Impact of Weather Conditions onRenewable Energy Consumption

MSc Research Project
Data Analytics

Shivani Shrikant Nandanikar
Student ID: x22197168

School of Computing
National College of Ireland

Supervisor: Dr. Catherine Mulwa

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Shivani Shrikant Nandanikar……………………………….

Student ID: X22197168…………………………………………………..……

Programme: MSc. Data Analytics………………………… **Year:** 2023-2024

Module: MSc. Research Project…………….………

Supervisor: Dr. Catherine Mulwa…………………….………

Submission Due Date: 12/08/2024…………………………….………

Project Title: Impact of Weather Conditions on Renewable Energy Consumption

Word Count: 7388 **Page Count** 28

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Shivani Shrikant Nandanikar……………………

Date: 12/08/2024…………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Impact of Weather Conditions on Renewable Energy Consumption

Shivani Nandanikar
x22197168

Abstract

The proposed study examines the impact of the weather conditions like temperature, pressure, humidity, rain, wind and solar energy on renewable energy consumption, focusing on studying the energy consumption patterns and impact of combination of complementary energy sources like solar and wind energy on energy consumption. Use of machine learning models along with performing the cross-validation on ML models such as Polynomial Regression, Random Forest, and Gradient Boosting, also time series analysis is being performed by using the combination of time series models like ARIMA, SARIMA + GARCH, and the multivariate VAR model, the research analyses energy consumption patterns over time. The key findings showed that amongst the applied three ML algorithms Random Forest outperformed with an 89% accuracy in predicting energy consumption based on weather variables like temperature, pressure, humidity, etc. The VAR model highlights the complementary relationship between solar and wind energy, which suggested a more stable energy supply when these two complementary energy sources are combined. This research provides significant insights into optimizing renewable energy consumption, which will help policymakers to create and develop sustainable energy strategies to reduce climate change impacts.

1 Introduction

In recent years, the topic of energy around the world has been shifted to focus on the importance of renewable energy to obtain sustainable development, thereby finding a solution for the same. This shift has been brought about due to the rapid depletion of non-renewable energy sources and the urgent need to deal with environmental issues, especially climate change. The depletion of fossil fuels and other non-renewable resources leads to a significant increase in greenhouse gas emissions, contributing to air pollution and climate change respectively. Investigating and making use of renewable energy sources, such as solar, wind, hydroelectric, and geothermal power, has therefore become an appropriate option. These energy sources not only help reduce the negative environmental impacts of fossil fuels but also provide a path for a more secure and sustainable energy future.

Switching to renewable resources is essential as traditional energy sources are becoming less feasible. Since renewable energy sources can naturally replenish themselves over time, they can supply energy continuously without depleting the planet's natural resources. Making this switch to renewable energy has many advantages. The main benefit is that it significantly decreases greenhouse gas emissions, which are the main contributor to climate change. Also, by decreasing pollutants that cause smog and respiratory diseases, the renewable energy

sources enhance air quality. From an economic point of view, sustaining renewable energy can lower the financial risks related to unpredictable fossil fuel markets and open job opportunities in new industries. Another important benefit is increased energy security, as different nations can use readily available renewable resources and reduce their dependency on imported fuels.
The goal of this research paper is to present a thorough study of the consumption of renewable energy over time. The study will find patterns in consumption of renewable energy sources by examining the collected data. The proposed research implements a comprehensive study of various machine learning algorithms like Polynomial regression, Random Forest, and Gradient Boosting Models; time series models like ARIMA and SARIMA + GARCH models, and multivariate model like VAR model to find the best fit model amongst the three ML algorithms, forecast the energy consumption patterns, and study the impact of combining the complementary renewable energy sources like solar and wind energy on the energy consumption respectively.

Hence, the main research question guiding this study is:

Research Question: *How do consumption patterns of renewable energy sources vary over time, considering their dependencies on external factors of weather conditions like temperature, pressure, humidity, rain, wind speed to support sustainable energy management strategies and enhance energy forecast accuracy?*

And the sub-research question examined in this study is as follows:

Sub-RQ1: *What trends and seasonal variations can be identified in the usage of energy, and how can time series analysis be used to predict future consumption patterns effectively?*

Sub-RQ2: *What role does complementary renewable energy sources play in reducing intermittency issues, and how well can a multivariate time series model improve forecasting accuracy?*

Objectives and Contributions

The objective of this study is to identify the variables that affect the consumption patterns of renewable energy and evaluate to understand which external factors impact these consumption patterns.

Table 1 explains those objectives, methods used to fulfill them, and the evaluation metrics used to measure them.

Table 1: Project Objectives and Contributions

Sr. No	Objectives	Methods Used	Evaluation Metrics
1	Apply regression models for consumption prediction, along with applying Random Forest and Gradient Boosting for comparison.	Linear/polynomial regression, feature selection	R^2 value, sum-squared residuals, prediction accuracy
2.1	Analyze year-by-year renewable energy consumption trends	Time series analysis	Yearly consumption trends, descriptive statistics
2.2	Analyze seasonal (monthly) consumption patterns	Time series analysis, Seasonal decomposition	Monthly consumption trends, seasonal patterns
2.3	Identify hourly consumption patterns and forecast	Hourly data analysis, SARIMA + GARCH models	Hourly consumption trends, peak usage times
3	Develop ARIMA model for time-based prediction	ARIMA modelling, parameter selection (p, q values)	Forecast accuracy, residuals
4	Develop VAR model for integrating multiple factors and time	Vector Auto Regressive (VAR) modelling, feature selection	Prediction accuracy, residuals, feedback error minimization
5	Validate model predictions against test data	Model validation, 70-30 data split	Test data prediction accuracy, comparison with actuals
6	Evaluate overall model effectiveness	Comparison of models, residual analysis	Model performance metrics, residuals, forecast accuracy

Project Contributions:

This research significantly contributes to the field of renewable energy consumption analysis in several ways. Firstly, it provides a complete understanding of the time patterns in the consumption of renewable energy using complex time series analysis methods such as VAR and ARIMA models. The study offers detailed understanding of the dynamics of consumption of energy by identifying and analyzing the hourly, seasonal, and yearly consumption patterns of renewable energy. Secondly, the study focuses on the important variables that affect the consumption of renewable energy, including laws and regulations, economic conditions, technological developments, and social views on sustainability. The research explains the relationships between these inter-related variables and how they affect the patterns of energy consumption by using some complex statistical methods such as feature extraction and correlation matrices respectively. Also, the process of developing and evaluating predictive models improves the ability to accurately predict future energy requirements, which thereby helps the industry leaders and policymakers in planning their strategy. A comparative study of various modelling techniques not only contributes to our understanding of how to include a variety of variables that influence the predictive modelling, but also helps to select the best modelling technique to study the energy consumption patterns. Therefore, ultimately this research encourages the global transition to renewable energy sources by providing useful insights and complex statistical models that encourage sustainable energy practices respectively.

2 Related Work

The literature review of the proposed study will give a better understanding of various studies related to the topic of renewable energy consumption. This section will be divided into 4 sub-sections, namely- Introduction, Analysis of Usage of Traditional Machine Learning Regression Methods for Prediction of Renewable Energy Consumption, Study of Time Series Model for Time-Based Prediction of Renewable Energy, and Comparison of Different Machine Learning Models for Checking Integrated Model Performance.

2.1 Renewable Energy Background

Prediction of the use of renewable energy sources has become a major area of research due to its growing importance in reducing global warming and guaranteeing energy security. As renewable energy sources—like solar and wind—are unpredictable and dependent on the weather, accurate forecasting is crucial to effective energy management and grid integration. A study by (Sharvini et al., 2018) has investigated the use of machine learning algorithms to forecast energy use. The author has stressed upon how important it is to consider a variety of weather factors to increase the accuracy of energy consumption forecasts. The study demonstrates how machine learning models can help with improved energy planning and management by identifying complex patterns in data related to renewable energy sources.

In the same way, Khalid et al., 2021) investigated the application of advanced regression approaches to forecast energy usage in relation to external variables like wind speed, humidity, and temperature using the Technology Acceptance Model (TAM). According to their findings, adding these variables significantly improves the models' ability to predict the future and provides a more thorough understanding of how weather patterns affect energy consumption patterns.

Furthermore, there is research conducted by (Kurbatova and Perederii, 2020) which focuses on the use of time series analysis in predicting the use of renewable energy. The study shows how well seasonal fluctuations and trends in consumption of energy can be captured by models such as SARIMA. The researchers were able to increase the accuracy of their forecasts by including time-based parameters, which gave grid operators and energy regulators new important data. The proposed study has used the Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) Model along with the SARIMA model to handle the issue of heteroskedasticity and potential autocorrelation. Also, research conducted by (Qazi et al., 2019) focuses on examining how well the different machine learning models forecast energy use. In comparison to traditional regression models, the authors have focused on the benefits of using ensemble techniques like Random Forest and Gradient Boosting to achieve improved accuracy. The study proposes the use of these modern techniques in energy forecasting, to effectively handle the issues raised by renewable energy sources. The proposed study involves use of both the Random Forest and Gradient Boosting ML algorithms along with the Polynomial Regression for comparison of the R-squared value to find the best fit amongst the 3 models for prediction of energy consumption.

Therefore, introduction of the literature section of the report gives a thorough explanation of the significance of forecasting the renewable energy consumption accurately, focusing on the use of time series analysis and machine learning to produce reliable results. Collectively, the mentioned studies demonstrate the progress made in weather variable integration and predictive modelling to improve the precision and resilience of energy consumption predictions.

2.2 Analysis of using Traditional ML Regression Methods for Prediction of Renewable Energy Consumption

As machine learning regression techniques use the previous data to identify the trends caused by a variety of external factors, they are essential for forecasting the consumption of renewable energy. According to the studies, these techniques can greatly improve energy forecast accuracy, which is crucial for maximizing energy production and consumption (Sharifzadeh et al., 2019). According to the literature, various machine learning techniques, like decision trees, neural networks, and polynomial regression, have shown a different level of efficiency in the domain of renewable energy consumption. A significant study has been conducted in which he examined the different machine learning models for forecasting solar energy generation. The researchers observed that polynomial regression models and support vector machines (SVM) showed excellent accuracy in capturing the nonlinear relationships present in solar energy data (Ahmad and Chen, 2018). However, the performance of these models can be heavily dependent on the quality and quantity of the training data, and the specific environmental variables considered, such as temperature and solar irradiance.

In another study, (Pasari et al., 2020) focused on the use of artificial neural networks (ANNs), which is a deep learning technique for wind energy production prediction. According to the research, ANNs are capable of accurately predicting the unpredictable nature of wind direction and speed, producing energy forecasts that are more accurate. But they need a significant amount of processing power and knowledge to modify their hyperparameters. There is research conducted by (Mohandes et al., 1998) wherein there is a comparative study of polynomial and linear regression. It was found that polynomial regression can capture more complex relationships, especially when the model degree is too high. Because of overfitting of the model, there is possibility of inaccurate predictions on untested data, which thereby highlights the need for robust model selection and validation. On the other hand, in the proposed study the polynomial regression is used along with the other machine learning algorithms (Random Forest and Gradient Boosting); and finally, the applied model is Cross – Validated to avoid model overfitting and improve the predictive performance of the model to study the renewable energy consumption patterns.

(Mohammed and Kora, 2023) has investigated the prediction of solar power using ensemble learning methods like Gradient Boosting and Random Forests. The research showed that, in comparison to single models, ensemble techniques could more accurately predict solar energy data due to their capacity to manage its unpredictable nature. The study showed that combining several weak learners to create a robust predictive model is beneficial for ensemble approaches, especially when handling noisy and high-dimensional data.

Therefore, altogether the above studies demonstrate the advantages and disadvantages of traditional machine learning regression techniques for forecasting the consumption of renewable energy sources. Although models with high accuracy, such as neural networks and polynomial regression, also need a great deal of validation and tuning know-how to prevent problems like overfitting. Including a variety of weather-related factors can improve model performance, indicating that an integrated approach for predicting energy use is advantageous, which is implemented in the proposed study by applying the VAR model which is a multivariate time series model used for studying the role of complementary renewable energy sources and thereby improve the model performance.

2.3 Study of Time Series Model for Time-Based Prediction of Renewable Energy

As time series models can consider the seasonality and changes present in energy data, they are an important technique for forecasting renewable energy use. For this reason, many studies have been done to understand the advantages and disadvantages of different time series models. Following is some literature related to the usage of time series model for time-based prediction of renewable energy. A study conducted by (Zhang et al., 2014) investigated the use of SARIMA model for wind energy prediction and discovered that these models could more accurately represent the seasonal patterns in wind energy data than simple ARIMA model. The researchers also identified two major limitations of the SARIMA model alone: its complexity in terms of parameter tuning and its inability to capture unexpected variations in wind speed. In the proposed study the SARIMA model is used with GARCH Model to handle the heteroskedasticity and capture the complexities of the hourly data which simple ARIMA model could not capture. In another paper, (Forecasting, n.d.) explained the application of Exponential Smoothing State Space (ETS) models for forecasting renewable energy. The researchers have focused on how flexible the

ETS models are for changing the seasonality and trends, thereby making them appropriate for a wide range of renewable energy data sources. They also observed that ETS models might not work well with data that shows unexpected changes or complex seasonal patterns, which would limit their use in some situations. It was examined by (Gers et al., 1999) the effectiveness of Long Short-Term Memory (LSTM) networks and recurrent neural networks (RNNs) for time series prediction of solar energy. They concluded that LSTM networks performed better at identifying long-term dependencies and non-linear patterns in the data than conventional RNNs and linear models. Even though LSTMs have a high accuracy rate, they are less suitable for real-time applications because to their high computational cost and need for significant data preprocessing. The application of the Vector AutoRegressive (VAR) model to forecast the interdependencies between various renewable energy sources, such solar and wind, was studied by (Contreras et al., 2003) They discovered that the VAR model offered a thorough understanding of energy production by successfully capturing the interactions between various time series. But the VAR model's computing requirements and intricacy were cited as drawbacks, especially for big datasets. In summary, time series models—each with distinct benefits and particular difficulties—are essential for forecasting the amount of renewable energy used. LSTM networks are superior at capturing non-linear dependencies, but SARIMA and ETS models perform well for linear and seasonal trends. Hybrid models integrate the best features of multiple techniques to improve prediction accuracy, whereas VAR models shed light on the relationships between various energy sources.

2.4 Comparison of Different ML models for checking Integrated Model Performance

(Babu and Reddy, 2014) demonstrated the combination of the time series model, ARIMA model and the neural networks to enhance the predictive accuracy for forecasting the consumption patterns of the renewable energy. This approach used the strengths of both statistical and machine learning models to improve the multivariate time series models. Then thereafter, (Jain et al., 2014) used the Deep Learning (DL) models for predicting the solar power generation, thereby focusing on the ability of both these models for capturing the complex patterns which are related to the weather conditions and seasonal patterns. The relevance of this research to the proposed study is that its application to understand the renewable energy consumption patterns. Also, (“Feature Selection for Predicting Building Energy Consumption Based on Statistical Learning Method,” n.d.) performed the Johanson co-integration, Larsson panel co-integration, and DH causality approaches to the time series and panel; which provided insights into the combination of the non-renewable and renewable energy sources and how much it depends on the external factors. Also, (“View of Renewable and Non-Renewable Energy and its Impact on Environmental Quality in South Asian Countries,” n.d.) developed a combined model of a machine learning model with LSTM (Long Short-Term Memory) networks, which is a deep learning algorithm for the prediction of wind energy. This study explained the efficiency of use of the LSTM model for time series analysis, which made it beneficial for prediction of the future consumption patterns. (Bansal et al., 2022) combined the deep learning models with machine learning models to forecast the energy consumption in smart grids. This demonstrated the combination of modern and traditional methods and improved predictive accuracy. In another study, (Sangwan and Herrmann, 2020) investigated the impact of both renewable and non-renewable energy sources on the environment by using panel data analysis. This research linked the energy consumption patterns to achieve sustainable goals. Lastly, (Mirjalili et al., 2023) used advanced regression models like random forest and support vector regression (SVR) for predicting solar and wind energy. The research showed the ability of machine learning models to handle the non-linearity in the data and study the integrated model performance. Table 2 gives an overview of some research papers having relevance to the proposed research.

Table 2: Review of Papers having relevance to proposed study

Author	Publication Year	Methodolog y Used	Relevance to proposed study
Babu and Reddy	2014	Combined ARIMA and neural networks for forecasting renewable energy consumption.	Used advanced hybrid methodsto improve prediction accuracy,relevant for enhancing multivariate time series models.
Jain et al. (2021)	2021	Used DL models for predicting solar power, focused on seasonalpatterns and weather dependencies.	Highlights effectiveness of DLin capturing complex dependencies, useful for studying renewable energy consumption patterns.
Zhao and Magoulès (2017)	2017	Applied Johanson co-integration,Larsson panel co-integration, andDH causality approach on time series and panel data.	Provides insights into co-integration of renewable and non-renewable energy sources, relevant for understanding it's dependency on external factors.
Ali, Anwar, and Nasreen (2020)	2020	Developed a robust ML model combining LSTM to predict windenergy.	Implemented application of LSTM in time series analysis, beneficial for predicting future consumption patterns.
Kulisz et al. (2019)	2019	Combined statistical models withDL for energy consumption forecasting in smart grids.	Shows the collaboration of traditional & modern techniquesin energy forecasting, helping inthe developing accurate predictive models.

Kuldip Singh Sangwan (2017)	2017	Investigated impact of renewable& non-renewable energy on environmental quality using panel data analysis.	Focuses on environmental impact of energy sources, thereby linking consumption patterns to broader sustainability goals.
Mirjalili et al. (2023)	2022	Used advanced regression modelsfor solar and wind energy prediction, including SVR and random forest.	Demonstrates the effectivenessof ML models in handling nonlinearities in energy data, useful for studying integrated model performance.

Hence, the literature review concludes by focusing on the critical role which forecasting plays in the consumption of renewable energy, considering the unpredictable nature of renewable energy sources like solar and wind. Although the modern models such as neural networks and ensemble techniques offer improved accuracy in capturing complicated, nonlinear relationships, traditional regression methods remain essential. Though the complex nature of the data might affect how well time series models work, they are especially good at explaining seasonality and temporal trends. Overall, the literature indicates that improving the accuracy of forecasts for renewable energy requires an integrated approach which involves different predictive models and weather-related variables.

3 Research Methodology and Data Pre-processing

3.1 Renewable Energy consumption prediction Methodology Approach

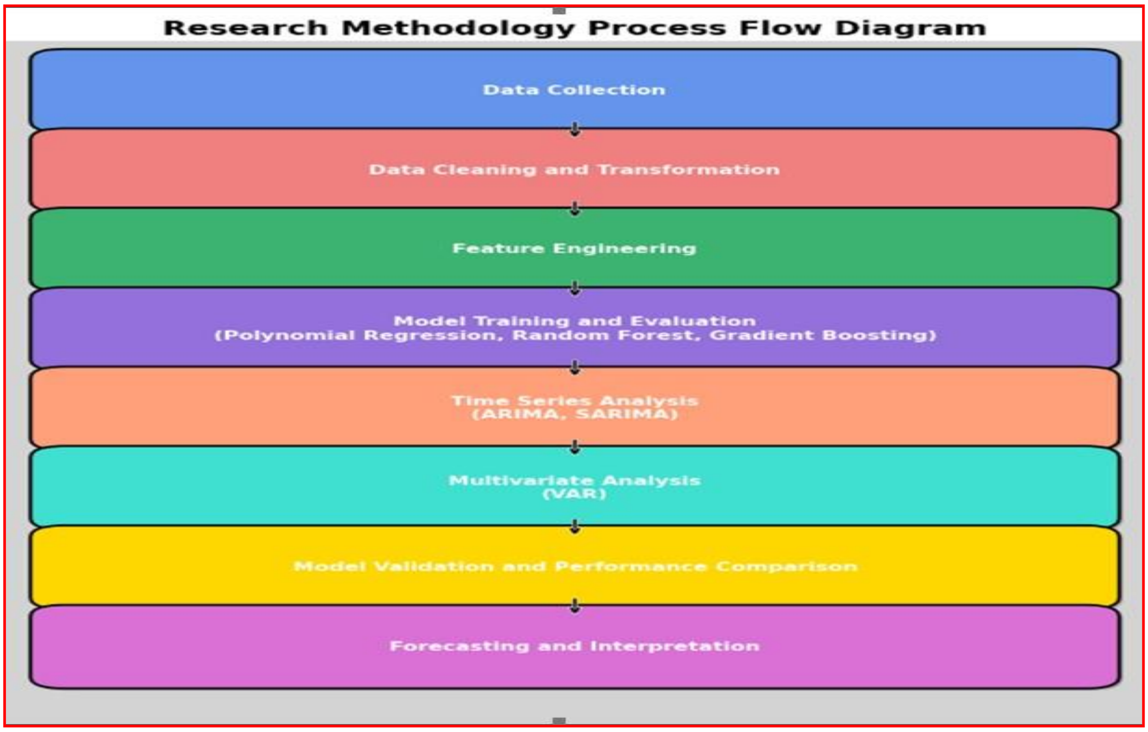


Figure 1: Methodology Approach used

The proposed study includes a set of actions which are intended to systematically answer the research topic of the consumption of renewable energy and how it is related to the weather conditions. Every step of the methodology followed follows the KDD that is- knowledge discovery in a database framework of data science. The first stage of methodology was collection of the dataset. The selected dataset was foundfrom a public dataset website- Kaggle. The selected dataset consists of columns like time, energy; it is comprised of a range of other external factors like temperature, humidity, pressure, wind speed, rain, sun, snow, etc. Then thereafter the collected data was then pre-processed and cleaned for further analysis. This process of the methodology included elimination of missing values and removing of the outliers for smooth evaluation. The collected dataset was standardized using data normalization techniques, allowing comparison between various machine learning models to find out the best fit for the data. These preprocessing procedures are necessary to maintain the data integrity and ensure that further analysis gives accurate and trustworthy results. Also feature engineering was performed on the collected data wherein selection of the important features and the target variables was done for improving model's predictive accuracy. The objective of performing the feature was to enhance the model performance by deriving significant patterns and correlations from the data. These characteristics were found and validated using a variety of statistical methods and visualizations.

In the next step of the methodology, several machine learning regression models, such as Polynomial Regression, Gradient Boosting, and Random Forest were used on the pre-processed dataset as training data. These three models were trained to forecast renewable energy consumption according to weather conditions like temperature, pressure, humidity, rain, and wind speed to know their impact on the renewable energy consumption respectively. The evaluation metric used to find out the best fit model amongst the three models was the R-squared value. The R-squared value is the value which calculates the percentage of the dependent variable's variation that can be predicted from the independent variables. Therefore,the model having high R-squared value will be the best fit model for predicting the energy consumption. The collected dataset was also divided into 80 and 20 training and testing data for robust evaluation process. Also, Cross-validation techniques were used to make sure the models performed adequately in the absence of observed data and to prevent overfitting. The results showed that the Random Forest model performed better than the other two, byaccurately predicting 89% of the variance in the data.

ARIMA and SARIMA models were used in time series analysis to get a clearer understanding of the temporal patterns of energy usage. These models were used to investigate the patterns of renewable energy consumption on a monthly, annual, and hourly basis. Different models were compared using a variety of evaluation criteria, including the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC). The ARIMA model was successful in determining how weather conditions affected annual and monthly energy consumption. But it was unable to fully capture the complexities of the hourly data.

To solve this, the GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) model and the SARIMA (Seasonal ARIMA) model were combined to produce more precise and reliable hourly energy consumption patterns. The last stage of the methodology was using the Vector Autoregressive (VAR) model, which is used for multivariate analysis. This model studied the impact of complementary renewable energy sources, namely solar and wind energy, by including extra weather variables like temperature, pressure, and humidity. The interdependencies between various energy sources and meteorological factors were well captured by the VAR model, enhancing the model's forecasting accuracy. The results of the model suggested that a more consistent and predictable renewable energy supply could be achieved by programmatically combining wind and solar energy, which would thereby help lessen the intermittency issues.

In conclusion, the overall research process used in this study followed the KDD principles. A thorough collection of data, data cleaning, feature engineering, polynomial regression along with Gradient Boosting and Random Forest analysis, time series models implementation and evaluation, and multivariate analysis by using the VAR model were all steps included in the process. To guarantee the authenticity and accuracy of the findings, each stage was carefully carried out, which finally resulted in a thorough understanding of the patterns of renewable energy use and how they are related to the weather conditions.

3.2 Data Pre-Preprocessing

In this step of the methodology involves all the steps performed for preprocessing of the data,required for the implementation purpose. So initially, the dataset was available publicly and itwas collected from the Kaggle website. The programming for implementation is done in Python by using the Jupyter Notebook. As explained in the methodology approach, various machine learning (Polynomial Regression, Gradient Boosting, and Random Forest) and time series (ARIMA and SARIMA) models are being used for the implementation purpose.

So, for each model, the collected data was programmatically loaded in the Jupyter Notebook. And before applying the respective model on the dataset, the collected dataset was loaded on the Jupyter Notebook by using appropriate functions in python. Then thereafter the collected data was then pre-processed and cleaned for further analysis. This process included elimination of missing values and removing of the outliers for smooth evaluation. The collected dataset was standardized using data normalization techniques, allowing comparison between various machine learning models to find out the best fit for the data. It was important to performall these preprocessing steps for the smooth implementation of all the selected machine learningand time series models respectively.

4 Design Specification

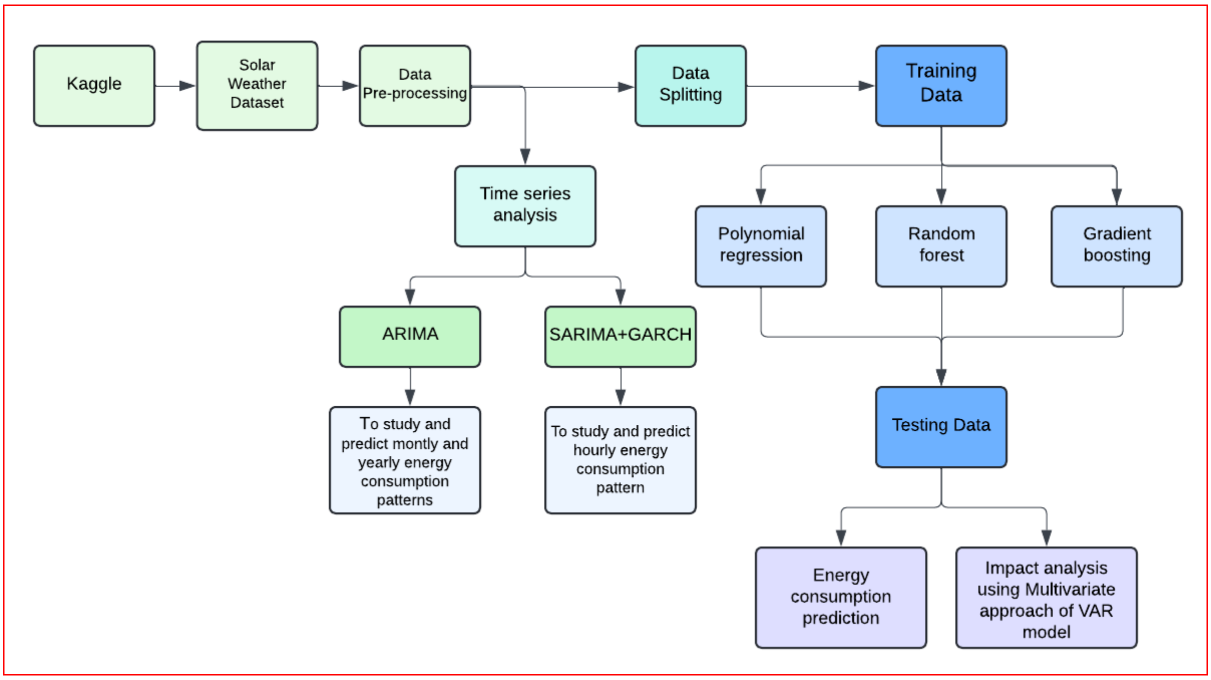


Figure 2: Design Specification

Figure 2 is the diagrammatic representation of the overall implementation of the proposed study. The diagram explains the flow of methodology performed. Firstly, the dataset named 'solar_weather.csv' was collected from Kaggle. Thereafter the collected dataset was pre-processed, which included data cleaning, eliminating the missing values, removing the outliers if present. Then the data was ready for implementation. Then in the next step, the selected machine learning algorithms (Polynomial Regression, Random Forest algorithm, Gradient Boosting Algorithm) and time series models (ARIMA & SARIMA + GARCH, VAR Model) were applied to find the best fit model and predict the renewable energy consumption.

During the implementation, the pre-processed data was split into training data and testing data. Then the polynomial regression model was applied to the dataset along with Random Forest and Gradient Boosting Models to check which model amongst the three models are best fit for prediction of the energy consumption. Evaluation metric used for prediction was the R-squared values. Also, cross validation was performed to avoid the over-fitting of the model. Therefore, after implementing all the three models, Random Forest Algorithm was found to bethe best fit for energy prediction with the highest R-squared value of 0.89.

Then the time series models, ARIMA & SARIMA were applied on the dataset. The ARIMA model was used to study and predict the monthly and yearly consumption patterns while SARIMA model was used along with GARCH model to examine and predict the hourly consumption patterns. The GARCH model was used to handle the problem of heteroskedasticity and improve the overall predictive performance of the model. The implementation of ARIMA model solved the research question partially and the implementation of the SARIMA model justified the research question completely.

Then finally the VAR model was applied to the testing data to study and analyse the impact ofcombining the complementary energy resources in the data, namely wind and solar energy, on renewable energy consumption. The evaluation metric used in this case is the Mean Squared Error (MSE) value. The results obtained showed that both, the sun and wind energy have low MSE Values, which indicates that the multivariate approach of the VAR model has improved the forecasting accuracy of the model by considering their collective impact and other weather conditions like temperature, humidity, pressure, rain, etc. The implementation ofthe VAR Model justified the sub-research question completely.

5 Implementation, Evaluation and Results of ML and Timeseries models used for forecasting Renewable Energy

In this section of the report, implementation of various machine learning regression models, and time series models used for the justification of the research question will be discussed. This section includes sub-sections like: Implementation of the Polynomial regression along with advanced models like Random Forest and Gradient Boosting, Implementation of the ARIMA and SARIMA time series models, and finally the Implementation of Vector Autoregressive (VAR) Model respectively.

Along with the explanation of implementation of the applied models, theevaluation and results of each model namely the polynomial regression, random forest, gradient boosting, ARIMA, SARIMA, and VAR Models will also be discussed in detail. Also, the summary of the results obtained from each model and to what extent the research or the sub- research question is justified will be explained at the end of implementation of individual model.

5.1 Implementation, Evaluation and Results of the Polynomial Regression, Random Forest and Gradient Boosting

The overall implementation of the polynomial regression was done by performing feature engineering for improving the model performance, cross-validation to avoid overfitting of the model. And finally including the advanced machine learning algorithms like Random Forest and Gradient Boosting for comparing the evaluation metrics used. The evaluation metric used in this case is the R-squared value. Along with the calculation of the evaluation metrics, various visuals have also been plotted for more detailed understanding of the prediction of the renewable energy consumption according to the weather conditions. Following are the values of the evaluation metrics of the selected 3 algorithms:

Table 3: R-squared values of selected 3 ML algorithms

Models used	R-squared values
Polynomial Regression	0.3313
Random Forest	0.8997
Gradient Boosting	0.3629

The table clearly shows that polynomial regression shows 33% variance in the data, which indicates its predictive capability is of moderate level. Same in case of the Gradient Boosting model, which shows 36% variance in the data. The Random Forest algorithm is significantly better, with a Test R2 score of 0.8997. This suggests that the model captures most of the variance (89%) in the data and provides accurate predictions of energy consumption based onthe weather conditions. Therefore, we can say that the Random Forest model successfully captures the complex and non-linear relationships between energy consumption and external factors of weather variables like temperature, humidity, pressure, wind speed, and rain respectively. Following are the visualizations (scatterplot) of the Actual vs Predicted Energy Consumption and Residuals of the Polynomial Regression, and Comparison of predictions of the two selected models- Random Forest and Gradient Boosting.

(i) Actual vs Predicted Energy Consumption of the Polynomial Regression: The following scatterplot explains that, since there is a wide spread of datapoints around zero that is, the line of perfect prediction, this means that the applied model has significant errors, and it is not so good in making predictions.

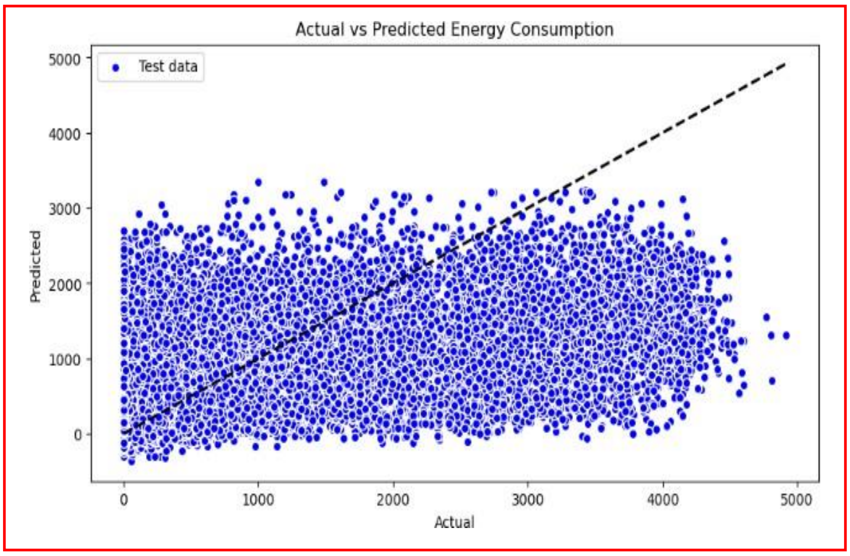


Figure 3: Scatterplot of Actual vs Predicted Energy Consumption

(ii) Residuals of Polynomial Regression: The following residual plot displays the heteroskedasticity and a pattern, which explains that this model might be missing some important information or trends in the data.

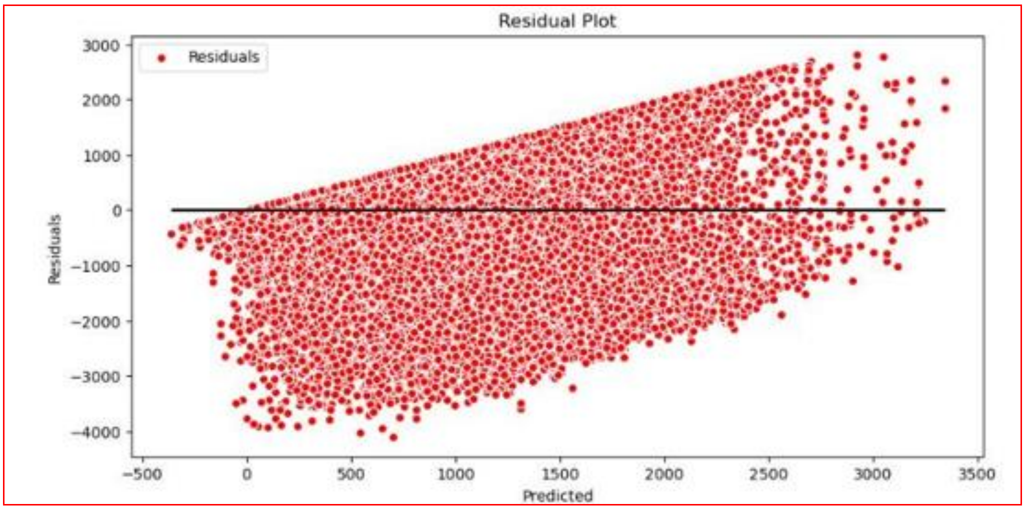


Figure 4: Residual Plot of Polynomial Regression

(iii) Comparison of Model Prediction: The following figure 5 is the scatterplot of the comparison of the selected advanced ML algorithms namely- Random Forest and Gradient Boosting Algorithms. The scatterplot clearly shows that the predictions made by the Random Forest Algorithm are much closer to the actual values as compared to that of the Gradient Boosting Algorithm respectively.

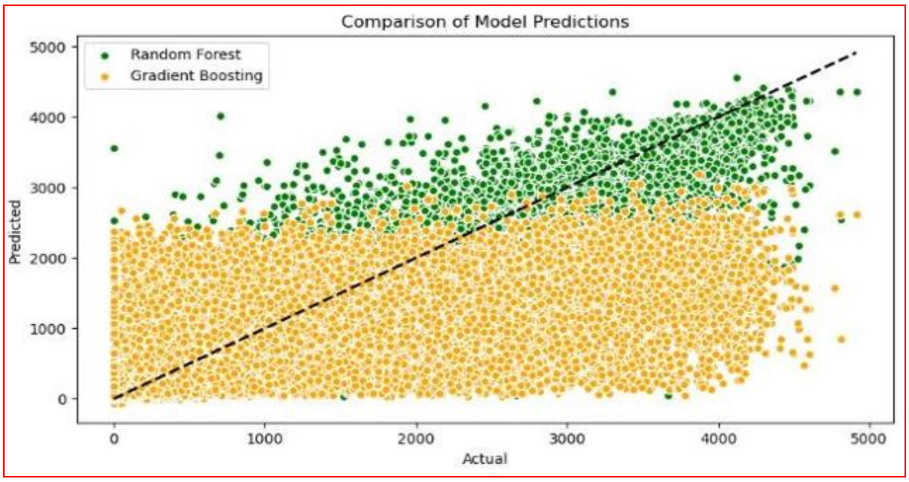


Figure 5: Comparison of Predictions of Random ForestandGradient Boosting algorithms

Therefore, in this part of the implementation of the Polynomial Regression we can conclude that, Random Forest algorithm is the best fit as compared to the other two algorithms (Polynomial regression and Gradient Boosting Algorithms), for the prediction of renewable energy patterns considering its dependencies on the weather conditions as it has the highest accuracy and predictive performance amongst the other tested models respectively.

5.2 Implementation, Evaluation and Results of ARIMA and SARIMA Models

The implementation of the ARIMA Model was done initially, to study the monthly, yearly and hourly consumption patterns of renewable energy. Various evaluation metrics and aspects like ADF Statistic to check the stationarity of the data, Akaike Information Criterion (AIC) & Bayesian Information Criterion (BIC) values for comparison with different models, Jarque- Bera (JB) and Ljung-Box (Q) for checking the normality of data and autocorrelation in the time series respectively, Mean Squared Error, along with some exogenous weather variables like temperature, pressure, humidity, wind speed, sun; were used for prediction of the consumption patterns of monthly, yearly, and hourly data respectively. Also, other evaluation metrics like root mean squared error (RMSE) and mean absolute error (MAE) are being calculated for evaluating the model which will be explained later in the Discussion and Conclusion sections of the report. Then, the plots of yearly, monthly, and hourly energy consumption, decomposition yearly, monthly, and hourly data; and the ARIMA forecast for yearly, monthly, and hourly energy consumption are being done to visually understand the results of the model. Table 4 is the tabular representation of the results of the ARIMA Model.

Table 4: Results of the ARIMA Model

Aspect	AIC	BIC	JB	Q	MSE	ADF Statistics
Monthly	1844	1864	2.78	3.41	113041002375.7	-7.501632
Yearly	-66.75	-70.26	0.24	4.40	1.89478062	-3.612292
Hourly	873241.4	873320.8	408244.86	191.43	48299456.9	-45.110877

From the table, it is clearly understandable that the ARIMA model can effectively identify andstate the impact of the weather conditions on renewable energy consumption only for the yearly and monthly data. The significant exogenous weather variables like temperature, pressure, humidity, etc. well explain the dependencies of the energy consumptions on the external factors. The seasonal variations in the collected data are well captured by the model in case of the monthly and yearly data by providing insights of how the renewable energy consumption patterns fluctuate with time. And because of the high values of AIC, BIC, JB, and MSE; the hourly model cannot capture all the complexities of the data, thereby indicating the probability of model overfitting.

Following is the visual representation of the hourly, monthly, and yearly consumption patterns of the renewable energy, decomposition yearly, monthly, and hourly data; and the ARIMA forecast for yearly, monthly, and hourly energy consumption:

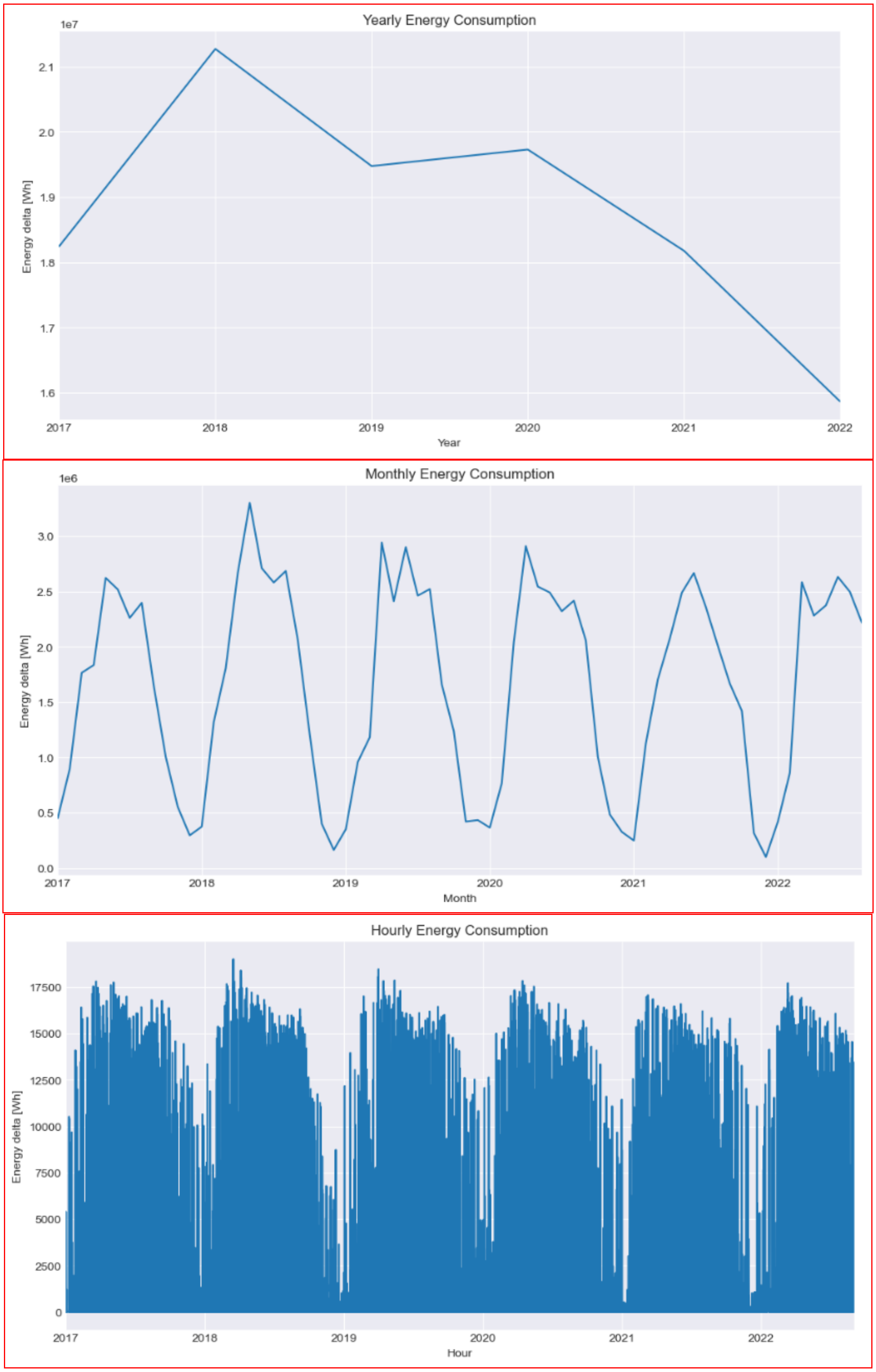
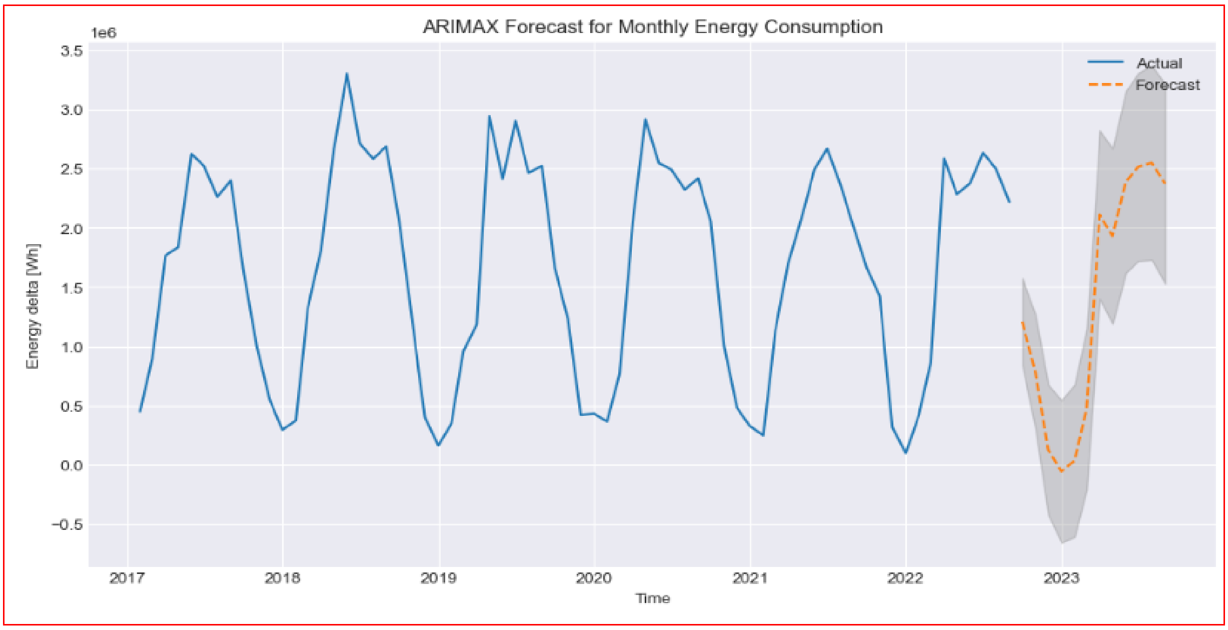
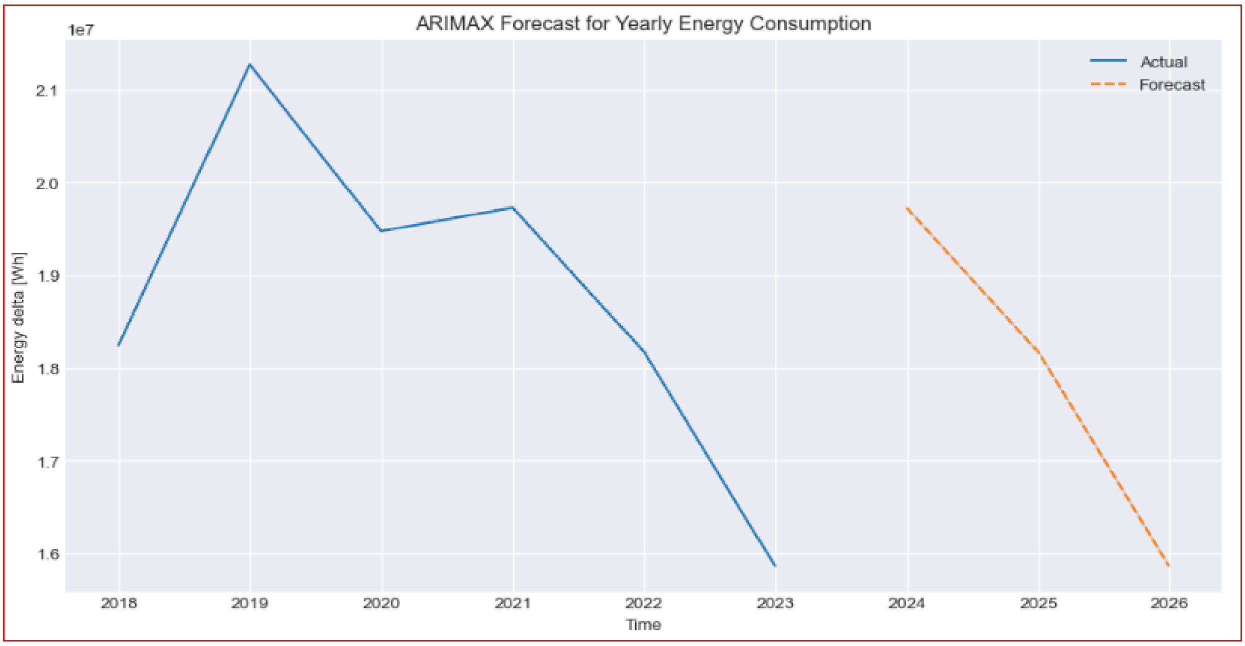


Figure 6: Hourly, Monthly, and Yearly energy consumption patterns before applying ARIMA Model

After the plotting of consumption patterns, the decomposition of Yearly, Monthly and Hourly data is being performed to improve the predictive performance of the model, and check if there are any outliers present in the data. Thereafter the ARIMA Model is applied to predict the energy consumption patterns yearly, monthly, and hourly respectively. Following are the visuals of model forecast.





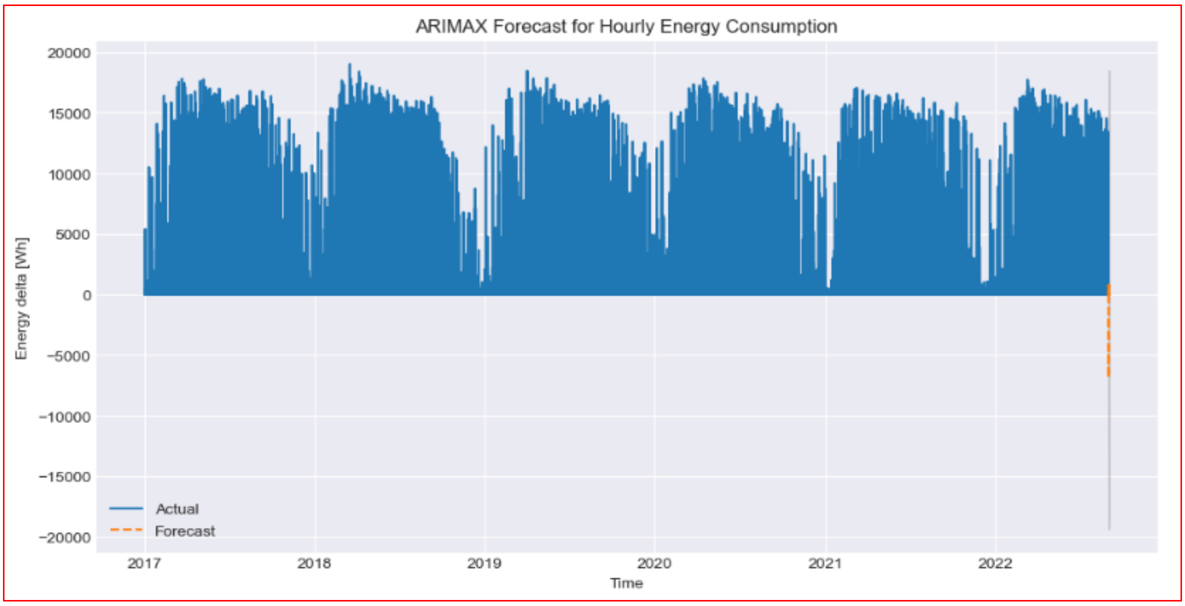


Figure 7: Model forecast for Monthly, Yearly and Hourly data after implementation of ARIMA

Therefore, from the above forecasts and values of the evaluation metrics, it is clear understood that implementation of ARIMA Model can only study the consumption patterns of the yearly and monthly data, but it could not predict the hourly consumption of renewable energy.

We can conclude that in case of the hourly model, the research question and the sub-research question 1 are partially justified, as it could just study the yearly and monthly energy consumption patterns, and not hourly consumption, as the sub-research question majorly contributes to the main research questions. Therefore, the Seasonal ARIMA (SARIMA) will be used to better capture the hourly consumption patterns. In the next part of the section, implementation of SARIMA model for analysing the hourly data will be explained in detail.

5.3 Implementation, Evaluation and Results of SARIMA Model and SARIMA+GARCH Hybrid Model

In this section the implementation of the Seasonal ARIMA and combination of both SARIMA & GARCH Models is done to analyse the hourly consumption of renewable energy and to capture the complexities in the hourly data. The SARIMA Model is used with the GARCH model too because the fitting of alone ARIMA model showed an issue of overfitting, residual heteroskedasticity, and autocorrelation. Also, cross-validation is performed to validate the applied model. Following are the results obtained from the implementation of SARIMA model along with GARCH model. And then the alone SARIMA Model was applied to the hourly data to check for the result by evaluating them using the evaluating metrics like MSE, RMSE, and MAE respectively. Table 5 shows the result of both the models.

Table 5: Results of SARIMA and SARIMA + GARCH model

Model	MSE	RMSE	MAE
SARIMA	1937482.8	1391.9	811.2
SARIMA + GARCH	2464655.2	1569.9	1202.8

Following is the forecast of the SARIMA Model for hourly energy consumption.

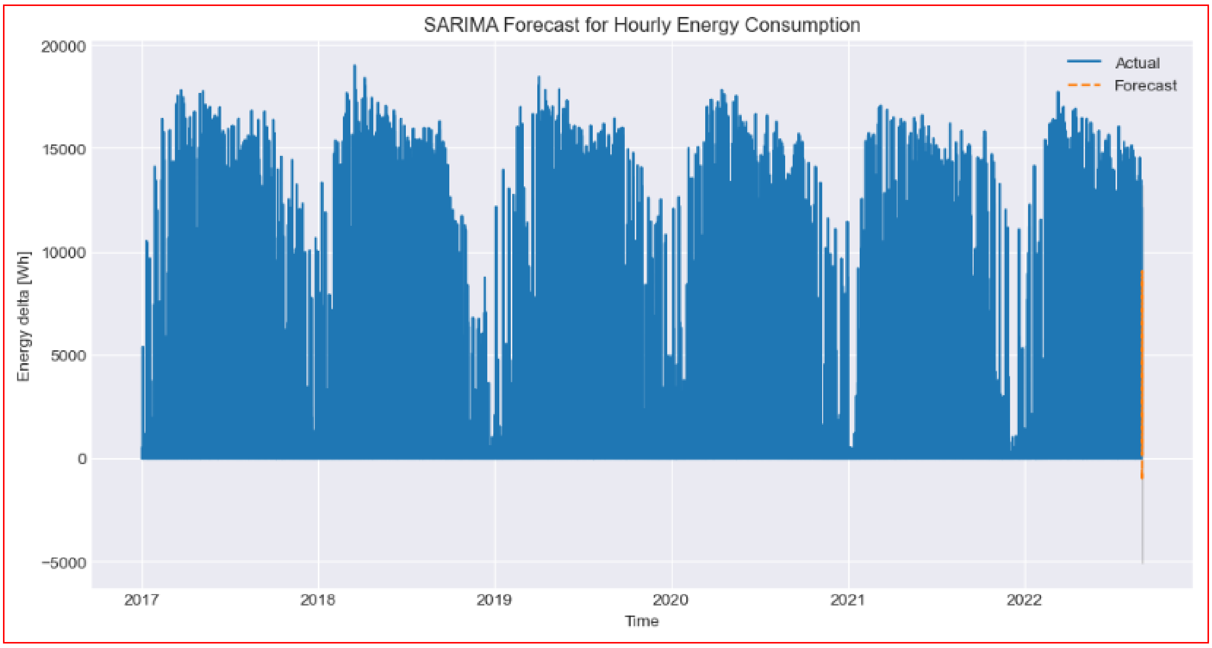


Figure 8: SARIMA for Hourly Energy Consumption

Therefore, according to the cross-validated MSE, the SARIMA model with GARCH gives more robust and accurate forecasts, as compared to a simple ARIMA model in terms of identifying complex patterns in the data. When it comes to considering the heteroskedasticity, SARIMA model performs better than a standard ARIMA model, especially when it is used with GARCH model; and is thereby able to capture the hourly energy consumption more effectively than the simple ARIMA Model. The explanatory power and accuracy of the model in describing hourly consumption patterns are significantly improved due to the addition of weather variables. Therefore, we can say that the research question and the sub-research question 1 is being justified completely by applying the ARIMA and SARIMA Model to study the trends and seasonal variations in the usage of the renewable energy, and how can time series analysis be used to predict future consumption patterns effectively. In the next part of this section, the results obtained from the implementation of the VAR Model,that is the Vector Autoregressive Model will be explained.

5.4 Implementation, Evaluation and Results of VAR model:

After the implementation of the SARIMA model, the VAR model that is the Vector Autoregressive Model is being applied by including the additional variables like temperature, pressure, humidity, etc. to study the role of complementary renewable energy sources and to improve the model performance, and thereby help in the justification of the sub-research question 2. Following are the results obtained from the implementation of VAR model:

Table 6: Results of VAR Model

Complementary Energy Source	Correlation of Residuals	Optimal Lag Order (AIC)	Optimal Lag Order (BIC)	MSE
Solar Energy	0.000223	15	8	0.020862
Wind Energy	0.067862	15	8	0.067171

From the table it is clearly understood that both the sun and wind energy have low MSE Values. Therefore, this indicates that the multivariate approach of the VAR model has improved the forecasting accuracy of the model by considering their collective impact and other weather factors. Also, the significant lags of the wind speed in the solar energy equation and solar energy in wind speed equation, indicates that both energy sources are complementary. There is low correlation in residuals which means that the model is sufficient to capture the interdependencies between the series, which will thereby improve the robustness of the forecasts respectively.

Figure 9 explains the forecasts of the Solar and Wind energy after the implementation of the VAR model respectively.

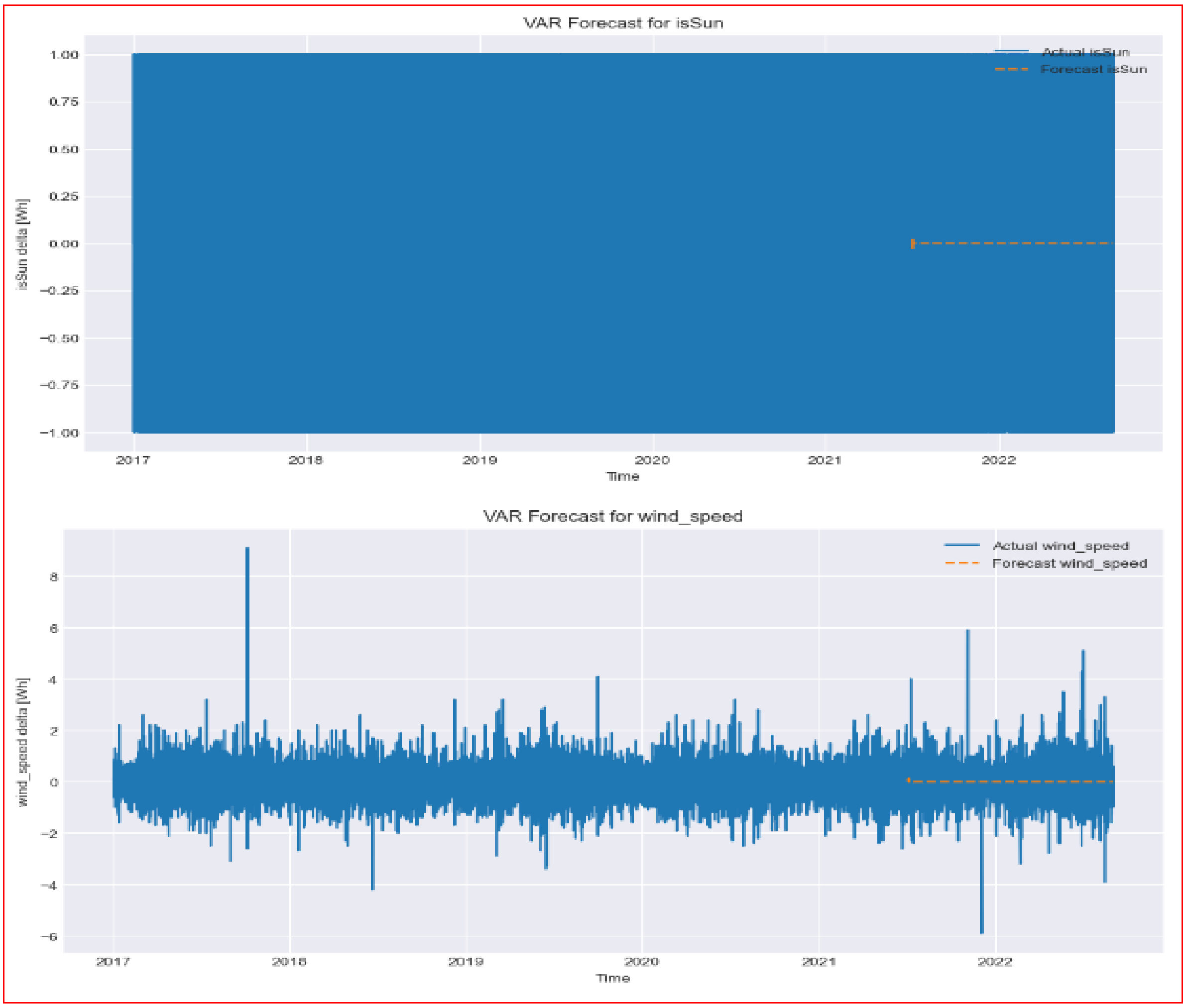


Figure 9: Solar and Wind energy forecasts after implementation of VAR model

The model has successfully captured all the interdependencies between solar and wind energy, along with weather conditions, which a single-variable model might miss capturing. Temperature, pressure, humidity, and rain have a significant impact on both solar and wind energy consumption patterns. Hence, here it can be concluded that if both the complementary renewable energy sources (solar and wind energy) are integrated programmatically, the model can provide more stable and predictable supply of renewable energy, which will thereby help in reducing the intermittency issues. Therefore, the research question, sub-research question 1 and sub-research question 2 are being justified to its fullest.

Discussion

The result of the proposed study provides significant insights into the patterns of the renewable energy consumption and effectiveness of various ML and time series models in capturing these patterns. Considering the hypothesis that the SARIMA + GARCH Model would perform the best as compared to alone SARIMA Model and that it would improve the forecasting accuracy, but the result showed exactly opposite- with SARIMA alone outperforming the SARIMA-

GARCH model. This suggests that the collected data may not show that level of volatility required for the implementation of the GARCH model, or a different GARCH variant could be more appropriate to use. Following Table 7 shows the comparison of the ARIMA, SARIMA, and SARIMA + GARCH models based on the evaluation metrics like MSE, RMSE, and MAE to better understand the stated limitation.

Table 7: Comparison of Time Series model's result

Evaluation Metrics	ARIMA	SARIMA	SARIMA+GARCH
MSE	48299456.9	1937482.8	2464655.2
RMSE	6949.7	1391.9	1569.9
MAE	6183.8	811.2	1202.8

Despite of this limitation in the implementation, the proposed study successfully highlighted the use of alone SARIMA model for capturing the seasonal variations in energy consumption; and the ability of the Random Forest Algorithm to effectively handle the non-linear relationships between the renewable energy consumption and weather variables. Then thereafter the implementation of the VAR Model to check its ability to understand the impact of integrating the complementary energy sources of solar and wind energy on the energy consumption to enhance the stability of the supply.

6 Conclusion

The proposed research studied and analysed the impact of weather conditions on renewable energy consumption by applying various machine learning regression models and time series models, thereby focusing on effectiveness of prediction of those models. Amongst the three ML algorithms namely, Polynomial regression, Random Forest algorithm, and Gradient Boosting; the Random Forest algorithm was found to be the best fit for prediction of energy consumption as it could capture 89 percent of variance in the data, and it could capture complex relationships between energy consumption and weather conditions. The ARIMA model could effectively capture the yearly and monthly energy consumption patterns but could not study the hourly consumption patterns properly because of its complexities. The SARIMA model could provide better results for forecasting the hourly consumption patterns and successfully captured the seasonal variations, but when it was combined with the GARCH model, it did not show any expected significant improvement in forecasting. The multivariate approach of VAR model experimented the importance of combining the complementary energy sources like wind and solar energy, which suggested that their integration can lead to a more stable and predictable supply of energy. Overall, the proposed study effectively experimented to justify the research and sub-research questions by providing valuable insights into energy consumption and use of predictive models.

Future Scope

The future research could focus on the improvement of the SARIMA + GARCH Model, by experimenting with different GARCH variants to better capture the volatility in hourly data. Also, by exploring hybrid models by combining different machine learning models with time series models could further improve the forecasting accuracy of the model for predicting the energy consumption. Additionally, if the dataset is expanded by including diverse locations and external factors like policy changes could improve the predictive performance and generalizability of the developed model. Another future work can include real-time forecasting models for practical applications in energy management, especially for combining renewable energy in the grid. Managing the intermittency issue using advanced energy storage solutions, could also further contribute to a more comprehensive energy management system.

Acknowledgement

I would like to thank my research project guide Dr. Catherine Mulwa for her guidance and motivation towards my research. I would like to thank my college, National College of Ireland for providing me with the relevant resources which played a major role in completing my research. I would like to extend my thanks to my colleagues for giving me guidance and support throughout my research. Finally, I would like to thank my parents for morally supporting me throughout the process.

References

Ahmad, T., Chen, H., 2018. Potential of three variant machine-learning models for forecasting district level medium-term and long-term energy demand in smart grid environment. Energy 160, 1008–1020. https://doi.org/10.1016/j.energy.2018.07.084

Khalid, B., Urbański, M., Kowalska-Sudyka, M., Wysłocka, E., Piontek, B., 2021. Evaluating Consumers' Adoption of Renewable Energy. Energies 14, 7138. https://doi.org/10.3390/en14217138

Kurbatova, T., Perederii, T., 2020. Global trends in renewable energy development, in: 2020 IEEE KhPI Week on Advanced Technology (KhPIWeek). Presented at the 2020 IEEE KhPI Week on Advanced Technology (KhPIWeek), pp. 260–263. https://doi.org/10.1109/KhPIWeek51551.2020.9250098

Mohammed, A., Kora, R., 2023. A comprehensive review on ensemble deep learning: Opportunities and challenges. Journal of King Saud University - Computer and Information Sciences 35, 757–774. https://doi.org/10.1016/j.jksuci.2023.01.014

Mohandes, M., Rehman, S., Halawani, T.O., 1998. Estimation of global solar radiation using artificial neural networks. Renewable Energy, 6th Arab International Solar Energy Conference: Bringing Solar Energy into the Daylight 14, 179–184. https://doi.org/10.1016/S0960-1481(98)00065-2

Pasari, S., Shah, A., Sirpurkar, U., 2020. Wind Energy Prediction Using Artificial Neural Networks, in: Sangwan, K.S., Herrmann, C. (Eds.), Enhancing Future Skills and Entrepreneurship. Springer International Publishing, Cham, pp. 101–107. https://doi.org/10.1007/978-3-030-44248-4_10

Qazi, A., Hussain, F., Rahim, N.ABD., Hardaker, G., Alghazzawi, D., Shaban, K., Haruna, K., 2019. Towards Sustainable Energy: A Systematic Review of Renewable Energy Sources, Technologies, and Public Opinions. IEEE Access 7, 63837–63851. https://doi.org/10.1109/ACCESS.2019.2906402

Sharifzadeh, M., Sikinioti-Lock, A., Shah, N., 2019. Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and Gaussian Process Regression. Renewable and Sustainable Energy Reviews 108, 513–538. https://doi.org/10.1016/j.rser.2019.03.040

Sharvini, S.R., Noor, Z.Z., Chong, C.S., Stringer, L.C., Yusuf, R.O., 2018. Energy consumption trends and their linkages with renewable energy policies in East and Southeast Asian countries: Challenges and opportunities. Sustainable Environment Research 28, 257–266. https://doi.org/10.1016/j.serj.2018.08.006

Babu, C.N., Reddy, B.E., 2014. A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. Applied Soft Computing 23, 27–38. https://doi.org/10.1016/j.asoc.2014.05.028

Bansal, M., Goyal, A., Choudhary, A., 2022. A comparative analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory algorithms in machine learning. Decision Analytics Journal 3, 100071. https://doi.org/10.1016/j.dajour.2022.100071

Contreras, J., Espinola, R., Nogales, F.J., Conejo, A.J., 2003. ARIMA models to predict next-day electricity prices. IEEE Transactions on Power Systems 18, 1014–1020. https://doi.org/10.1109/TPWRS.2002.804943

Feature Selection for Predicting Building Energy Consumption Based on Statistical Learning Method [WWW Document], n.d. URL https://journals.sagepub.com/doi/epdf/10.1260/1748-3018.6.1.59 (accessed 8.10.24).

Forecasting: Principles and Practice (2nd ed), n.d.

Gers, F.A., Schmidhuber, J., Cummins, F., 1999. Learning to forget: continual prediction with LSTM, in: 1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470). Presented at the 1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470), pp. 850–855 vol.2. https://doi.org/10.1049/cp:19991218

Jain, R., Smith, K., Culligan, P., Taylor, J., 2014. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. Applied Energy 123, 168–178. https://doi.org/10.1016/j.apenergy.2014.02.057

Mirjalili, M.A., Aslani, A., Zahedi, R., Soleimani, M., 2023. A comparative study of machine learning and deep learning methods for energy balance prediction in a hybrid building-renewable energy system. Sustainable Energy Research 10, 8. https://doi.org/10.1186/s40807-023-00078-9

Sangwan, K.S., Herrmann, C. (Eds.), 2020. Enhancing Future Skills and Entrepreneurship: 3rd Indo-German Conference on Sustainability in Engineering, Sustainable Production, Life Cycle Engineering and Management. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-44248-4

View of Renewable and Non-Renewable Energy and its Impact on Environmental Quality in South Asian Countries [WWW Document], n.d. URL http://ojs.fccollege.edu.pk/index.php/FJES/article/view/122/140 (accessed 8.10.24).

Zhang, Y., Wang, J., Wang, X., 2014. Review on probabilistic forecasting of wind power generation. Renewable and Sustainable Energy Reviews 32, 255–270. https://doi.org/10.1016/j.rser.2014.01.033