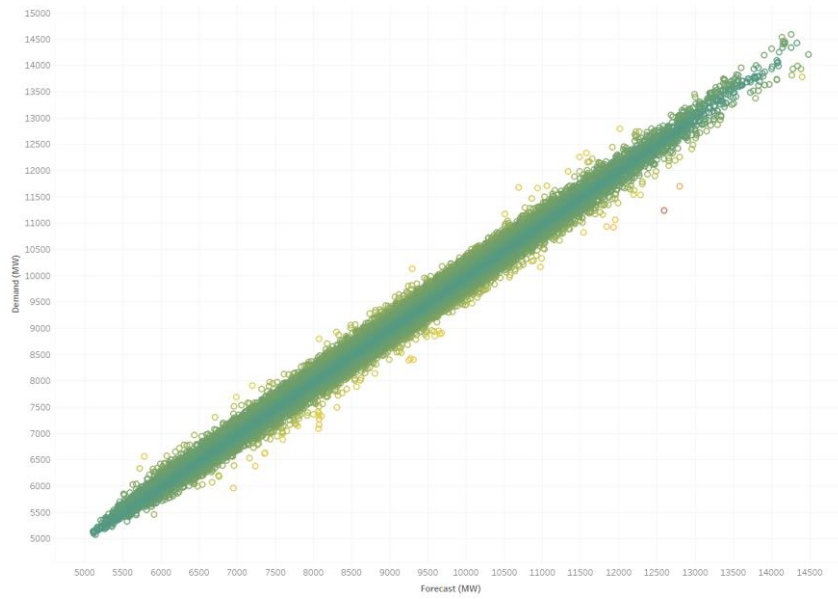

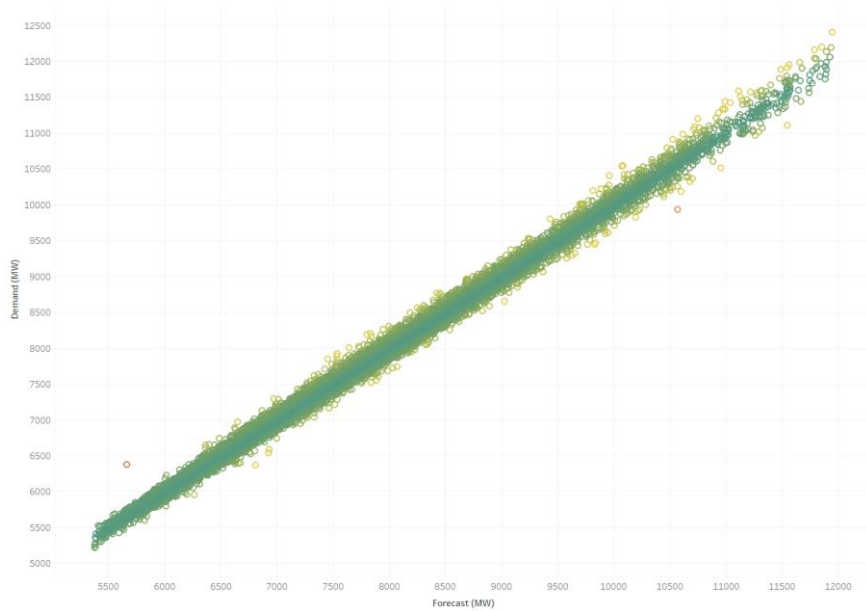Figure 13: Benchmark – 30-minute forecast vs demand scatterplot

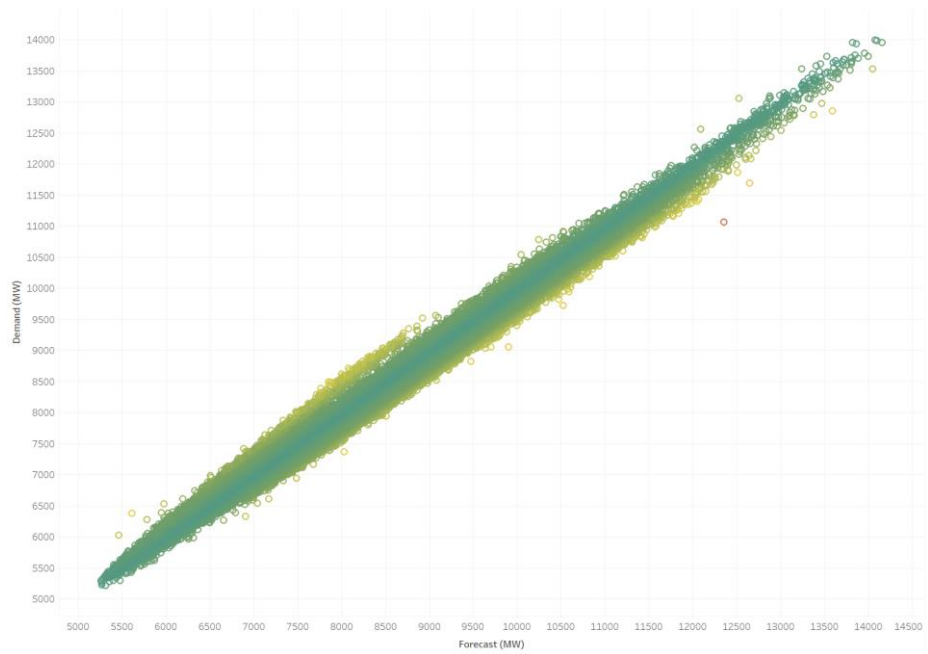Figure 14: Random Forest actual vs prediction scatterplot

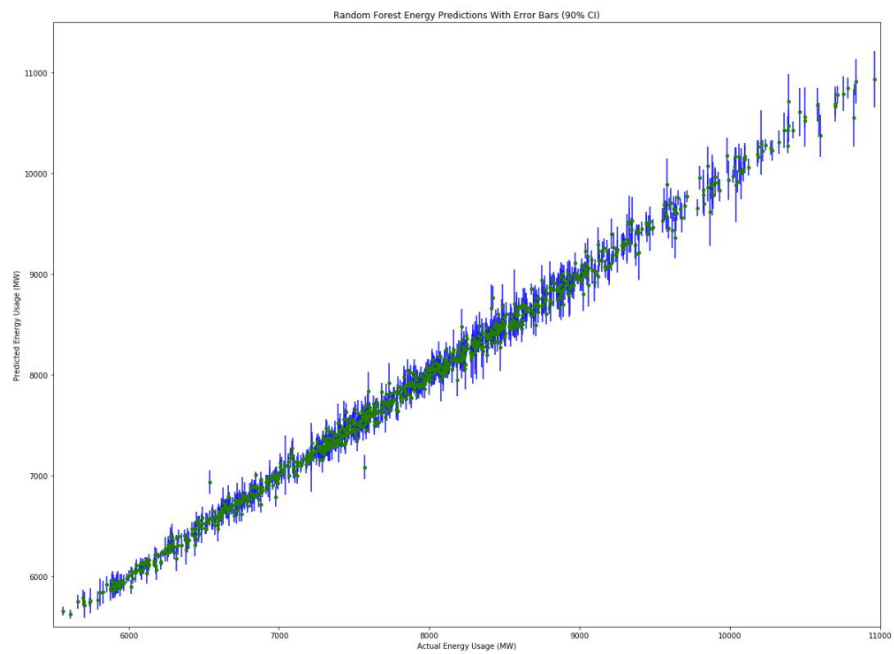Figure 15: SVM actual vs prediction scatterplot

Figure 16: Random Forest actual vs prediction scatterplot with confidence intervals

6   Discussion, Conclusion, and Further Issues

Based on our understanding of what is required by energy generators and other market participants, we believe that our random forest model is good enough for commercial release. It shows meaningful improvement from the benchmark data provided, in the form of the previous demand forecasts provided. The Neural Network model also showed promise if we had more time to explore alternative feature selection.

We then stepped back to consider the real-world implications of the model and how it could be used. Based on our research and our AEMO interview, we know that sometimes a point forecast is helpful. This is particularly true in long-term forecasting—if a generator is considering a major capital investment to increase capacity, or considering when to close a power station, then understanding the overall shape of demand is often enough. However, in short- and medium-term forecasting, such as our models were built to deliver, usually a range is more appropriate, and the "worst-case" scenario (usually meaning the highest possible realistic demand is often more important than the point estimate. The reason for this is that demand significantly more than supply can lead to price spikes and significant instability, sometimes even leading to blackouts and/or equipment damage. Therefore, in the scenario above, where the model forecasts 8000 mWh demand, the AEMO will be more interested in the demand two or three standard deviations above the mean forecast. That is, they will want to understand the 8200 mWh number and have contingency plans around it, directing generators to have capacity on standby (for example, starting the burn in coal generators but not turning on the dynamos yet) if needed (AEMO 2020).

The right way to think about the appropriate level of uncertainty to accept (e.g., whether to plan around a two- or a three- or a four-standard-deviation event) is a mathematical concept called the "newsvendor model." This approach looks at the cost of making a forecast that is too high (so that money is wasted on spare capacity that isn't needed) or too low (leading to price spikes and stability issues) (Cigdem 2017). By looking at the statistical distribution of outcomes, the AEMO or other market participants can minimize the overall expected cost to the system across both kinds of error. In practice, it seems to us that the AEMO looks at a two-standard-deviation event, although this is quite a complicated topic given the different response times of different sources of supply and demand.

Our random forest model runs in a similar manner, producing a 90% confidence interval to produce more utility in the situations where a single point forecast isn't sufficient. We were able to fit lines at 1.65 standard deviations above and below the predicted values and check how often the error exceeded those lines (validated at 10% of the time). We could then conclude that the model did not have fat tails, and that therefore it is robust for scenarios and for contingency-planning use as noted above, in addition to point-estimate use.

We believe that we have successfully built a model that could have real-world commercial applications, with market operators, generators, or traders as clients. In

practice, if we were to try to market this as a commercial product, we would invest more time in improving model performance, as noted below.

Our observations on the data and then the models are as follows:

- Electricity demand is highly variable, but it varies in relatively predictable ways, making it easier to forecast than many other economic variables
- The key to successful demand forecasting is to understand the complexity of the time input. More specifically, we found that time was extremely difficult to use as a single input, but much easier when we broke it down into its component parts: time of day, date in the year, year, and weekend/weekday. Modelling with these inputs separately was much simpler and more effective.
- Ultimately, if the client is most interested in a worst-case scenario, then the value of marginal improvements in model accuracy may not be very high. The reason for this is that the biggest decision (by far) the client has to make is what level of uncertainty they are willing to tolerate (e.g. two standard deviations vs. three), and small improvements in the point forecast will not be very consequential in comparison.

We have a few reflections on the models themselves, and how additional work could be conducted in this area.

First, there is more that could be done to improve the accuracy of these models if more time were available. The first step would have been to include more variables from the data we were given. This could have been done in two ways: first, breaking down time into multiple variables, as discussed above (we did this to a certain extent but not fully across all modelling efforts). The second would have been to create more variables based on previous demand and previous temperature at different points in time—30 minutes ago, one hour ago, 4 hours ago, and so on.

The second step we could have taken to improve accuracy, had we had more time, would have been to integrate more data from external sources into the model. For example, in our AEMO interview, we learned that in the summer, humidity is quite a significant drive, due to its impact on air-conditioner use (Wonhas 2022).

The third step we could have taken would have been to spend additional time on tuning the models, although as we did a fair bit of this, we think the incremental improvements would have been modest.

The final step to take in this project is to remember that modelling is an iterative process. We can see from the exploratory data analysis that trends in temperature and demand change over time and a model that we create now is not necessarily a model that will perform well in x amount of years. This should be kept in mind and the models should be revisited after 2 to 3 years, or once the difference between predictions and measured energy demand starts to show trends towards inaccuracy.

We have one additional reflection on the overall modelling effort. As discussed throughout this report, we have worked hard to keep a practical and commercial lens on our work. As noted above, some clients would need a point forecast, while others would need a worst-case scenario. It is easy to see that different models would be appropriate for these two different clients—for the first client, a model that minimizes MAE would be appropriate, while for the second client, a model that balances MAE with the fatness of the tails would be more helpful. Our random forest model achieved both deliverables and we believe is ready to market as a commercial product.

7   Final Thoughts

Finally, while we understand that this report is meant to be written as a consultant would write a report for a client, we wanted to step outside that framework briefly to discuss what we have learned.

The process of researching the electricity market in Australia, and in New South Wales in particular, as well as the process of building these models and analysing their results, has helped understand the practical applications of data science in significantly more depth than before. It has taught the team a great deal about the work required to ensure that data science is useful in the real world, and not just a technical exercise in model-building.

As a team, we also have learned a great deal about project management and working collaboratively. We have all learned a few lessons, ranging from the importance of having the right tools and software, to the importance of proper team roles and responsibilities, and finally the importance of a strong operating rhythm, supported by appropriate agendas and minutes. These lessons will stand us in good stead in future data-science projects, and indeed more broadly in our careers and lives.

The entire team—Shuba, Abdul, Chris, David, Sonal, and Jamie—would all like to thank our instructors for the opportunity this course provided, and indeed for the opportunity provided to us by the entire master's program.

References

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

Australian Energy Market Operator, Electricity Demand Forecasting Methodology Information Paper, August 2020. https://www.aemo.com.au/-/media/files/electricity/nem/planning_and_forecasting/inputs-assumptions-methodologies/2020/2020-electricity-demand-forecasting-methodology-information-paper.pdf

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

Bennett, C., Stewart, R.A. and Lu, J., 2014. Autoregressive with exogenous variables and neural network short-term load forecast models for residential low voltage distribution networks. *Energies*, *7*(5), pp.2938-2960.

BBC, "Can you have a big 'switch off'?". BBC News. 6 September 2007. http://news.bbc.co.uk/2/hi/uk_news/magazine/6981356.stm

Cigdem, E, "A newsvendor approach to energy imbalance mechanisms in a day-ahead electricity market," Bilkent University, 2017. http://repository.bilkent.edu.tr/bitstream/handle/11693/33380/10152663.pdf?sequence=1&isAllowed=y

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.

Ghalehkhondabil, I, Ardjmand, E, Weckman, G, and Young, W, An overview of energy demand forecasting methods published in 2005-2015. Department of Industrial and Systems Engineering, Russ College of Engineering and Technology, Ohio University, Athens, OH. https://www.researchgate.net/profile/Iman-Ghalehkhondabi/publication/301665131_An_overview_of_energy_demand_forecasting_methods_published_in_2005-2015/links/5728e47a08aef7c7e2c0ca8e/An-overview-of-energy-demand-forecasting-methods-published-in-2005-2015.pdf

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.

Knez, S., Šimić, G., Milovanović, A. *et al.* Prices of conventional and renewable energy as determinants of sustainable and secure energy development: regression

model analysis. *Energ Sustain Soc* **12,** 6 (2022). https://doi.org/10.1186/s13705-022-00333-9

Mahdavi, N., 2020, November. Probabilistic forecasting of operational demand in Australia. In *2020 International Conference on Smart Grids and Energy Systems (SGES)* (pp. 139-144). IEEE.

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.

Saddler, Hugh. Power Down Why is electricity consumption decreasing? The Australia Institute Paper No. 14 December 2013 ISSN 1836-8948. https://australiain-stitute.org.au/wp-content/uploads/2020/12/IP-14-Power-down_0.pdf

Sims, R.E.H.. Renewable energy: a response to climate change. Solar Energy, Volume 76, Issues 1–3, 2004, Pages 9-17, ISSN 0038-092X. https://doi.org/10.1016/S0038-092X(03)00101-4.

Wonhas, A, former Deputy CEO of the Australian Energy Market Operator, interview conducted 25 March 2022.

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.