APPLICATION OF MACHINE LEARNING ALGORITHMS FOR ELECTRICITY PRICE PREDICTION

Project submitted to the

SRM University – AP, Andhra Pradesh

for the partial fulfillment of the requirements to award the degree of

Bachelor of Technology

In

Computer Science and Engineering

School of Engineering and Sciences

Submitted by

Pudi Arjun Sumanth | AP19110010251

Nunna Lakshmi Manasa | AP19110010382

Balla Jayanth Venkata Durga Bhuvaneswar | AP19110010394

Kreethi Kumari Mishra | AP19110010424

Under the Guidance of

Dr. Priyanka Singh

SRM University–AP

Neerukonda, Mangalagiri, Guntur

Andhra Pradesh – 522 240

May-2023

Certificate

Date: 25/05/2023

This is to certify that the work present in this Project entitled "**Application of Machine Learning Algorithms for electricity price prediction**" has been carried out by **Pudi Arjun Sumanth, Nunna Lakshmi Manasa, Balla Jayanth, Kreethi Kumari Mishra** under my supervision. The work is genuine, original, and suitable for submission to the SRM University – AP for the award of Bachelor of Technology in **School of Engineering and Sciences**.

Supervisor

Dr. Priyanka Singh,

Assistant Professor,

Department of Computer Science and Engineering,

SRM University – AP

Andhra Pradesh 522502, India

Acknowledgement

Without acknowledging those who made it possible for a task to be completed effectively, the happiness that results from doing so would be absent. Success is the result of tenacity and abstract hard work, but steadfast of all is good counsel. We thank everyone whose advice and support acted as a lighthouse and helped to make our efforts successful.

We want to sincerely thank our guide, **Dr. Priyanka Singh**, Assistant Professor, Department of Computer Science and Engineering, SRM University, Amaravati for her constant encouragement, guidance, and moral support throughout the project.

We want to extend our sincere gratitude and appreciation to each team member of this group for their unwavering support and dedication throughout the project. It is through our combined efforts and collaboration that we have achieved the successful completion of this project. We extend our sincere heartfelt thanks to one another for their contributions.

Table of Contents

Abstract

Electricity price forecasting plays a vital role in enabling market participants to make informed decisions and develop effective bidding strategies in the dynamic energy market. This research focuses on predicting electricity price fluctuations using machine learning techniques. The study employs various algorithms, including random forest, convolutional neural network, artificial neural network, gated to forecast future electricity prices.

The dataset used in this study is carefully organized and contains multiple columns capturing relevant factors such as settlement date, retail electricity prices (RRP), hour of the day, day of the week, 24-hour average load, 24-hour lagged load, previous 24-hour average load, previous week information, and weekday. These features provide valuable insights into temporal patterns, load fluctuations, and external influences affecting electricity prices.

Performance metrics, including mean square error, mean absolute error, mean absolute percentage error, R2 Score, and root mean square error are employed to evaluate the accuracy and effectiveness of the forecasting models. The experiments compare the performance of different algorithms, providing a comprehensive assessment of their predictive capabilities.

The research contributes to the development of effective forecasting models for the energy market, enabling stakeholders to make informed decisions, optimize revenue, and mitigate risks in a competitive electricity price market.

In conclusion, this study demonstrates the utility of machine learning techniques, specifically the Random Forest Regressor, which accurately predict electricity price fluctuations. The comprehensive dataset and rigorous evaluation of various performance metrics enhance the reliability and effectiveness of the forecasting models. The research findings provide valuable insights for market participants and contribute to the advancement of the energy industry by fostering competition, improving economic efficiency, and delivering better services.

Keywords: Random Forest Regression, Ridge Regression, Convolutional Neural Network, Long-Short Term Memory, Radial Basis Function, Artificial Neural Network, Gated Recurrent Unit

Abbreviations

RFR	Random Forest Regression
RR	Ridge Regression
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network
RBF	Radial Basis Function
ANN	Artificial Neural Network
GRU	Gated Recurrent Unit
SVR	Support Vector Regression
SVM	Support Vector Machine
OLS	Ordinary Least Squares
ARIMA	Autoregressive Integrated Moving Average
PPO	Proximal Policy Optimization
PSO	Particle Swarm Optimization
MSE	Mean Square Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Square Error
RRP	Recommended Retail Price

List of Tables

List of Figures

1. Introduction

A method for predicting the future based on current data outputs is called forecasting. In order to forecast future advancements, it is rigorously examined for both past and present tendencies. It employs statistical methods and equipment. It is also referred to as statistical analysis as a result. It is particularly useful as a planning tool to help organizations get ready for potential future unpredictability. Sharing management's expertise and experience is the first step in the process [1].

Electricity price forecasting is a requirement for all market participants in order to select bidding strategies and establish bilateral agreements that maximize revenues while minimizing risks. Electricity prices exhibit unique characteristics due to the physical nature of the commodity. Unlike other markets, electricity is non-storable, meaning that small changes in load or generation can result in significant price fluctuations within hours or even minutes. Energy price fluctuations, severe volatility, and seasonality are all features of the market. In addition, a number of factors, including the weather, fuel costs, and electricity usage, affect energy pricing. These characteristics make it difficult to predict energy price changes. The goal of these reforms is to promote competition, achieve economic efficiency, improve service quality, and lower consumer electricity prices. As a result, understanding and forecasting electricity prices have become crucial for market participants, enabling them to manage risks, enhance competitive strategies, and make informed decisions. Numerous forecasting methods have been documented in the literature recently. The three groups into which energy price prediction models are categorized are game theory models, fundamental models, and time series models [1].

Based on game theory, the first category investigates how players' tactical decisions impact the price of electricity using a key point equilibrium as a standard. The second category is based on theoretical or computer simulation models that precisely reproduce the power system's physical model. This method bases power market energy pricing on marginal generation costs while taking transmission congestion, losses, and other ancillary service requirements into account [2]. The planning of power generation and the growth of the electric power system are both aided by electric load forecasting. Accurate electric load forecasting is essential for the safe, secure, and reliable functioning of power systems. Power system planning requires precise projections of the electric load over the long and short periods. It addresses load forecasting with sufficient foresight to prepare for long-term and mid-term maintenance, planning the building of new generation facilities, buying generating units, and creating transmission and distribution systems.

Long-term load forecasts are established will have a big impact on future generating and distribution plans. Customers will come from underestimating the load demand, while overestimating it will necessitate sizable additional investments in the infrastructure that supports the production of power. The long-term and mid-term forecasting time frames from a few weeks to many years. Unfortunately, it could be challenging to anticipate load demand with sufficient accuracy over such a long planning period. This is due to the fact that forecasting is inherently unreliable. Numerous significant, many unknowable, and essential components make up the basic forecasting process.

Reforms in the energy industry have allowed for the transition from a vertically integrated monopoly market structure to a competitive wholesale retail system with marketplaces that resemble power exchanges. By creating a structured market in this way, Power Exchange is able to provide recently created European markets based on this concept with standardized goods. Nations with different power arrangements share comparable goals. They are all working to boost competition in the electrical sector in order to improve economic efficiency, supply better services, and lower customer electricity rates. Price analysis is now the first priority for all market participants. Power cost background knowledge is required for efficient risk management. For a rival in the market, it can be favourable. All players in the market, including those who are engaged in cash flow analysis, capital. Long-term energy price forecasting is advantageous to all sector participants, including those engaged in cash flow analysis, capital budgeting, financial procurement, regulatory rule-making, and integrated resource planning [2].

Periodic dynamic long-memory regression models with were introduced to capture the seasonality of electricity spot prices using sinusoids and weekday dummies. Other models have utilized the hyperbolic decay of autocorrelation or the sum of exponential functions to capture the long memory exhibited by electricity spot price time-series. electricity markets. To address these limitations, forecasting models that combine different techniques. The motivation behind models is to leverage the strengths of each individual model and improve forecasting accuracy. For instance, models like CNN and ANN models have been applied to various domains, including stock prices and electricity markets. These models have shown improved forecasting accuracy. The development of robust and accurate electricity price forecasting models is essential for market participants to make informed decisions, manage risks, and maximize profits in the competitive [3].

2. Related Works

In recent years, numerous studies have focused on electricity price forecasting using various methodologies and techniques. This section provides an overview of related works that have contributed to the field, considering the introduction, methodology, and dataset description. Forecasting electricity prices is crucial for market participants to optimize their bidding strategies and mitigate risks. The introduction highlights the challenges associated with energy price fluctuations, volatility, and the influence of factors such as weather, fuel costs, and electricity usage. To address these challenges, researchers have explored different forecasting methods and models.

One category of models discussed in the introduction is game theory models, which investigate how players' tactical decisions impact electricity prices. These models often use a key point equilibrium as a reference and examine the strategic interactions between market participants. They aim to capture the dynamics of electricity pricing based on players' behaviours and decision-making processes. Another category of models mentioned is fundamental models, which are based on theoretical or computer simulation models that reproduce the physical characteristics of the power system. These models consider factors such as supply-demand dynamics, transmission constraints, and market conditions to forecast electricity prices. They provide a more detailed understanding of the underlying mechanisms and physical constraints governing the electricity market. The third category of models discussed is time series models, which employ statistical and machine learning techniques to capture patterns and trends in historical electricity price data. These models leverage algorithms such as RFR, RR, CNN, LSTM, RNN, RBF, ANN, and GRU to predict future electricity prices.

For instance, the RFR is an ensemble learning algorithm that combines multiple decision trees to create a robust and accurate predictive model. It can handle a large number of input variables, capture complex nonlinear relationships, and provide insights into feature importance. Studies such as "An Online Electricity Market Price Forecasting Method Via Random Forest" [4] have utilized RFR to forecast the electricity market prices in online settings with real-time data availability. RR is another technique discussed in the methodology, which addresses multicollinearity and overfitting by introducing a penalty term to the ordinary least squares method. It preserves the interpretability of the model and is suitable for datasets containing correlated input features. Researchers have employed RR to improve the overall performance and interpretability of electricity price forecasting models.

Deep learning algorithms, such as CNN, LSTM, RNN, and GRU, have also been extensively explored for electricity price prediction. CNNs excel at capturing spatial patterns and relationships, making them suitable for analysing electricity price data. LSTM and GRU, as types of RNNs, have the ability to model temporal dependencies and capture long-term patterns. These algorithms have been used in research papers such as “An Electricity Price Forecasting Model by Hybrid Structured Deep Neural Networks” [8] and” Effective long short-term memory with differential evolution algorithm for electricity price prediction” [9] to predict electricity prices accurately.

Furthermore, ANN have been employed in electricity price forecasting due to their ability to capture complex relationships and dependencies in the data. ANNs can learn from historical data and make predictions based on patterns and relationships captured during the training phase. Studies such as” Electricity price forecasting using artificial neural networks” [11] have utilized ANN models to forecast electricity prices accurately.

The study titled “Electricity Price Forecasting Using Machine Learning Techniques: A Survey” provides a comprehensive overview of different machine learning approaches employed in electricity price forecasting. The authors review various algorithms such as SVR, ANNs, and Random Forests, among others. They highlight the strengths and limitations of each method and discuss their applicability in different forecasting scenarios. [4]

The paper titled” Short-Term Electricity Price Forecasting: A Review of State-of-the-Art Models” [4] focuses specifically on short-term electricity price forecasting. The authors review and compare different models, including ARIMA, ANN, and hybrid models that combine multiple techniques. They analyse the performance of these models in terms of accuracy and computational efficiency and provide insights into their suitability for short-term price forecasting. [5] The paper discusses electricity price forecasting “Gated recurrent unit network-based short-term photovoltaic forecasting” This paper focuses on short the term photovoltaic (PV) forecasting using a Gated Recurrent Unit (GRU) network. The objective is to improve the accuracy of predicting PV power generation within a short-term time horizon. The GRU network, a type of recurrent neural network, is utilized to model the sequential and temporal dependencies in the PV time series data. The network incorporates gating mechanisms that enable it to selectively capture relevant information and discard irrelevant 3 or redundant information during the forecasting process. By training the GRU network on historical PV generation data, it learns to capture the patterns and dynamics of the PV system, thus enhancing its predictive capability. The proposed approach is evaluated using real-world PV generation data, and its performance is compared with other forecasting methods. The results demonstrate that the GRU-based approach achieves improved accuracy in short-term PV forecasting, contributing to better predictions of electricity generation from photovoltaic sources [6].

In the paper” A review of deep learning methods applied on load forecasting” [6], the authors examine the use of machine learning algorithms for electricity load forecasting. They discuss the challenges associated with load forecasting, such as seasonality, weather patterns, and demand fluctuations, and review various models such as Decision Trees, SVMs, and Gradient Boosting. The study provides a comprehensive overview of the existing techniques and their performance in load forecasting. [7]

A study titled" Electricity Price Forecasting in Deregulated Energy Markets: A Review" [7] presents a comprehensive review of electricity price forecasting in deregulated energy markets. The authors discuss the characteristics of these markets, such as volatility and nonlinearity, and review different forecasting models, including statistical methods, econometric models, and machine learning techniques. They highlight the importance of accurate price forecasting for market participants and compare the performance of different models in terms of accuracy and computational efficiency. [8]

In addition to the above works, numerous other studies have explored the application of machine learning techniques for electricity price and load forecasting. These studies have investigated various algorithms, data pre-processing techniques, and feature selection methods to improve the accuracy and reliability of the forecasts. The research in this field continues to evolve, with an emphasis on developing hybrid models that combine different techniques to capture the complex dynamics of electricity markets.

Overall, the related works in electricity price and load forecasting highlight the significance of accurate predictions for market participants and the challenges posed by the volatility and complexity of energy markets. The application of machine learning algorithms has shown promising results in capturing nonlinear relationships, temporal dependencies, and external factors affecting electricity prices and load patterns. However, there is still ongoing research to further enhance the accuracy and robustness of these forecasting models.

3. Methodology

In this section, we present the methodology employed for predicting electricity price fluctuations using machine learning techniques. The following algorithms were utilized in our study:

3.1 Random Forest Regression

It is an ensemble learning algorithm widely utilized for regression tasks. By combining multiple decision trees, it creates a robust and accurate predictive model. Each decision tree in the forest is trained on a random subset of the training data and selects features randomly at each split. This process helps to reduce overfitting and enhance the model's generalization capability. The final prediction is obtained by averaging the predictions from individual trees. This algorithm can handle many input variables without requiring feature selection or dimensionality reduction techniques. Additionally, it can capture complex nonlinear relationships between the input features and the target variable, making it suitable for modeling electricity prices. Its ability to handle outliers, resistance to noise, and computation efficiency make it a powerful tool for accurate electricity price predictions. Moreover, the random forest regressor provides insights into feature importance, aiding in the interpretation of the model's results. The algorithm is implemented in popular machine learning libraries such as scikit-learn, allowing for easy integration and utilization within the project's framework.

A method for predicting electricity market prices using a machine learning algorithm called Random Forest. The authors focus on the forecasting of electricity prices in online markets, where prices can fluctuate based on various factors such as demand, supply, and market conditions. Accurate price forecasting is important for market participants to make informed decisions about buying and selling electricity. The RF algorithm is a type of machine learning model that uses a combination of decision trees to make predictions. In one paper, the authors propose using RF to forecast electricity market prices in an online setting, where real-time data is continuously available. The authors explain the process of training the Random Forest model using historical data, which includes information about past electricity prices and relevant market factors. They discuss how the model learns patterns and relationships in the data to make predictions about future prices. Additionally, the authors describe the online nature of the proposed method, which means that the model can adapt and update its predictions as new data becomes available. This allows for real-time forecasting, which is crucial in dynamic electricity markets.

3.2 Ridge Regression

This is a linear regression algorithm that adds a penalty term to the ordinary least square method. By doing so, it helps to reduce the impact of multicollinearity and prevent overfitting by shrinking the coefficients towards zero. The algorithm achieves this by introducing a regularization parameter, known as alpha, which controls the strength of the penalty term. RR is particularly useful when dealing with datasets containing correlated input features. It addresses the issue of multicollinearity by effectively reducing the influence of highly correlated predictors. Moreover, it preserves the interpretability of the model by shrinking less important features closer to zero. RR is computationally efficient and suitable for handling large datasets. Its simplicity and ease of implementation, especially in libraries such as scikit-learn, make it a popular choice for regression tasks. Within the context of our project, the RR algorithm played a crucial role in predicting electricity prices by handling multicollinearity and improving the model's overall performance and interpretability.

3.3 Convolutional Neural Network

These are deep learning algorithms commonly employed for image analysis and recognition tasks. They excel at capturing spatial patterns in data by applying convolutional layers, which convolve input data with learnable filters. This process allows the network to automatically extract relevant features from the input. CNNs also utilize pooling layers to down sample the feature maps, reducing computational complexity and increasing translation invariance. With their ability to learn hierarchical representations, CNNs can effectively capture complex relationships and patterns in electricity price data. By leveraging multiple convolutional and pooling layers, the algorithm can discern meaningful patterns at different scales. CNNs have demonstrated remarkable performance in various domains and are particularly useful for tasks where spatial or temporal relationships are crucial. In our project, the CNN algorithm played a pivotal role in accurately predicting electricity prices by leveraging its ability to extract and analyze relevant spatial patterns in the input data.

A paper presents a method for predicting electricity prices using a combination of structured deep neural networks and model for predicting electricity prices using a combination of two types of neural networks, which are computer programs designed to learn and make predictions. The first type is called a "structured" neural network, which means it uses specific patterns and relationships in the data to make predictions. The second type is a "deep" neural network, which is a more complex model that can learn from large amounts of data. By combining these two types of neural networks, the researchers aim to create a more accurate and reliable electricity price forecasting model. This model takes into account various factors that influence electricity prices, such as historical data, weather conditions, and market trends. By analyzing these factors, the model can make predictions about future electricity prices. The goal of the research is to develop a model that can help individuals and businesses make informed decisions about their electricity usage and plan accordingly. By knowing the expected electricity prices in advance, people can optimize their energy consumption, save costs, and make more sustainable choices.

3.4 Long Short-Term Memory

It is a type of RNN specifically designed to handle sequential data. It overcomes the limitation of traditional RNNs in capturing long-term dependencies by incorporating memory cells with gating mechanisms. These gates allow the network to selectively remember or forget information over time, making it well-suited for modeling sequences with long-range dependencies. LSTM networks effectively capture patterns and trends in electricity price data by considering historical price values as sequential inputs. The memory cells enable the model to retain relevant information from past time steps while updating and adapting to new inputs. This ability to capture temporal dependencies and preserve context makes LSTM particularly powerful for time series forecasting tasks. By leveraging its memory cells and gating mechanisms, LSTM helps in accurately predicting future electricity prices based on historical data.

In one paper, the authors propose a method for predicting electricity prices using a combination of two techniques: LSTM and differential evolution algorithm (DE) [16]. Let's break down what these terms mean:

- Electricity price prediction: This refers to the task of estimating the future prices of electricity. It is important because it helps energy market participants, such as consumers and producers, make informed decisions about buying or selling electricity.
- Long short-term memory: LSTM is a type of artificial neural network, which is a computer model inspired by the human brain. LSTM is particularly good at processing sequences of data, such as time series. In the context of electricity price prediction, LSTM can analyze historical electricity prices and find patterns or relationships that can be used to make predictions about future prices.

So, it suggests a method to forecast electricity prices in the future. It combines a powerful type of neural network called LSTM with an optimization algorithm called DE. By using historical data and fine-tuning the model's parameters, to improve the accuracy of the predictions, which can help people in the energy market make better decisions.

3.5 Artificial Neural Network

These are computational models inspired by the structure and functionality of the human brain. They consist of interconnected nodes, called artificial neurons or perceptron, organized in layers. Each neuron receives input signals, applies a weighted sum and an activation function, and passes the output to the next layer. ANNs can learn complex nonlinear relationships by adjusting the weights through a process called backpropagation. With their ability to learn from large amounts of data, ANNs can capture intricate patterns and dependencies in electricity price data. By adjusting the architecture, activation functions, and training parameters, ANNs can model and predict electricity prices accurately. ANNs are versatile and can handle various types of data, making them suitable for price forecasting tasks. In our project, the ANN algorithm played a crucial role in predicting electricity prices by leveraging its ability to learn and generalize from historical data, ultimately providing accurate price predictions for future time periods.

In another paper predict electricity prices using a type of artificial intelligence called artificial neural networks. They want to develop a model that can forecast electricity prices accurately. Artificial neural networks are computer systems designed to mimic the human brain's ability to learn and process information. They consist of interconnected nodes called neurons that work together to solve complex problems. These networks can learn patterns and relationships from historical data and use that knowledge to make predictions. Electricity prices can fluctuate based on various factors such as supply and demand, weather conditions, and market dynamics. To forecast these prices, the authors feed historical data, such as past electricity prices and related information, into the artificial neural network. The network then processes this data and learns the patterns and relationships between the input variables and the electricity prices. Once the network has learned from the historical data, it can be used to predict future electricity prices. By inputting relevant information about the current conditions, such as weather forecasts and energy market trends, into the trained neural network, it can generate forecasts of electricity prices. The authors are likely exploring different types of neural network architectures and training algorithms to find the most accurate model for electricity price forecasting. Their goal is to improve the accuracy of these predictions, which can be valuable for various stakeholders in the electricity industry, such as power producers, consumers, and policymakers.

3.6 Gated Recurrent Network

It is a variant of the RNN architecture that addresses some of the limitations of traditional RNNs. It introduces gating mechanisms to selectively update and reset the hidden state, allowing for better long-term memory retention and gradient flow. GRUs have a simpler structure compared to other RNN variants like LSTM but still perform well in capturing temporal dependencies. By leveraging gates that control the flow of information, GRUs can effectively capture patterns in time series data, such as electricity prices. The gating mechanisms enable the model to adaptively update and forget information over time, which is crucial for accurate price predictions. GRUs strike a balance between model complexity and performance, making them computationally efficient and suitable for practical applications. In our project, the GRU algorithm played a crucial role in predicting electricity prices by effectively modeling and capturing long-term dependencies in the data, ultimately providing accurate and reliable price forecasts.

3.7 Recurrent Neural Network

They have proven to be a powerful tool in various fields, including time series analysis and natural language processing. In the context of your project, the Electricity

Price Prediction Model, RNNs offer a promising approach. RNNs are designed to process sequential data by maintaining a hidden state that captures information from previous time steps. This makes them well-suited for capturing temporal dependencies and patterns in electricity price data.

By using RNNs, your model can learn to analyze the historical trends and patterns in electricity prices, allowing it to make predictions for future prices. The ability of RNNs to handle variable-length input sequences makes them flexible for dealing with different time series lengths and capturing long-term dependencies. One of the most popular variants of RNNs is the LSTM network. LSTMs address the vanishing gradient problem, which can occur when training deep neural networks. They achieve this by incorporating memory cells, which can retain information over long sequences and selectively forget or remember information based on the context. Another variant is the GRU, which is simpler than LSTM but still capable of capturing long-term dependencies. GRUs have fewer gates and parameters, making them computationally more efficient and faster to train compared to LSTMs, while still delivering competitive performance.

RNNs can be trained using various optimization techniques such as gradient descent, backpropagation through time, or more advanced algorithms like Adam or RMSprop. These techniques aim to find the optimal set of weights and biases for the network, minimizing the prediction error. However, it's important to note that while RNNs can be effective, they also have some limitations. For instance, they can struggle with long sequences due to vanishing or exploding gradients. Additionally, RNNs are computationally more expensive than other algorithms, which can pose challenges for training and deployment in large-scale applications.

In conclusion, by leveraging RNN like LSTMs and GRUs, we developed an electricity price prediction model that can capture temporal dependencies and patterns in electricity price data. These algorithms provide a powerful framework for forecasting future prices, enabling decision-makers to make informed choices in the electricity market. Nevertheless, it's crucial to consider the limitations and challenges associated with RNNs while developing and deploying such models.

3.8 Radial Basis Function

It is a type of artificial neural network that uses radial basis functions as activation functions. It is particularly useful for solving regression problems, such as predicting electricity prices. RBF networks consist of three layers: an input layer, a hidden layer with radial basis functions as activation functions, and an output layer. The hidden layer computes the distance between the input data and a set of centers, and the output layer produces the final prediction. RBF networks are known for their ability to approximate complex nonlinear relationships in data. They can capture intricate patterns and dependencies in electricity price data, allowing for accurate price predictions. RBF networks are relatively simple to train and have fewer parameters compared to other neural network architectures. In our project, the RBF algorithm played a crucial role in accurately predicting electricity prices by leveraging the power of radial basis functions to model and generalize from historical price data.

Another paper discusses a type of mathematical model called an "enhanced radial basis function network" and how it can be used for predicting short-term electricity prices. Let's break it down into simpler terms. When we talk about short-term electricity price forecasting, we mean predicting how much the price of electricity will be soon. This is important because it helps power companies and consumers plan their energy usage and make decisions accordingly. The model "enhanced radial basis function network," is a way to make these predictions. It's a mathematical approach that uses a network of nodes, called radial basis functions, to analyze historical data and find patterns. These functions help the model understand how different factors, like time of day, weather conditions, and energy demand, affect electricity prices. What makes this model "enhanced" is that the researchers have made some improvements to it. They have introduced additional features or techniques to make the predictions more accurate. For example, they might have included new variables or found a better way to train the model using the available data. Overall, the paper describes how this enhanced radial basis function network can be used as a tool for forecasting short-term electricity prices. It's a way to make predictions based on historical data and take into account various factors that influence electricity prices. Each algorithm used in our study holds specific strengths and advantages, enabling accurate prediction of electricity price fluctuations. By leveraging their respective capabilities, we aimed to achieve the most precise and reliable predictions, as demonstrated by the evaluation metrics employed in our analysis.

4. Dataset Description

The data used in this paper contains the electricity prices taken from Australian Energy Market Operator website presented on half-hourly basis. The data consists of half-hourly electricity prices for everyday from Jan 2019 to Mar 2023. The dataset consists of multiple columns that capture various relevant factors and parameters related to electricity prices and load. The detailed description of each column is as follows:

Settlement Date: This feature represents the specific date on which the electricity price was settled. This column allows for the temporal analysis of price fluctuations over time and enables the identification of seasonal patterns or trends.

RRP: This contains the retail electricity prices recorded at specific times. These prices represent the cost of electricity in the retail market and are essential for understanding the dynamics of electricity pricing and forecasting future price movements.

Hour of the Day: The hour of the day feature provides the precise hour at which the electricity price was recorded. This information is crucial for capturing diurnal patterns and analyzing how prices vary throughout the day.

Day of the Week: This feature in the dataset represents the day of the week corresponding to each observation. It is encoded with numerical values ranging from 0 to 6, where 0 represents Sunday, 1 represents Monday, and so on until 6, which represents Saturday. This numerical representation enables machine learning models to capture cyclic patterns and dependencies between electricity prices and specific days of the week. The research can explore variations in electricity prices based on weekdays and weekends, analyze day-based effects to implement models that consider the day of the week as an important factor in electricity price forecasting.

24-hr Average Load: This feature represents the average electricity load over a 24-hour period. This information helps to understand the overall electricity consumption patterns and their relationship with price fluctuations. Analyzing load patterns is crucial for accurate price forecasting as high-demand periods often correspond to higher prices.

24-hr Lagged Load: This feature captures the electricity load at the same hour on the previous day. Lagged load information is useful for identifying dependencies and correlations between current and past load patterns, which can provide valuable insights into price behavior and help develop predictive models.

Previous 24-hr Average Load: This feature represents the average electricity load over a 24-hour period on the previous day. This feature allows for the examination of load trends and the impact of historical load patterns on current price variations.

Previous Week: The Previous Week feature provides additional information about the previous week, which include previous week electricity load. This data allows for the analysis of weekly seasonality and the impact of external factors on price dynamics.

Weekday: The day of the week feature indicates the specific day corresponding to the settlement date. This categorical information enables the examination of weekly patterns.

The dataset provides a comprehensive and rich set of features necessary for conducting advanced research in electricity price prediction. It offers a unique opportunity to explore the intricate relationship between price variations, load patterns, temporal factors, and external influences, ultimately contributing to the development of effective forecasting models for the energy market.

5. Experiments and Results

The models were implemented using popular libraries such as scikit-learn and TensorFlow. Each model was trained and evaluated multiple times to account for randomness in the initialization. For each experiment, we employed different machine learning models. The results and various analysis are obtained are like this:

5.1 Ordinary Least Squares Regression Analysis:

A statistical method known as OLS regression is used to represent the relationship between the dependent variable and any number of independent variables [21]. It is performed on the entire dataset. The OLS regression results refer to the output generated from running an OLS regression analysis on a given dataset.

The OLS regression results typically provide information about the estimated coefficients, statistical significance, and the regression model's ability to fit data well. The following are the common components of OLS regression results:

Table 2 – OLS Analysis

	r	**Std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	-29.7681	2.514	-11.842	0.000	-34.695	-24.841
Hour Of Day	1.1056	0.064	17.142	0.000	0.979	1.232
Day of the Week	-0.0483	0.448	-0.108	0.914	-0.927	0.830
24-hr Average Load	0.5895	0.012	47.958	0.000	0.565	0.614
24-hr Lagged Load	0.0453	0.004	12.526	0.000	0.038	0.052
Previous 24-hr Average Load	0.5891	0.014	41.080	0.000	0.561	0.617
Previous Week	-0.1955	0.019	-10.429	0.000	-0.232	-0.159

Correlation coefficient (r)

These values represent the estimated impact of the independent variables on the dependent variable. Each independent variable has a coefficient that, represents the change in the dependent variable that results from a one-unit change in the corresponding independent variable, while holding other variables constant [22].

Standard Error (Std err)

These are measures of the variability or uncertainty associated with the estimated coefficients. They help determine the precision of the coefficient estimates. More precise forecasts are shown by smaller standard errors [23].

t-statistics

They are calculated by dividing the coefficient estimates by their respective standard errors. They indicate the number of standard deviations the estimated coefficient is away from zero. Larger absolute t-values generally suggest more significant relationships [24].

p-values

The values in the "P>|t|" column indicate the p-values associated with each coefficient. The p-values [25] associated with the t-statistics indicate the probability of observing the estimated coefficient if the null hypothesis (there is no correlation between the independent and dependent variables) [26] is true. Lower p-values (<0.05) suggest stronger evidence against the null hypothesis and indicate a statistically significant relationship. In this case, the p-values for all features, except "Day of the Week," are very close to zero, indicating that they are statistically significant.

From Table 2,

- Intercept: It represents the estimated value of the dependent variable when all other independent variables are held constant at zero. Here, (r=-29.7681) value suggests that when all other variables are zero, the expected value of the dependent variable is r. The std err for the intercept is (2.514) which means that the estimated coefficient (r) of the intercept is expected to vary by approximately (2.514) units. The t-value obtained by dividing r value with std err indicates that the estimated intercept is significantly different from zero.

- Hour of the Day: The correlation coefficient (r=1.1056) for this feature indicates that for every one unit increase in the hour of the day, the expected value of the dependent variable RRP increases by r units, assuming all other variables are held constant. The std err (0.064) suggests that the estimated coefficient of (r) is expected to vary by approximately (0.064) units indicating a precise estimate. The calculated t-value (17.142) indicates a statistically significant relationship between this feature and the dependent feature.

- Day of the Week: The correlation coefficient (r=-0.0483) suggests that there is a very small negative relationship between this feature and the dependent variable RRP. The larger std err (0.448) indicates a higher level of uncertainty associated with this estimated coefficient. The t-value (-0.108) implies that the estimated coefficient is not significantly different from 0. This is supported by the high p-value (0.914), which suggests that there is no statistically significant relationship between this feature and the dependent feature.

- 24-hr Average Load: The correlation coefficient (r=0.5895) value indicates that for every one unit increase in the 24-hour average load, the expected value of the dependent variable RRP increases by r units, assuming all other variables are held constant. The std error for this feature (0.012) suggests that the estimated coefficient of r is expected to vary by approximately (0.012) units indicating a high level of precision in the estimate. The large t-value (47.958) indicates the coefficient is significantly different from 0, showing a strong and significant relationship between this feature and the dependent variable.

- 24-hr Lagged Load: The correlation coefficient (r=0.0453) suggests that for every one unit increase in the 24-hour lagged load, the expected value of the dependent variable RRP increases by r units, assuming all other variables are held constant. The std err here (0.004) means that the estimated coefficient (r) is expected to vary by approximately (0.004) indicating a precise estimate. The larger t-value (12.526) indicates that the estimated coefficient (r) is significantly different from 0, suggesting a statistically significant relationship between this feature and the dependent variable.

- Previous 24-hr Average Load: The correlation coefficient (r=0.5891) suggests that for every one unit increase in the previous 24-hour average load, the expected value of the dependent variable RRP increases by r units, assuming all other variables are held constant. The std error (0.014) suggests that the estimated coefficient (r) is expected to vary by approximately (0.014) units indicating a high level of precision in this estimate. The large t-value (41.080) indicates that the estimated coefficient is significantly different from zero, indicating a strong and statistically significant relationship between this feature and the dependent variable

- Previous Week: The correlation coefficient (r=-0.1955) suggests that there is a negative relationship between the previous week and the dependent variable. For every one unit increase in the previous week, the expected value of the dependent variable RRP decreases by r units, assuming all other variables are held constant. The std err (-0.019) suggests that the estimated coefficient (r) is expected to vary approximately by 0.019 units suggesting a precise estimate. This t-value indicates that the estimated coefficient (-0.1955) is significantly different from zero, suggesting a statistically significant relationship between this variable and the dependent variable

Omnibus

This statistic measures the overall departure from normality of the residuals. In the output we got, the Omnibus value is (212188.831) which indicates a significant departure from normality [27, 28].

Skewness

It measures the asymmetry of the distribution of residuals. In this case, the skewness value is reported as (38.722) which is a positive skewness value that indicates a right-skewed distribution, meaning that the residuals have a tail that extends more towards larger positive values [28, 29].

Kurtosis

It measures the heaviness of the tails of the distribution of residuals and the kurtosis value is reported as (1890.875) and this high value suggests that the residuals have heavy tails and exhibit more extreme values than a normal distribution [29].

5.2 T-test

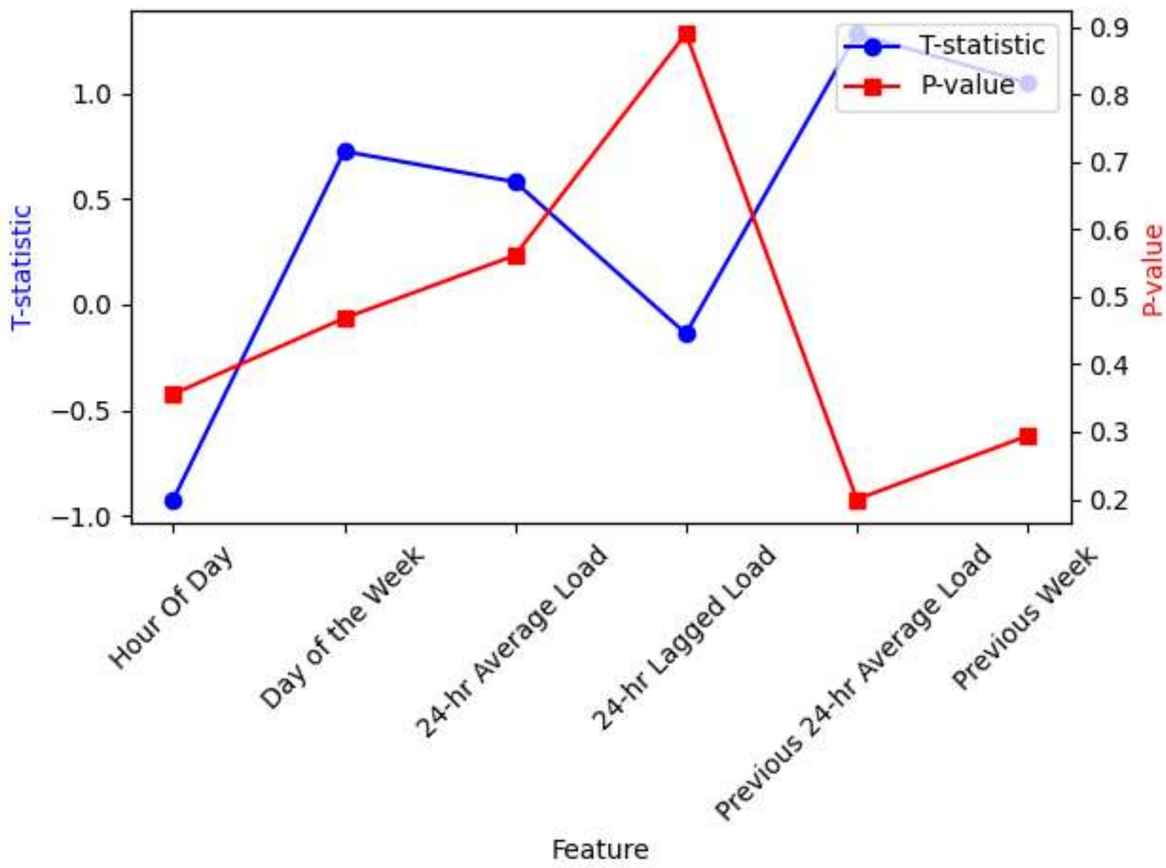

Figure 1: The graph of t-statistic and p-value for different parameters

To determine if there is a crucial difference between the means of two groups or samples, this statistical hypothesis T-test [30] that is used. Here the T-test is performed on the test dataset which is 30% of the entire dataset by dividing it into 2 groups that are group1 and group2.

Assumptions are made in the t-test, including the assumption of normality and the assumption of equal variances between the groups. Violations of these assumptions can impact the validity of the test results. Overall, the t-test provides a framework for comparing means and determining if observed differences are statistically significant, helping to draw meaningful conclusions about the groups being compared.

The t-statistic measures the difference between the means of the two groups, relative to the variability within each group. The p-value indicates the probability of observing such a difference in means if the null hypothesis (no difference between the groups) were true. A smaller p-value (<0.05) suggests stronger evidence against the null hypothesis.

In Figure 1, t-statistic, std error and p-value for each feature indicates:

- For the feature 'Hour Of Day' the t-statistic value (-0.92282864) indicates the difference in the mean of the feature between the two groups. The p-value (0.35610651) indicates that the relationship between 2 groups is are statistically significant at conventional significance level means there is not enough evidence to reject the null hypothesis.
- For the feature 'Day of the Week' the t-statistic value (0.7250981) suggests that there is a positive relationship between the two groups. The p-value (0.46839946) indicates that the relationship between 2 groups are not statistically significant at conventional significance level.
- For the feature '24-hr Average Load' the t-statistic value (0.58050355) suggests that there is a positive relationship between the two groups. The p-value (0.56158097) indicates that the relationship between 2 groups are not statistically significant at conventional significance level.
- For the feature '24-hr Lagged Load' the t-statistic value (-0.1382742) suggests that there is a negative relationship between the two groups. The p-value (0.89002497) indicates that the relationship between 2 groups is not statistically significant at conventional significance level.
- For the feature 'Previous 24-hr Average Load' the t-statistic value (1.28191522) suggests that there is a positive relationship between the two groups. The p-value (0.19988573) is relatively low but still above the conventional significance level of 0.05. Therefore, while the relationship is not statistically significant, there is some weak evidence to suggest that the previous 24-hour average load may have a potential impact on the differ across groups.

- For the feature 'Previous Week' the t-statistic value (1.04915157) suggests that there is a positive relationship between the two groups. The p-value (0.29411975) indicates that the relationship between 2 groups are not statistically significant at conventional significance level.

5.3 Wilcoxon Test

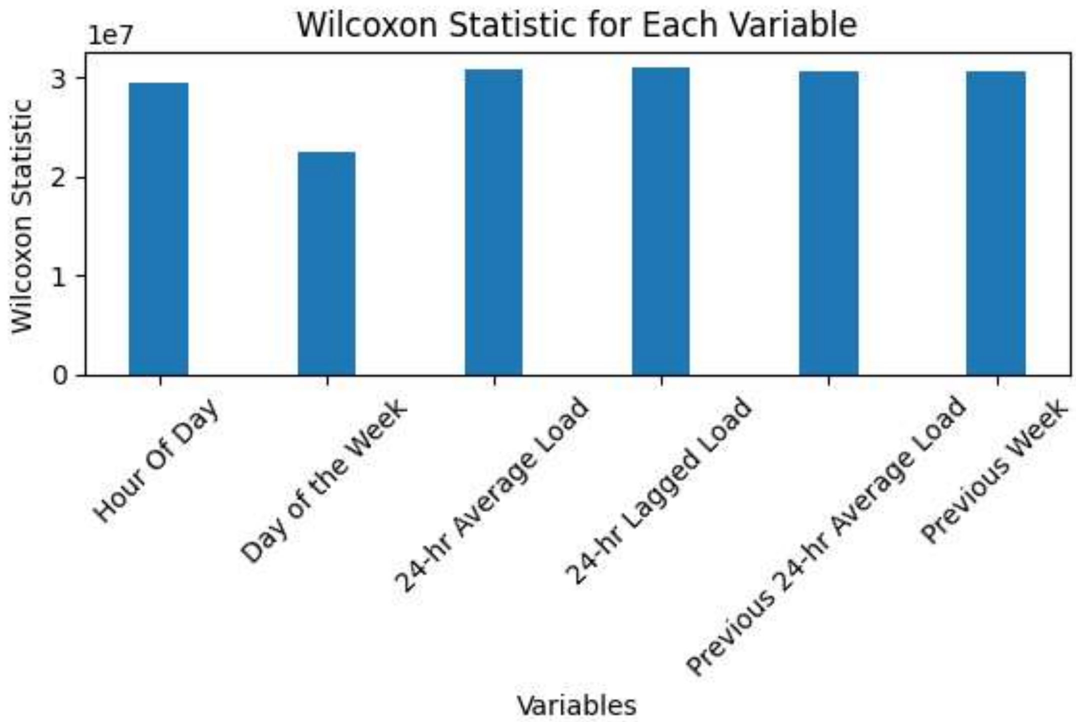

Figure 2: Graph of Wilcoxon static value for different parameters

A non-parametric statistical test which is Wilcoxon signed-rank test [31] is used to compare two related samples or paired data. It is especially helpful for analysing data that might not satisfy the assumptions of parametric tests, such as normality. This report presents the findings of a Wilcoxon signed-rank test performed in Python and the SciPy package on a dataset. To see if there was a significant difference between the two groups, the Wilcoxon signed-rank test was used. The test findings provided the Wilcoxon statistic and p-values for each feature variable. Here the T-test is performed on the test dataset which is 30% of the entire dataset. The Wilcoxon statistic evaluates the size and direction of the difference between the paired observations.

In Figure 2, the Wilcoxon signed-rank value for different features is given as:

- Hour Of Day – 29,500,579
- Day of the Week – 22,546,282.5
- 24-hr Average Load – 30,749,908.5
- 24-hr Lagged Load – 30,964,680.5

- Previous 24-hr Lagged Load – 30,686,490
- Previous Week – 30,601,721.5

These values indicate the ranks assigned to the differences between paired observations for every feature.

5.4 Results

The table below presents the performance metrics of different models used in the research study. These metrics provide insights into the models' accuracy, precision, and predictive power.

Table 1 – Performance metrics obtained for different models

Model	MSE	MAE	MAPE	R2 Score	RMSE
RFR	***0.000103778***	***0.001***	***0.017904629***	***0.4008146***	***0.010187131***
RR	0.000147049	0.003	0.034267947	0.150976087	0.012126382
RNN	0.000151463	0.003	0.040517103	0.125493475	0.012307018
ANN	0.000173206	0.004	0.053788207	-4.72E-05	0.013160778
GRU	0.000147871	0.003	0.033076812	0.146229015	0.012160236
CNN	0.000138267	0.003	0.033288539	0.201684143	0.011758683
LSTM	0.00013764	0.002	0.032128902	0.205303028	0.011732001
RBF	0.000147514	0.003	0.038397667	0.148289825	0.012145551

Note: Bold – italic values indicate the best values of performance metrics.

The following details are provided by the performance metrics [32, 33]:

From Table 1, we can derive that the metrics indicates as follows:

Mean Squared Error

The average squared difference between the expected and actual numbers. RFR has the lowest MSE of 0.000103778, indicating the most accuracy [34].

Mean Absolute Error

The average absolute difference between the expected and actual values is what is measured. Greater precision is indicated by lower numbers, with RFR having the lowest MAE of 0.001 [34].

Mean Absolute Percentage Error

The average percentage difference between the expected and actual values is calculated. RFR has the lowest MAPE of 0.017904629, indicating more precision [34].

R2 Score

The coefficient of determination, often known as the R2 score, is a statistical indicator that shows how much of the variance in the dependent variable can be accommodated by the independent variables in a regression model. It has a value between 0 and 1, where a value of 0 means that the model does not explain any variance and a value of 1 means that the model perfectly fits the data and fully accounts for all variance. Higher values indicate a better fit to the data, with RFR having the highest R2 Score of 0.4008146 [34].

Root Mean Squared Error

The average size of the prediction error is represented by the square root of the MSE. RFR has the lowest RMSE of 0.010187131 and lower values indicate better accuracy [34].

The ***Random Forest Regression (RFR)*** model, which has the lowest MSE, MAE, and MAPE among the models, exhibits the highest overall performance according to the performance measures.

6. Concluding Remarks

In this study, we focused on electricity price forecasting using machine learning techniques. We explored various algorithms and evaluated their performance in predicting electricity prices accurately. Our dataset included relevant factors such as settlement date, retail electricity prices, hour of the day, day of the week, average load, lagged load, previous 24-hour average load, previous week, and weekday.

The methodology employed in our study involved the use of machine learning algorithms, including RFR, RR, RNN, ANN, GRU, CNN, LSTM and RBF. Each algorithm was trained and evaluated multiple times, and performance metrics such as MSE, MAE, MAPE, R2 Score, and RMSE were used to assess their effectiveness.

Among the algorithms we tested, ***Random Forest Regressor*** demonstrated the ***best*** performance. This algorithm proved to be robust, accurate, and capable of capturing the complex nonlinear relationships between input features and the target variable. Our study contributes to the field of electricity price forecasting by showcasing the efficacy of machine learning techniques in predicting price fluctuations.

In conclusion, our research highlights the importance of accurate electricity price forecasting and the potential of machine learning algorithms to contribute to this field and the findings can guide market participants in making informed decisions, improving risk management, and optimizing their operations in the dynamic energy market.

7. Future Work

In future we will plan to focus on exploration of hybrid models by combining different machine learning algorithms, application of optimization algorithms to enhance the training and performance of electricity price prediction models, incorporate external factors that influence electricity prices into prediction models by considering factors such as weather data, renewable energy generation, economic indicators and grid load. We also plan to think in the direction to develop techniques to quantify the uncertainty associated with electricity price predictions. To extend the analysis on real-time prediction scenarios and addressing the challenge of model interpretability by incorporating different techniques. The following are the things we

- Investigate the integration of CNN and LSTM networks.
- Explore the combination of GRU and LSTM networks.
- Ensemble multiple machine learning models, such as ANN, CNN, RF, RR, GRU and LSTM.
- Investigate the use of genetic algorithms to optimize the hyperparameters of machine learning models.
- Explore the application of PSO algorithms to optimize the parameters and architecture of machine learning models.
- Investigate the potential of reinforcement learning techniques for training electricity price prediction models.

Implementing these future works can provide valuable insights on energy market and facilitate informed decision-making processes.

8. References

1. N. Chaˆabane, A hybrid arfima and neural network model for electricity price prediction, International Journal of Electrical Power Energy Systems 55 (2014) 187–194
2. U. B. Tayab, A. Zia, F. Yang, J. Lu, M. Kashif, Short-term load forecasting for microgrid energy management system using hybrid hho-fnn model with best-basis stationary wavelet packet transform, Energy 203 (2020) 117857.
3. W. Li, D. M. Becker, Day-ahead electricity price prediction applying hybrid models of lstm-based deep learning methods and feature selection algorithms under consideration of market coupling, Energy 237 (2021) 121543.

4. A. Arif, N. Javaid, M. Anwar, A. Naeem, H. Gul, S. Fareed, Electricity load and price forecasting using machine learning algorithms in smart grid: A survey, in: Web, Artificial Intelligence and Network Applications: Proceedings of the Workshops of the 34th International Conference on Advanced Information Networking and Applications (WAINA-2020), Springer, 2020, pp. 471–483
5. R. Weron, Electricity price forecasting: A review of the state-of-the-art with a look into the future, International journal of forecasting 30 (4)(2014) 1030–1081
6. Y. Wang, W. Liao, Y. Chang, Gated recurrent unit network-based short-term photovoltaic forecasting, Energies 11 (8) (2018) 2163.
7. A. Almalaq, G. Edwards, A review of deep learning methods applied on load forecasting, in: 2017 16th IEEE international conference on machine learning and applications (ICMLA), IEEE, 2017, pp. 511–516.
8. S. K. Aggarwal, L. M. Saini, A. Kumar, Electricity price forecasting in deregulated markets: A review and evaluation, International Journal of Electrical Power & Energy Systems 31 (1) (2009) 13–22.
9. R. Shreyas, D. Akshata, B. Mahanand, B. Shagun, C. Abhishek, Predicting popularity of online articles using random forest regression, in: 2016 Second International Conference on Cognitive Computing and Information Processing (CCIP), IEEE, 2016, pp. 1–5.
10. P. Wang, K. Xu, Z. Ding, Y. Du, W. Liu, B. Sun, Z. Zhu, H. Tang, An online electricity market price forecasting method via random forest, IEEE Transactions on Industry Applications 58 (6) (2022) 7013–7021.
11. G. C. McDonald, Ridge regression, Wiley Interdisciplinary Reviews: Computational Statistics 1 (1) (2009) 93–100.
12. S. Albawi, T. A. Mohammed, S. Al-Zawi, Understanding of a convolutional neural network, in: 2017 international conference on engineering and technology (ICET), Ieee, 2017, pp. 1–6.
13. P.-H. Kuo, C.-J. Huang, An electricity price forecasting model by hybrid structured deep neural networks, Sustainability 10 (4) (2018) 1280.
14. A. Graves, A. Graves, Long short-term memory, Supervised sequence labelling with recurrent neural networks (2012) 37–45.
15. L. Peng, S. Liu, R. Liu, L. Wang, Effective long short-term memory with ifferential evolution algorithm for electricity price prediction, Energy 162 (2018) 1301–1314.
16. S. Agatonovic-Kustrin, R. Beresford, Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical re- search, Journal of pharmaceutical and biomedical analysis 22 (5) (2000) 717–727.
17. D. Singhal, K. Swarup, Electricity price forecasting using artificial neural networks, International Journal of Electrical Power & Energy Systems 33 (3) (2011) 550–555.
18. U. Ugurlu, I. Oksuz, O. Tas, Electricity price forecasting using recurrent neural networks, Energies 11 (5) (2018) 1255.
19. S. A. Billings, G. L. Zheng, Radial basis function network configuration using genetic algorithms, Neural Networks 8 (6) (1995) 877–890.

20. W.-M. Lin, H.-J. Gow, M.-T. Tsai, An enhanced radial basis function network for short-term electricity price forecasting, Applied Energy 87 (10) (2010) 3226–3234.
21. L. Leng, T. Zhang, L. Kleinman, W. Zhu, Ordinary least square regression, orthogonal regression, geometric mean regression and their applications in aerosol science, in: Journal of Physics: Conference Series, Vol. 78, IOP Publishing, 2007, p. 012084.
22. E. W. Weisstein, Correlation coefficient, https://mathworld. wolfram. com/ (2006).
23. D. G. Altman, J. M. Bland, Standard deviations and standard errors, Bmj 331 (7521) (2005) 903.
24. F. J. Gravetter, L. B. Wallnau, Introduction to the t statistic, Essentials of statistics for the behavioral sciences 8 (252) (2014).
25. J. P. Ioannidis, The proposal to lower p value thresholds to. 005, Jama 319 (14) (2018) 1429–1430.
26. R. L. Hagen, In praise of the null hypothesis statistical test. (1997).
27. S. Raschka, Model evaluation, model selection, and algorithm selection in machine learning, arXiv preprint arXiv:1811.12808 (2018).
28. K. V. Mardia, Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies, Sankhy ̄a: The Indian Journal of Statistics, Series B (1974) 115–128.
29. M. J. Blanca, J. Arnau, D. L ́opez-Montiel, R. Bono, R. Bendayan, Skewness and kurtosis in real data samples, Methodology (2013).
30. H. Keselman, A. R. Othman, R. R. Wilcox, K. Fradette, The new and improved two-sample t test, Psychological Science 15 (1) (2004) 47–51.
31. D. Rey, M. Neuh ̈auser, Wilcoxon-signed-rank test, in: International encyclopedia of statistical science, Springer, 2011, pp. 1658–1659.
32. M. Naser, A. H. Alavi, Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences, Architecture, Structures and Construction (2021) 1–19.
33. M. Naser, A. Alavi, Insights into performance fitness and error metrics for machine learning, arXiv preprint arXiv:2006.00887 (2020).
34. Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 7, e623.