**ANL488 Project Proposal**

**Forecast Model on Electricity Price and Demand in Singapore from 2023 to 2050**

A picture containing icon

Description automatically generated

**Submitted By:**

**Yim Wai Ling**

**H1882319**

**SCHOOL OF BUSINESS**

**Singapore University**

**of Social Sciences**

**Presented to Singapore University of Social**

**Sciences in partial fulfilment of the**

**requirements for the**

**Degree of Bachelor of Science**

**In Business Analytics**

**2022**

**TABLE OF CONTENTS**

Chapter One Introduction 3

Purpose

Background

Chapter Two Literature Review 5

Literature 1

Literature 2

Literature 3

Comparison and Conclusion

Chapter Three Data Understanding and Preparation 11

Data Understanding

Data Preparation

Chapter Four Proposed Modelling and Evaluation 14

Proposed Model 1

Proposed Model 2

Performance Evaluation Metrics

Chapter Five Proposed Schedule 16

Project Objective

Project Schedule and Methodology

Project Considerations

References 18

**1. Introduction**

**1.1 Purpose**

The purpose of this proposal is to research and evaluate the feasibility of forecasting Singapore’s electricity prices and demand from 2023 to 2050. It also investigates into the correlation between renewables growth and electricity prices in Singapore. The proposal also includes a literature review of three research papers on the topic of forecasting electricity prices using various models and datasets. Followed by data description and preparation of the proposed dataset sourced from the Energy Market Authority (EMA). Lastly, this proposal contains proposed models for developing a time series forecasting model for the electricity price and demand in Singapore, as well as the project’s future plans.

**1.2 Background**

Forecasting electricity prices has grown increasingly crucial in recent years, especially for competitive electrical market players. As such, time series forecasting model for electricity prices has become an important tool for driving strategic decisions and optimizing the scheduling of its energy resources.

Time series forecasting can be defined as “the process of analysing time series data using statistics and modelling to make predictions and inform strategic decision-making” (Tableau, 2022). The results of the forecast can vary greatly especially when dealing with the commonly fluctuating variables and other inevitable factors in the time series data. Forecasting has a wide range of other practical applications in various industries, including weather forecasting, finance forecasting, environmental studies forecasting, and many more, as long as consistent historical data is available. Typically, the accuracy of the forecast is determined by how comprehensive the given data is.

Accurate electricity price forecasting is vital because it enables better optimization of electricity supply and demand, promoting market stability and efficiency. With the current global COVID-19 pandemic, the price of raw materials (e.g. coal) for electricity production has skyrocketed, exacerbating price fluctuations and increasing the need for electricity price forecasting. In addition, this also accelerate the transition toward clean energy, particularly in Asia. In 2020, Japan and South Korea announced to achieve net zero by 2050 while Singapore has pledged to heaving peak emissions by 2050 (Andrews-Speed, 2021).

While renewables have grown in popularity over the years, it has been reported that global electricity demand is growing faster than renewables, resulting in a significant increase in electricity generation from fossil fuel and electricity prices (iea, 2021). Despite that, renewables still remain as one of the major variables influencing electricity pricing.

In Singapore, the main drivers of electricity demand are growing population, existing and new industries, and possibly the transition from combustion engines to electric automobiles in the transportation sector (Tan, 2019). Although Singapore has progressively increased the proportion of solar energy in its fuel mix in order to develop a future with a reliable, sustainable, and affordable energy supply, the country still faces challenges in transitioning to renewables due to space constraints and cloud cover (Andrews-Speed, 2021).

Singapore will need to analyse if its renewable growth can keep up with growing electricity demand while keeping electricity prices low for customers by 2050, and whether it can rely exclusively on solar energy to fulfil future demands. The electricity price forecast model helps Singapore not just in predicting prices based on historical pricing data, but also analysing how much renewables impact electricity prices.

**2. Literature Review**

The guiding principle for the literature review in this proposal is to deepen the understanding and knowledge of the impact of renewable energy and other market variables on electricity pricing, as well as to examine the methodologies used to forecast electricity prices and power demand. With this in mind, three research papers were chosen and reviewed to help build this knowledge.

**2.1 Literature 1**

This purpose of this paper is to present a comparative model for forecasting day-ahead electricity prices using different ANN models and to select the most suitable model for the prediction. It mainly looks into three ANN models, namely Multi-Layer Neural Network (MLNN) with Levenberg-Marquardt (LM) algorithm, Generalized Regression Neural Network (GRNN) and Cascade-Forward Neural Network (CFNN). The dataset used consists of historical Uniform Singapore Energy Price (USEP) and system demand, both of which are presented monthly.

The authors deemed that price prediction using soft-computing models, such as ANN models, are more suitable due to its robustness and its ability to learn non-linear patterns. While hard-computing models, such as Autoregressive Integrated Moving Average (ARIMA) and generalized auto-regressive conditional heteroskedastic (GARCH) can generate excellent prediction accuracy, they require a massive amount of information and have high computational cost. Given that electricity prices generally have a non-linear pattern, the authors decided to solely investigate the aforementioned soft-computing models and recommend the best model to the end user based on the simulation results.

The ANN models are trained to produce a specific target output depending on a given input. The training and testing process used in this study consists of four steps: (1) assembling the training data, (2) creating the network, (3) training the network and (4) computing the network response to new inputs (Anbazhagan & Kumarappan, 2011).

The first ANN model for electricity price forecasting is the MLNN model. This MLNN structure have one input layer composed of 33 neutrons, one hidden layer composed of 23 hidden layer neurons and one output layer composed of one neuron. Back-propagation (BP) algorithm training is commonly used in the training of MLNNs because it provides high degrees of robustness and generalization (Anbazhagan & Kumarappan, 2011). However, it has a slow convergence rate and may produce suboptimal solutions as it relies on the steepest descent method to update the weights. As such, LM algorithm is selected to train MLNN as it is one of the most effective algorithms for tackling this problem.

The second ANN model for electricity price forecasting is the GRNN model. The GRNN structure has a multilayer structure comprising an input layer, a hidden layer, an output layer, and a regression layer that has one unit more than the output layer and only contains linear units. In this system, the feature's vector is a real-valued input vector and has only one output to forecast electricity prices (Anbazhagan & Kumarappan, 2011).

The third ANN model for electricity price forecasting is the CFNN model. The CFNN structure have one input layer with 33 neutrons, one hidden layer with 23 neurons and one output layer with one neuron for forecasting electricity prices. Similar to MLNNs, CFNN employ the BP algorithm to update weights, however each neuron in the network is related to all neurons in the previous layers. Nonetheless, it has been observed that CFNN with BP training are extremely compatible, since they provide the best performance in terms of convergence time, optimum network structure and recognition performance (Anbazhagan & Kumarappan, 2011).

In this study, it is concluded that the CFNN model is the recommended model for price prediction. The result from the comparisons shows that the CFNN model possesses better forecasting abilities and requires lesser computation time as compared to the other ANN models. Furthermore, the volatility had the least impact on its performance.

Overall, the study is useful for this study since it helps to seek a comprehensive understanding of soft-computing models on forecasting electricity price in Singapore, their characteristics, and how well each of the three ANN models fits the data set. It is also worth noting that before making a final recommendation, an ensemble of different models is thoroughly reviewed. This knowledge can be crucial in developing the model for this proposal.

**2.2 Literature 2**

The purpose of this paper is to evaluate the effectiveness of Long-Short Term Memory (LSTM) model in forecasting electricity price for the next hour in both Australia and Singapore market. This result from this model is then compared against the other four popular forecasting models: conventional BP based multilayer feedforward network (BP-ANN), ANN with the wavelet transformation (WTANN), seasonal ARIMA (SARIMA) and adaptive neuro fuzzy inference system optimized with particle swarm optimization algorithm (PSO-ANFIS) (Jiang & Hu, 2018). In addition to the historical electricity pricing, exogenous variables such as day of the week, holidays, weather conditions, oil prices, hour of the day and historical price and demand are used as inputs to the model.

The author stated that time series-based models, such as ARIMA and GARCH, have been widely employed to forecast electricity prices. However, they can be problematic when there are frequent changes in prices and rapid variations (Jiang & Hu, 2018). As such, robust non-linear modelling techniques such as ANN and AWNN should be included for comparison.

At the pre-processing stage, it is observed that there is a need to refine the prices due to the infrequency of price spikes. Negative and extreme prices will need to be refined into specific values in order to minimize the effect of abnormal events on the performance of the predictive model (Jiang & Hu, 2018).

The gradient descent optimization algorithm in RNN often uses BP to modify weights between network layers during training (Jiang & Hu, 2018). However, such approach may prevent the neural network from further training (Jiang & Hu, 2018). As a result, LSTM is presented as a solution to this problem. The training algorithm for LSTM is backpropagation through time (BPTT), which is an extension of the classic BP for RNN. After the LSTM network has been trained, the model's weights and biases can be acquired, allowing the model to be used to forecast the electricity price.

It is observed that the proposed multilayer LSTM based model is the optimal model for forecasting the day-ahead electricity prices due to the aforementioned advantages and its ability to bridge long time lags of inputs as well as remembering the historical trend information in time series. The results show that for both Australia and Singapore markets, the LSTM model outperforms other compared methods.

Overall, the study provides a better understanding of the LSTM model, its characteristics, merits, and approaches to the given dataset. However, more research and analysis of the other four forecasting models is required to evaluate whether they should be included in the project, as they were not extensively emphasized and investigated in this research paper.

**2.3 Literature 3**

This research paper focuses on employing machine learning algorithms to forecast electricity spot market prices in Germany. The forecasts make use of spot market order book data from EPEX as well as other exogenous variables such as renewable infeed and expected total demand. Following on from existing literature, appropriate feature extraction for the dataset is built. Random forests and neural networks are fitted to the data using cross-validation to optimize hyperparameters. The sample performance of these models is then compared to statistical reference models.

The paper discussed that renewable electricity generation is highly volatile and has a substantial impact on the day-ahead electricity price. Using multivariate regression approaches, it is inferred from a review of existing literatures that increased renewable infeed generally leads to lower market pricing. As such, expected solar and wind infeed are included as features for price forecasting in this research paper.

At the data preparation stage, feature extraction is being used on the dataset to reduce dimensionality while preserving valuable information for forecasting. It emphasises the importance of practicing feature scaling before training a neural network as large-scale features might dominate on small-scale features and relevant information might be discarded by the net (Schnürch & Wagner, 2020).

Random forest network is a tree-based technique of Machine Learning (ML). Given that decision trees can perform poorly due to their reliance on the training data, random forests are introduced to overcome this limitation. It is capable of averaging the predictions of numerous decision trees that have been trained in a randomized way using the same data (Schnürch & Wagner, 2020). They can also provide a ranking of the features based on their relevance for predicting electricity prices, and this ranking can also be used for feature selection to train the following neural network approach (Schnürch & Wagner, 2020).

Feedforward neural network (FFNN) is a type of non-linear neural network that consists of several layers and uses a combination of non-linear activation functions and weighted sums to generate output. It has been proven that increasing the number of layers improves performance for many applications, notably deep learning applications. Aside from the number of hidden layers and neurons per layer, other so-called hyperparameters can also have a significant impact on forecasting performance (Schnürch & Wagner, 2020).

Finally, Hyperparameter Optimization via Cross-validation is used for the neural networks and random forests, with the training dataset divided into five equal-sized parts. The final results show that the FFNN model with only ten features chosen by random forest performs the best, but only slightly better than simpler models such as the ordinary linear regression. It is also concluded that while ML produces competitive results, it cannot significantly reduce the work effort required in model setup (Schnürch & Wagner, 2020).

Overall, this study emphasizes the importance of data preparation and feature extraction prior to neural network training, as well as whether ML techniques can be used to improve neural network predictive performance. Furthermore, it is unique as it considers the impact of renewable energy generation on electricity prices and provides an overview of the economics of the electricity markets. This is significant because it is something I can dive into for my own project.

**2.4 Comparison and Conclusion**

Overall, the three research papers chosen for this literature review were vital in laying the groundwork for this project on forecasting electricity price and demand in Singapore.

A unifying factor among the three research papers is that soft-computing models are better suited to forecasting electricity prices than hard-computing models, such as ARIMA and GARCH. (Anbazhagan & Kumarappan, 2011) stated that hard-computing models “requires a lot of information, and the computational cost is very high”, while (Jiang & Hu, 2018) pointed that they can also be “problematic when there are rapid variations and high frequency changes in the price”. On the contrary, soft-computing models offer greater flexibility and “better solution to model complex non-linear relationships” (Anbazhagan & Kumarappan, 2011).

One of the key differences among the three research papers is the types of soft-computing models (deep learning) used. (Anbazhagan & Kumarappan, 2011) solely focuses on providing a better understanding of ANN models as well as a thorough analysis of the characteristics and performance of all three suggested models (MLNN with LM algorithm, GRNN and CFNN). Ultimately, the paper provided a better overview of the different ANN approaches available for electricity price forecasting.

On the other hand, (Jiang & Hu, 2018) focuses on LSTM modelling and compares its results against the other four commonly used forecasting models: BP-ANN, WTANN, SARIMA and PSO-ANFIS. In comparison to (Anbazhagan & Kumarappan, 2011), (Jiang & Hu, 2018) lacks depth in evaluating the other four models and does not explain their methodology as thoroughly as the suggested LSTM model. However, (Jiang & Hu, 2018) used exogenous variables, such as day of the week, holidays, and weather conditions as additional inputs to improve the effectiveness of the predictive model. These additional inputs and data can also be considered for this project.

Similarly, (Schnürch & Wagner, 2020) also included other market variables, such as the total expected electricity demand and renewable infeed, as inputs to its FFNN model. Coupled with the support from other relevant studies, (Schnürch & Wagner, 2020) emphasis that it is important to take “expected solar and wind infeed as features for the price forecasts” as renewable electricity generation “has a substantial impact on the day-ahead electricity price”. (Schnürch & Wagner, 2020) also focuses whether the prediction of electricity prices can be further improved by employing an additional ML algorithm on top of the proposed neural network. In comparison to the other two papers, (Schnürch & Wagner, 2020) places the most emphasis on feature extraction prior to neural network training, which contributes to a better understanding of its significance. As a result, this project can also investigate feature extraction methodologies in order to improve forecasting results.

**3. Data Understanding and Preparation**

**3.1 Data Understanding**

There are three datasets proposed for use in this electricity price and demand forecasting project, all of which are obtained from EMA. EMA is a statutory board under the Ministry of Trade and Industry whose mission is to ensure a reliable and secure energy supply, promote effective competition in the energy market and develop a dynamic energy sector in Singapore (Government of Singapore, 2022). As such, the source of the datasets is highly credible, and EMA has made these datasets publicly available for educational and research purposes.

The first dataset is the average monthly Uniform Singapore Energy Price (USEP) in $/MWh (Government of Singapore, 2022).

Graphical user interface

Description automatically generated with medium confidence

The half-hourly energy price in the Singapore Wholesale Electricity Market is known as the Uniform Singapore Energy Price (USEP). However, the USEP is presented monthly from 2005 to 2021 in this dataset. The monthly data is preferable over the daily data as it can be used more effectively with the other datasets.

It is also observed that the data types are not consistent. Hence, there is a need to align the data types in this dataset.

Table

Description automatically generated

The second dataset is the annual installed capacity of residential and non-residential grid-connected solar PV systems from 2009 to 2021 by the different user types in MWp and MWac (Government of Singapore, 2022).

Table

Description automatically generated

Based on the above observation, it is notable that there is a lot of empty cells or missing values in the dataset as well as inconsistency in the data types. Hence, there is a need to conduct data preparation for this dataset.

The third dataset is the total final annual energy consumption by sector and energy product in ktoe from 2009 to 2019 (Government of Singapore, 2022).

Table

Description automatically generated

Similar to the second dataset, the dataset contains missing values and there is a need to replace or remove them.

Although the data type in this dataset is consistent, it may be necessary to change it in order for it to better suit the model.

Table

Description automatically generated

**3.1 Data Preparation**

For the first and third dataset, the data types have been changed to ensure consistency and to minimise issues when formulating the models at the later stage.

Graphical user interface, text, application

Description automatically generated Table

Description automatically generated

The rows that contain special characters and blank values are also removed for the third dataset.

Table

Description automatically generated

**4. Proposed Modelling and Evaluation**

The three literature reviews have assisted in the understanding of the various types of modelling that can implemented using a similar proposed dataset for electricity price forecasting. Since the reviews revealed that deep neural network approaches outperform conventional approaches, only deep neural network approaches will be included in this project and their performances will be compared. The two approaches that are selected for this project are LSTM and FFNN.

**4.1 Proposed Model 1**

The first proposed model is LSTM. When it comes to time series modelling, RNN models (i.e. LSTM) have been proven in (Jiang & Hu, 2018) to outperform ANN models. To generate an output, ANN usually assigns a weight matrix to the current input without considering past information as information only flows through once. As a result, ANN may not produce promising results when time context is involved. However, LSTM is able to resolve this issue as it has the ability to bridge long time lags of inputs and store past information using a memory cell. Therefore, LSTM is chosen to be analysed with the proposed dataset in this project.

**4.2 Proposed Model 2**

The second proposed model is FFNN. FFNN and LSTM are mathematically quite similar, except that LSTM incorporates a loop. (Schnürch & Wagner, 2020) also stated that FFNNs outperformed RNNs in time series applications and this may be “due to the high dimensionality of the multivariate time series under consideration” (Schnürch & Wagner, 2020). However, the paper lacks detailed discussion on this subject. Hence, it will be interesting to compare and the analyse the performance of both models in this project due to their similarities and strengths in time series forecasting.

**4.3 Proposed Performance Metrics**

We can see from Chapter 3 that the proposed datasets are not very large, therefore it is not necessary to reduce the dimensionality and the complete dataset can be used for modelling. The proposed performance metrics for evaluating the models for this project are Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE). These are the most common Key Performance Indicators (KPIs) used to measure the accuracy of forecasting models.

**5. Proposed Schedule**

**5.1 Project Objective**

The business problem is that Singapore will need a forecasting model to optimize the allocation of its energy resources and balance electricity supply and demand. As such, the business objective is to improve the stability of the electricity market while keeping electricity prices low for consumers. The data mining objective is to develop an electricity price forecasting model that can predict accurate electricity prices from 2023 to 2050.

**5.2 Project Schedule & Methodology**

This project will employ the Cross Industry Standard Process for Data Mining approach (CRISP-DM). This methodology will also be used to document the project.

Below is the proposed project schedule moving forward:

|  |  |
| --- | --- |
| **Week of** | **Proposed Deadlines** |
| 28 Feb | Data Prep and Modelling 1 |
| 7 Mar | Data Prep and Modelling 2 |
| 14 Mar | Data Prep and Modelling 3 |
| 21 Mar | Presentation Slides, Preparation and Presentation |
| 28 Mar | Review data prep and modelling with Mentor |
| 4 Apr | Review data prep and modelling with Mentor |
| 11 Apr | Finalize models |
| 18 Apr | Prepare final report |
| 25 Apr | Prepare final report |
| 2 May | Prepare final report |
| 9 May | Submission Deadline |

**5.3 Project Considerations**

The project can consider and examine other models than those proposed in Chapter 4 or even incorporate and test different ML training algorithms with the proposed models. If there is additional time, it is worth considering including hard-computing models to evaluate whether the results differ significantly from soft-computing models. Furthermore, time and computational costs must also be considered when experimenting and performing additional data preparation and modelling.

**References**

Anbazhagan, S., Kumarappan, N. (2011). DAY-AHEAD PRICE FORECASTING IN ASIA'S FIRST LIBERALIZED ELECTRICITY MARKET USING ARTIFICIAL NEURAL NETWORKS*.* *Atlantis Press.* Retrieved from <https://www.researchgate.net/profile/Anbazhagan-S/publication/254222210_Day-Ahead_Price_Forecasting_in_Asia%27s_First_Liberalised_Electricity_Market_Using_Artificial_Neural_Networks/links/578e89d708ae35e97c3f736e/Day-Ahead-Price-Forecasting-in-Asias-First-Liberalised-Electricity-Market-Using-Artificial-Neural-Networks.pdf>

Andrews-Speed (2021). *Commentary: Why hasn’t solar energy in Singapore taken off in a big way after so long?* Channel News Asia.Retrieved from <https://www.channelnewsasia.com/commentary/solar-energy-singapore-panels-cloudy-unpredictable-electricity-1882996>

Government of Singapore (2022). *Average Monthly Uniform Singapore Energy Price.* Energy Market Authority. Retrieved from <https://www.ema.gov.sg/statistic.aspx?sta_sid=201707106swQNahHjHnP>

Government of Singapore (2022). *Installed Capacity of Grid-Connected Solar Photovoltaic (PV) systems by User Type*. Energy Market Authority. Retrieved from <https://www.ema.gov.sg/statistic.aspx?sta_sid=20170711hc85chOLVvWp>

Government of Singapore (2022). *Our Roles.* Energy Market Authority. Retrieved from <https://www.ema.gov.sg/Our_Roles.aspx>

Government of Singapore (2022). *Total Final Energy Consumption.* Energy Market Authority. Retrieved from <https://www.ema.gov.sg/statistic.aspx?sta_sid=20141211NBOOXgSCZ0no>

iea (2021). *Global electricity demand is growing faster than renewables, driving strong increase in generation from fossil fuels.* iea. Retrieved from <https://www.iea.org/news/global-electricity-demand-is-growing-faster-than-renewables-driving-strong-increase-in-generation-from-fossil-fuels>

Jiang, L., Hu, G. (2018). Day-Ahead Price Forecasting for Electricity Market using Long-Short Term Memory Recurrent Neural Network. *A\*STAR & Nanyang Technological University.* Retrieved from <https://www.researchgate.net/profile/Lian-Jiang-6/publication/329878004_Day-Ahead_Price_Forecasting_for_Electricity_Market_using_Long-Short_Term_Memory_Recurrent_Neural_Network/links/5c1f6d6e299bf12be39435fc/Day-Ahead-Price-Forecasting-for-Electricity-Market-using-Long-Short-Term-Memory-Recurrent-Neural-Network.pdf>

Schnürch, S., Wagner, A. (2020). Electricity Price Forecasting with Neural Networks on EPEX Order Books. Applied Mathematical Finance, 27:3, 189-206 (2020). Retrieved from <https://doi.org/10.1080/1350486X.2020.1805337>

Tableau (2022). *Time Series Forecasting: Definition, Applications, and Examples.* Tableau. Retrieved from <https://www.tableau.com/learn/articles/time-series-forecasting#:~:text=Time%20series%20forecasting%20is%20the,and%20inform%20strategic%20decision%2Dmaking.&text=Forecasting%20then%20takes%20the%20next,might%20happen%20in%20the%20future>

Tan (2019). *Singapore to ramp up solar energy production to power 350,000 homes by 2030.* Straits Times. Retrieved from <https://www.straitstimes.com/singapore/environment/solar-energy-to-meet-4-of-singapores-energy-demand-by-2030-up-from-less-than-1>