sustainability

MDPI

*Article*

# Study of Potential Impact of Wind Energy on Electricity Price Using Regression Techniques

**Neeraj Kumar [1], Madan Mohan Tripathi [2], Saket Gupta [3], Majed A. Alotaibi [4,*], Hasmat Malik [5,6,*] and Asyraf Afthanorhan [7]**

1 Electrical and Electronics Engineering Department, Bharati Vidyapeeth's College of Engineering, New Delhi 110063, India; neeraj.kumar@bharatividyapeeth.edu
2 Electrical Engineering Department, Delhi Technological University, New Delhi 110042, India; mmmtripathi@dce.ac.in
3 Instrumentation and Control Engineering Department, Bharati Vidyapeeth's College of Engineering, New Delhi 110063, India; sguptamits@gmail.com
4 Department of Electrical Engineering, College of Engineering, King Saud University, Riyadh 11421, Saudi Arabia
5 Department of Electrical Power Engineering, Faculty of Electrical Engineering, University Technology Malaysia (UTM), Johor Bahru 81310, Malaysia
6 Department of Electrical Engineering, Graphic Era (Deemed to be University), Dehradun 248002, India
7 Faculty of Business and Management, Universiti Sultan Zainal Abidin (UniSZA), Gong Badak, Kuala Terengganu 21300, Malaysia; asyrafafthanorhan@unisza.edu.my
* Correspondence: majedalotaibi@ksu.edu.sa (M.A.A.); hasmat.malik@gmail.com (H.M.)

**Abstract:** This paper seeks to investigate the impact analysis of wind energy on electricity prices in an integrated renewable energy market, using regression models. This is especially important as wind energy is hard to predict and its integration into electricity markets is still in an early stage. Price forecasting has been performed with consideration of wind energy generation to optimize energy portfolio investment and create an efficient energy-trading landscape. It provides an insight into future market trends which allow traders to price their products competitively and manage their risks within the volatile market. Through the analysis of an available dataset from the Austrian electricity market, it was found that the Decision Tree (DT) regression model performed better than the Linear Regression (LR), Random Forest (RF), and Least Absolute Shrinkage Selector Operator (LASSO) models. The accuracy of the model was evaluated using the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The MAE values considering wind energy generation and without wind energy generation for the Decision Tree model were found to be lowest (2.08 and 2.20, respectively) among all proposed models for the available dataset. The increasing deployment of wind energy in the European grid has led to a drop in prices and helped in achieving energy security and sustainability.

**Keywords:** price forecasting; renewable energy; grid integration; machine learning; decision tree; random forest; linear regression; LASSO; MAE; MAPE



**Citation:** Kumar, N.; Tripathi, M.M.; Gupta, S.; Alotaibi, M.A.; Malik, H.; Afthanorhan, A. Study of Potential Impact of Wind Energy on Electricity Price Using Regression Techniques. *Sustainability* **2023**, *15*, 14448. https://doi.org/10.3390/su151914448

Academic Editors: Marc A. Rosen, Saad Mekhilef, Ahmed Fathy, Abdullrahman Abdullah Al-Shamma'a and Hassan M. Hussein Farh

Received: 9 June 2023
Revised: 24 September 2023
Accepted: 26 September 2023
Published: 3 October 2023



## 1. Introduction

The surge in worldwide demand for electrical energy, accompanying the rapid proliferation of Renewable Energy (RE) sources (wind and solar power), has brought about an intensified focus on energy security and sustainability. It is causing a tense scenario with regard to fossil fuels. Even understanding this, excessive usage of these fuels will significantly increase environmental pollution. The single major source of $CO_2$ emissions is power generation. India, the second-most populated nation in the world, is a developing nation that uses more energy sources than any other nation. Focus should be placed on developing an alternative energy source to meet energy demand because fossil fuels are a key contributor to pollution and global warming. Due to their endless supply and environmental friendliness, now is the right time to lift the capacity of renewable energy. Moreover, carbon dioxide and other atmospheric pollutants are not produced by wind energy.

Some developing nations like the Indian and European electricity markets are currently more focused on renewable energy resources to meet the rising demand for power at affordable costs. Wind energy has gained popularity as an alternative to fossil fuels due to its characteristics such as renewability, natural cleanliness, affordability, and low environmental effects. Because of its safety and good environmental attributes, wind energy has developed from a marginal activity to a multibillion-dollar industry in India's power production sector. Although multiple wind energy power plants generate electricity in India's various geographical regions, analyzing their performance is a challenging task and a priority for all stakeholders. By incorporating RE sources into the grid, consumers benefit from a cost-saving advantage when there is a surplus of power, whilst incurring a higher price in times of deficiency [1]. However, the intermittent nature of wind and solar makes it difficult to optimally allocate and generate power in an RES-integrated environment. By 2018, the total capacity of wind energy (both onshore and offshore) worldwide had attained a remarkable 564 GW [2] and is anticipated to ascend to 2000 GW by 2030 [3]. Accordingly, to ensure an effective, dependable, and secure electricity system, there is a requirement to establish sophisticated technologies and solutions. A review of the literature reveals that wind energy penetration is leading to a decrease in electricity prices [4–7]. Price forecasting with wind energy penetration has become an increasingly important part of the energy sector due to the increasing prevalence of wind energy. Wind penetration in the US has grown from 5.6% in 2011 to 16% in 2020 and is projected to continue to increase going forward. As wind penetration rises, there is an increased need to forecast electricity prices given fluctuating supply and demand. Price forecasting with wind energy penetration can reduce the unpredictability of electric markets, thereby helping to ensure the efficient and cost-effective operation of energy markets [8]. Case studies from regions such as Europe have demonstrated how price setting is more successful with increased wind energy penetration. Thus, price forecasting with wind energy penetration is key to the efficient operation of energy markets.

EPF is an essential factor in power system planning, as it helps in predicting future electricity prices based on past data and other variables (load, forecasted load, and meteorological data). Its significance lies in providing power producers with the necessary data to bid optimally in the market, optimize demand-side management, and improve congestion management. The complexity of the task is increased by the non-stable and fluctuating demand. RE grid integration poses a complex challenge, particularly due to the intermittent character of wind energy. To tackle this issue, EPF (Electric Power Flexibility) technologies can be employed to modify energy utilization concerning pricing signals. This effect is especially salient in European countries, which have high penetration levels of wind energy, as the effects of wind energy on electricity prices are discussed in [9].

As wind penetration increases, the electricity price decreases; however, the price volatility for small time frames (e.g., 5 min) increases compared to longer intervals (e.g., one hour). If wind power is over-forecasted, electricity prices will rise, whereas under-forecasting wind power will result in a price decrease. The level of curtailing will vary depending on the level of wind penetration, with higher levels leading to a decrease in electricity price instability.

Wind energy penetration has reduced electricity prices due to zero marginal cost and fuel requirements during periods of high wind power. Consequently, fuel-based plants are forced to limit their production, allowing wind power to supplant them and further reduce electricity prices [10]. Various studies in the literature have highlighted the significant impact of wind energy on pricing, covering aspects such as price forecasting [11–14]. Moreover, an economic analysis exploring the price forecasting error in wind-integrated markets was conducted, as was a quantile regression method to determine the effect of wind and solar on electricity price variability for Germany [15].

The utilization of machine learning algorithms within the context of wind energy and its resultant implications remain yet to be completely investigated. Regression techniques have been regularly employed for price prediction with varying levels of accuracy depending on the specific electricity market data, as they are comparatively straightforward to execute [16,17]. Although empirical evidence attests to the executive precision and dependability of Artificial Neural Network (ANN)-based techniques, the challenge of successfully managing the nonlinear and erratic performance of prices is made much more intricate in the context of an energy market incorporating renewable sources, particularly given the intermittent character of the wind.

Electricity price forecasting is a difficult and multi-faceted task owing to the increasing integration of renewable energy sources, such as solar and wind, into the electricity grid. Forecasting the prices of electricity generated from these renewable sources is critical for profit optimization and risk management. This literature review examines the current state of electricity price forecasting under the effect of renewable energy sources such as solar and wind. In a study of electricity price forecasting, the effect of wind energy generation must be taken into account. Wind energy is a renewable energy source, and its availability depends on a variety of factors, such as seasonality, geographical location, and weather conditions. Accurately forecasting the availability of wind energy generation is essential for optimal electricity price forecasting. Furthermore, understanding the relationship between electricity prices and wind energy availability is crucial for effective electricity price forecasts. Machine learning models are less explored for price forecasting under the effect of wind energy penetration. Electricity price forecasting is an important tool in the renewable energy (solar and wind) environment. It helps to improve the integration of renewable energy sources into the existing electricity market by predicting future electricity prices. This information is used to help energy producers and consumers make informed decisions about their energy investments, such as when to buy or sell energy. As renewable energy sources become more prevalent in the electricity market, the complexity of electricity price forecasting increases significantly. A limited literature exists that explores price forecasting in correlation to renewable energy sources [18,19]. Consequently, the authors constructed a modern model to forecast prices in a wind energy market setting and assess the influence of wind energy penetration on prices. The detailed literature review is provided in Table 1 to support the literature survey of this study.

The majority of academics have made an effort to investigate price forecasting using various machine learning models, with results of varying degrees of accuracy for the fuel-based power market. Price forecasting in such an interactive grid scenario is vital due to fluctuation in the wind and solar energy output in the current scenario, since the grid operations, their scheduling, and their dynamics are changing. However, because its non-stationary and stochastic nature, reliable price forecasting is a difficult endeavor. The interaction between price forecasting of fuel-based plants and Distributed Energy (DE)-based plants makes this work both more difficult and crucial at the same time. Price forecasting has been a popular topic among researchers due to the elevated capacity, expansion, and global growth of renewable energy. This research attempts to close the price forecasting gap between fuel-based plants and Distributed Energy (DE)-based plants.

This paper present a comparison of four robust widely adopted machine learning-based regression techniques by the research community. We analyze these algorithms for the task of electricity price forecasting under the influence of wind energy penetration which is an integral aspect of managing power generation through renewable energy sources. However, most of the existing research in this domain, specifically price forecasting under the influence of renewable energy, is very limited. We experiment with different kernel combinations, loss functions, and model setups for each of the five models given in the article and report only the best-performing algorithmic setting for each model. We also conduct extensive hyperparameter tuning of the models for obtaining optimum results. We share the best hyperparametric configuration of the best performing model. We take these steps so that other researchers do not have to go through the tedious task themselves, thus aiding research reproducibility, community adoption, and accelerating research processes in this domain.

**Table 1.** A brief literature review of some relevant works.

| Ref No. | Authors Name | Title of the Paper | Year | Method Used | Remark |
|---|---|---|---|---|---|
| 1 | Olukunle O. Owolabi et al. [20] | Role of Variable Renewable Energy Penetration on Electricity Price and its Volatility across Independent System Operators in the United States | 2023 | Quantile Regression techniques | Merit order effect on price and linearity effect has been considered |
| 2 | Kumar, Neeraj and Tripathi, M.M. [21] | Investigation on Effect of Solar Energy Generation on Electricity Price Forecasting | 2022 | LSTM | Effect of solar energy penetration on electricity price has been investigated |
| 3. | Anna Maria Oosthuizen, Roula Inglesi-Lotz, George Alex Thopil [22] | The relationship between renewable energy and retail electricity prices: Panel evidence from OECD countries | 2022 | Empirical results were presented (panel unit test) | Investigation of wind energy penetration on electricity price for 34 OECD countries were conducted |
| 4 | Anbo Meng et al. [23] | Electricity price forecasting with high penetration of renewable energy using attention-based LSTM network trained by crisscross optimization | 2022 | LSTM | Empirical wavelet transform and crisscross optimization is used to decompose the data features and retrain the data |
| 5. | Haolin Yang, Kristen R. Schell [24] | Real-time electricity price forecasting of wind farms with deep neural network transfer learning and hybrid datasets | 2021 | DNN | GRU transfer learning concept is used for improving the forecasting accuracy |
| 6. | Elisa Trujillo-Baute, Pablo del Río, Pere Mir-Artigues [25] | Analysing the impact of renewable energy regulation on retail electricity prices | 2018 | Statistical analysis | The impact on retail electricity prices is positive and statistically significant, although relatively small |
| 7 | Talari, S. et al. [26] | Price Forecasting of Electricity Markets in the Presence of a High Penetration of Wind Power Generators. | 2017 | Bivariate ARIMA-Wavelet and RBFN | Large scale wind generator effects on electricity price have been considered |
| 8. | Cristina Ballester, Dolores Furió [27] | Effects of renewables on the stylized facts of electricity prices | 2015 | Statistical and empirical analysis has been presented | Statistically negative relationship between wind energy share and marginal price has been derived |
| 9 | Shcherbakova, A. et al. [28] | Effect of increased wind penetration on system prices in Korea's electricity markets | 2014 | Seasonal correlation between wind output and load | Statistical analysis on wind energy penetration on system marginal price has been performed |
| 10 | Blanca Moreno, Ana J. López, María Teresa García-Álvarez [29] | The electricity prices in the European Union. The role of renewable energies and regulatory electric market reforms | 2012 | Empirical analysis | Deployment of RES increases prices paid by consumers in a liberalized market |

The authors' main contribution in this research paper is to analyze the effect of wind energy generation on electricity prices using regression methods. The authors conducted a detailed analysis of the data to explore the relationship between the two variables. The dataset used in the study includes mainly historical electricity prices and wind energy generation data. The authors have used multiple regression models (LR, RF, LASSO, and DT) to examine the effect of wind energy generation on electricity price, with results showing a positive and statistically significant correlation. The authors have also provided valuable insights into the implications of the study.

Decision tree models are advantageous for price forecasting due to their ability to handle nonlinear relationships between variables, their ability to capture interactions between variables, and their accuracy in predicting outcomes. Decision tree models are also highly interpretable, allowing for easy visualization of the relationships between the various features and the target variable. This makes decision tree models an ideal choice for price forecasting, as the insight obtained from the model can be used to inform decisions regarding pricing strategies. Hyperparameter tuning has been performed by adjusting the values of the hyperparameters of a model to minimize the training error and maximize the model performance on a given dataset. This process helps to optimize the hyperparameters of a model and improve the accuracy of the model on unseen data. Hyperparameter tuning can help to reduce the overfitting of a model, improve the generalization of the model, and also improve the accuracy of the model on unseen data. From the predicted electricity price information, a system planner can schedule resources and maintain the reliability of the system effectively. Scheduling resources for demand and supply management given the intermittent nature of wind is a challenging task. By providing information about price signals, this investigation provides an invaluable aid to those in the wind power industry in their efforts to maximize the efficiency of their resources, allowing for the optimization of their demand schedules.

The key summary of the main contributions of this research work is as follows:

- To analyze the effect of wind energy generation on electricity prices using regression methods;
- A detailed analysis of the data to explore the relationship between the two variables;
- Use of real-time data on Austria's electricity market;
- Implementation of multiple regression models (LR, RF, LASSO, and DT) to validate the performance;
- Hyperparameter tuning work has been performed by adjusting the values of the hyperparameters of a model in order to minimize the training error and maximize the model performance on a given dataset.

This paper is divided into five sections. The introduction section provides an overview of the literature related to the field. The second section describes the data features and their statistical analysis. The third section outlines the models used to inspect the effect of wind energy on electricity prices, their mathematical modeling, and their advantages and disadvantages for a specific task. The fourth section shows the results and comparisons of the models used. MAPE, RMSE, and MAE were calculated for LR, RF, LASSO, and DT models, respectively. The fifth section concludes the literature review, models used, and results.

## 2. Data Analysis

This research undertakes an appraisal of the influence of wind energy penetration into the Austrian electricity market grid on pricing [30]. To do so, relevant information concerning day-ahead (DA) price, actual load, forecasted load, and wind energy generation was procured. As wind energy generation is intermittent, the data were normalized and regularized via a min-max method to optimize dataset redundancy. Statistical analysis of the available dataset was conducted to show the variability and non-linearity of price data and wind energy generation. After this, a statistical analysis of the price and wind generation data was conducted. As there were some missing values, a 10-month dataset from 2018 was utilized to train the model [31]. Data splitting was executed in a 70 and 30 (%) breakup to train and test the model. A scatter plot of wind energy generation at 15-min intervals is depicted in Figure 1 and the arithmetical investigation of the dataset is given in Table 2.

**Table 2.** Data Statistical Summary.

| Name of Parameter | Electricity Price | Wind Power |
|---|---|---|
| No. of samples | 26,393 | 26,393 |
| Time interval in the dataset | 15 min | 15 min |
| No. of missing data | 0 | 0 |
| The mean value of the data | 41.84024 | 658.2215 |
| Root mean square deviation | 19.13648 | 583.0336 |
| Minimum | −149.99 | 0 |
| Maximum | 977 | 2678 |
| Dissymmetry | 4.115555 | 1.052050 |
| Kurtosis | 225.2525 | 0.389035 |

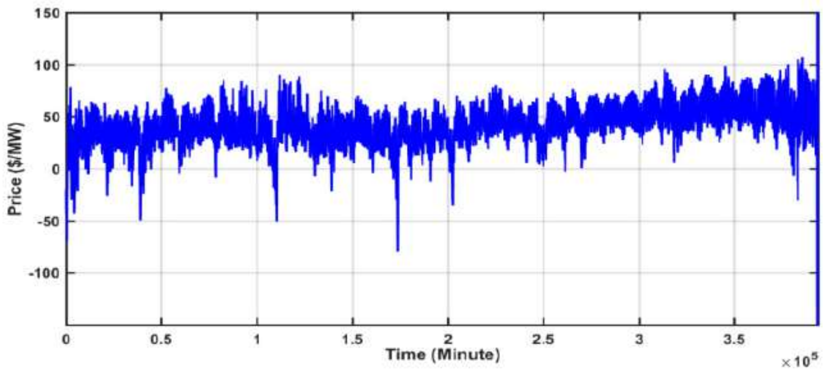


**Figure 1.** Variation in price with respect to time.

In Figure 1, the price versus time graph is shown, and the deviation of the price is attributable to fluctuations in demand. Notably, there are also periods of negative prices, which signify that power is abundant and there is reduced demand in the market. Since wind energy is intermittent, forecasting the price to deal with the price fluctuations and stochasticity is a challenge faced by researchers.

In Figure 2, the variation of wind energy output over time is depicted. To predict the price, two distinct regression techniques were modeled and compared, with wind production being taken into account as one of the limitations. An examination of the dataset utilized to train models through analytical means is presented in Table 2. This analysis is crucial for selecting the most appropriate model for price forecasting. It allows us to evaluate the dataset's symmetry and other features, thus helping us make a more informed decision.

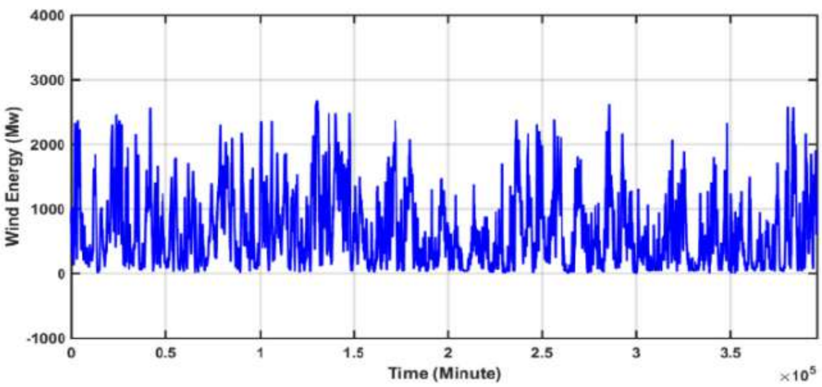


**Figure 2.** Variation of wind energy output with time.

## 3. Methodology

The methodology of this paper for examining the effect of wind energy on price forecasting is depicted in Figure 3. Descriptions of the various models employed are provided below. The proposed approach is comprised into seven steps, namely: (1) dataset collection, (2) dataset preparation (training and testing phase data file), (3) design of the regression model without wind energy, (4) design of the regression model with wind energy, (5) performance evaluation, (6) result comparison, (7) save model for further use.

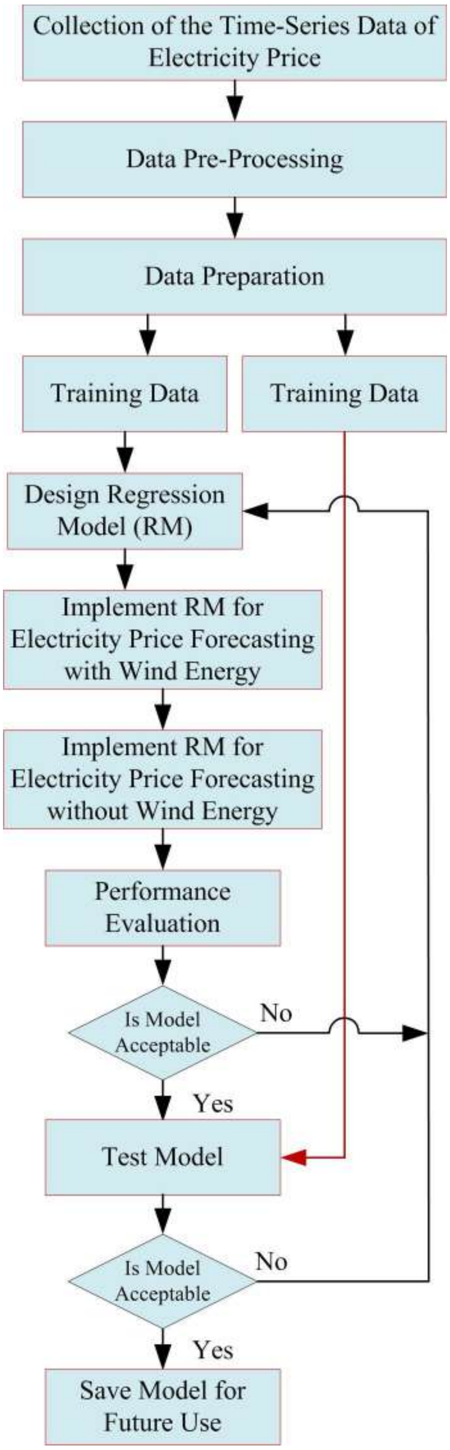


**Figure 3.** Flow diagram for methodology.

### 3.1. Decision Tree

DT is a supervised learning method employed to perform classification or regression problems by providing a set of features and a Boolean decision as output [32,33]. It creates a tree-like structure, breaking down the task into smaller subsets and progressively developing a decision tree. The resultant tree will comprise decision nodes, each signifying a particular attribute, and leaf nodes demonstrating the target value. The foremost decision node, representing the most optimal predictor, is known as the root node. The Iterative Dichotomiser 3 (ID3) algorithm of the Decision Tree is a powerful tool for supervised learning that utilizes a recursive partitioning methodology for constructing a decision tree from a dataset. This algorithm can generate a decision tree with ease and has been widely used in many real-world applications, such as data mining, knowledge discovery, and machine learning. At each step of the process, ID3 utilizes a heuristic measure to determine the best feature to split the data on, thereby reducing the complexity of the decision tree and allowing the user to better understand the underlying data. Standard deviation is utilized to assess the homogeneity of the dataset and will be zero if it is homogeneous. The decrease in standard deviation after segmenting the dataset is used to evaluate the maximum number of attributes that provide the highest standard deviation. In Table 3 specifications of the DT model are given.

**Table 3.** Details of hyper-tuned parameters for the Decision Tree model.

| The Objective of the Learning Task | Regression Model |
|---|---|
| Tuning linear model | Regression tree |
| No. of tree | 100 |
| Maximum depth of tree | 6 |
| Minimum sample at leaf node | 4 |
| Parameter tuning | Grid search with 5-Fold CV |

The Sum of Squared Error (SSE) equation is used in the context of regression with Decision Tree algorithms. It is used to calculate the error between the predicted values from the Decision Tree and the actual values of the target variable. The SSE equation is calculated by adding up the squares of the differences between the actual values and the predicted values for each data point in the dataset. The lower the SSE, the better the model is at predicting the target variable, following Equation (1). Therefore, the SSE equation is essential in determining the best decision tree model to use, as it can help us identify the model that will produce the best predictions.

$$S = \sum_{i=1}^{n} (Y_i - C)^2 \tag{1}$$

The S in Equation (1) is comprised of the number of observations (n) at the node, the mean outcome (C) which encompasses all the observations at the node, and the predicted value $Y_i$ of the i-th case.

Decision trees are also capable of generating robust and interpretable predictive models. Furthermore, these models allow for an easy implementation of feature engineering or selection, which can further increase forecasting accuracy. Finally, decision tree models usually require minimal data pre-processing and can work with both continuous and categorical data, making them ideal for wind energy-integrated markets [34]. The decision tree model has the advantages of interpretability, data preparation is less, it is non-parametric, versatile, and non-linear. At the same time, it suffers from data overfitting, data resampling, and feature reduction problems.

### 3.2. Random Forest

Random Forest (RF) is an ensemble learning method that is based on an estimator. To enhance the precision of the outcomes, an average of the entirety of the trees is taken and further refined. Bagging is a technique used in random forests to reduce overfitting. It works by taking a subset of the data, fitting a model to that subset, and then repeating the process with a different subset of the data. This creates multiple models which are then combined to form a single random forest model. Bagging is an important part of the Random Forest algorithm, as it is responsible for creating the individual decision trees that make up the forest. The bagging procedure allows for the trees to be independent of each other, and therefore, the forest is created with a wide variety of trees, thus increasing the accuracy of the model [35]. Bagging helps to reduce the variance of the model and increases the accuracy of the predictions. It is a powerful tool for improving the performance of random forest models. Random forest has numerous benefits; it does not experience the ill effects of overfitting even when the dataset is immense, and the random sample selection renders it a more effective predictor model. Table 4 furnishes the particulars of hyper-tuned parameters for the training and assessment of the RF model.

**Table 4.** Details of hyper-tuned parameters for Random Forest model.

| The Objective of the Learning Task | Regression Model |
|---|---|
| Tuning linear model | Random forest regression |
| Maximum depth of tree | 10 |
| Minimum sample at leaf node | 8 |
| Parameter tuning | Grid search with 5-Fold CV |

The algorithm works by randomly selecting a subset of the data, fitting a decision tree model to this subset, and then repeating this process multiple times. The final prediction is then the average of the predictions made by each decision tree. As a result, the Random Forest algorithm is an effective and efficient way to solve regression problems. Drawing from the initial dataset, n bootstrap samples are obtained, from which an unpruned tree is developed for each sample. Rather than selecting the optimal sample from the range of available predictors, an arbitrary selection of samples is taken from the predictors at every node, with the most appropriate split being selected from this selection. The Random Forest algorithm is an ensemble of decision tree classifiers that uses a bagging technique to create predictions. Each tree is created from a random sample of the data, and a prediction for a new data value is made by calculating the average of all the trees from the samples. This helps to reduce the variance from a single decision tree and improve the accuracy of the predictions. The Random Forest algorithm is an effective tool for both classification and regression tasks [36].

Random forests are an efficient function approximation tool that can be used to increase price forecasting accuracy by creating a collection of decision trees, each of which is trained on a different bootstrap sample of the original dataset and with different parameters. This allows random forests to improve variance and reduce bias as compared to basic decision tree models. Furthermore, by averaging out the predictions, random forests also reduce the risk of outliers in price forecasts and can capture a wider range of market dynamics. The RF algorithm outputs the relevance of features, which is a very valuable feature, and it is less prone to overfitting than DT and other algorithms. A small change in the data can cause the Random Forest method to change significantly, and its computations can become far more complicated than those of other algorithms.

### *3.3. Linear Regression*

Linear regression is a statistical algorithm used to predict a numerical outcome based on a dataset of input variables. It is one of the most widely used algorithms for solving regression problems. It works by finding a linear relationship between the input variables and the output variable, and then using this relationship to predict the output. It is a simple and powerful technique that can be used to solve a wide variety of problems.

$$Y = X \times \beta + \mathrm{r} \quad (2)$$

The Least Square Regression (LR) technique employs an approach to obviate the sum of squared discrepancies between the predicted and empirical values of Y, the response variable, by making use of X, the predictor matrix, and β, the relationship vector, in conjunction with r, the residual vector, as delineated in Equation (2) [37–39]. The mathematical representation of the LR model is depicted in Equation (3) and the particulars of the hyper-tuned parameters resulting from the training and validation stage of this model can be viewed in Table 5.

$$\mathrm{min}_{\beta}||\mathrm{X} \times \beta - \mathrm{Y}||^2 \quad (3)$$

**Table 5.** Details of hyper-tuned parameters for Linear Regression model.

| **The Objective of the Learning Task** | **Regression Model** |
|---|---|
| Tuning linear model | Linear Regression |
| Fit_intercept | true |
| Copy_X | true |
| normalize | false |

Linear regression is a good tool for price forecasting, as it is based on the assumption that property values change by a consistent amount over some time. Linear regression models can be easily transformed to allow for seasonality, wind direction, wind speed, or other impact factors on the electricity generated from wind power. It also allows for market responsiveness as well as for the incorporation of new energy sources into the trading portfolio.

Implementing linear regression is straightforward, and it is simpler to understand the output coefficients. However, because the bounds of the linear regression technique are linear, outliers can have a significant impact on the regression.

### *3.4. LASSO*

The LASSO (Least Absolute Shrinkage and Selection Operator) algorithm is a powerful technique for solving regression problems. It is a regularization method that penalizes the sum of the absolute values of the model coefficients. This helps to reduce the complexity of the model by shrinking some of the coefficients to zero and eliminating some of the features. The advantage of using the LASSO method is that the resulting models are usually simpler and easier to interpret [40]. Additionally, it can help to reduce the risk of overfitting by limiting the number of features used in the model. It is based on an optimization problem that has a convex objective function and a regularization term added to the cost function. This regularization term is a sum of the absolute values of the coefficients of the model, which ensures that the coefficients of unimportant features are set to zero and the important ones are retained. The equation for the LASSO model is given in Equation (4) [41].

$$£(\beta;\ \lambda) = \min\sum_{i=1}^{n}(Y_i - X_i\beta)^2 + \lambda_1\sum_{i=1}^{n}|\beta_k| \quad (4)$$

In Equation (4):
$Y_i$ = the result of the forecasted variable for the i-th instance;
$X_i\beta$ = remapped vector of forecasting where $X_i$ is the feature of the i-th variable;
B = assigned bias weight;
$\beta_k$ = the magnitude of the coefficient;
$\lambda_1$ = penalty factor.

The precision of the LASSO model for the current dataset is suboptimal owing to the insufficiency of features and an over-reliance on certain features, leading to overfitting. Table 6 refers to the specifications used in the training and validation of the LASSO model.

**Table 6.** Details of hyper-tuned parameters for the LASSO model.

| Objective for Learning Task | LASSO L1 Regularizer for the Linear Model |
|---|---|
| Tuning linear model | Lasso Regression |
| Fit_intercept | True |
| Copy_X | True |
| Alpha regularization parameter | $1\times10^{-15}$ |
| Normalize | True |

LASSO algorithms can also handle large-scale datasets and nonlinear functions which are commonly found in wind energy markets. Additionally, it can quickly adapt to changing input parameters which can be useful when forecasting prices in wind energy markets which are subject to sudden changes [42].

Automatic selection of features: a key benefit of the LASSO regression model is its ability to reduce the coefficients for features that are not interesting to zero. When we have correlated variables, the fundamental issue with LASSO regression is that it only keeps one variable and sets the other connected variables to zero. That might cause some information to be lost, which would impair our model's accuracy.

In Figure 4, the correlation plot for the features of the dataset is presented. It can be deduced that the target is highly correlated with the load and moderately correlated with wind generation. Hence the effect of wind energy on electricity prices is considered for the investigation of price forecasting in the wind energy integrated market. In the correlation plot, the correlation of data with target data is shown on the numerical values scale. One of the reasons for not running the LASSO model for the available dataset is the limited number of features and features that are not significantly correlated with the target value. Nevertheless, in a consolidated market where wind energy is a factor, the cost of electricity is impacted by wind energy. Hence, it is necessary to analyze the effect wind energy has on electricity prices. The correlation between the data and the goal data is displayed on a numerical values scale in a correlation plot. Because there are few features and they are not strongly correlated with the target value, the LASSO model cannot be run for the available dataset. However, wind energy affects the price of power in a consolidated market where it is a factor. Therefore, it is important to examine the impact wind energy has on the cost of power.

The tenuous relationship between load and day of the month does not conclusively indicate a significant correlation, yet it does suggest that electricity pricing is impacted by the presence of wind energy in an integrated market. The correlation coefficient between the two is not sufficient to validate a meaningful association, though it can be assumed that the price of electricity is to some degree contingent upon the availability of wind energy. Thus, it is imperative to assess the influence that wind energy has on the cost of electricity.

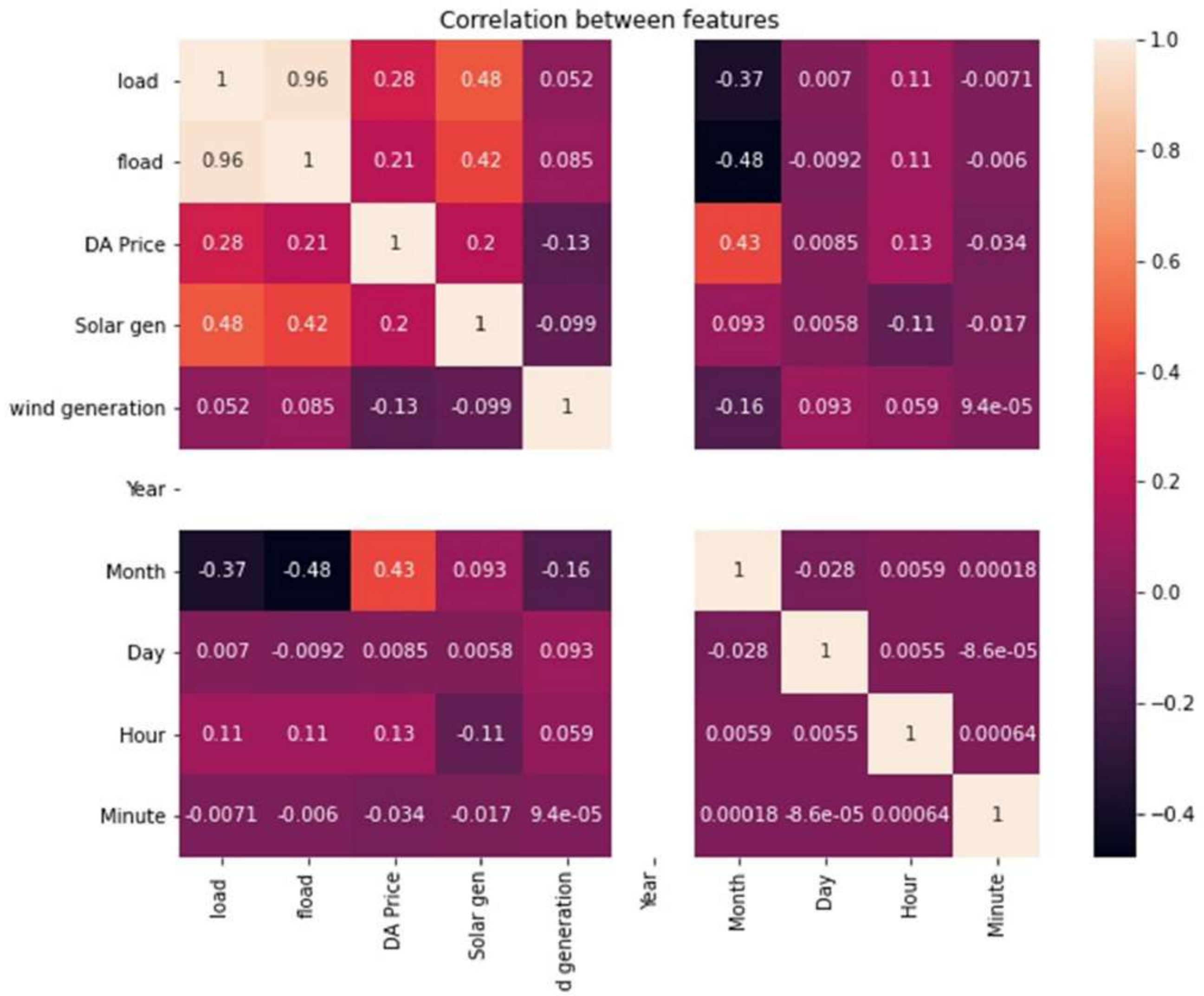


**Figure 4.** Correlation plot for the dataset used.

Although the shaky correlation between load and day of the month cannot prove a causal link, it does imply that the existence of wind energy in an integrated market affects electricity pricing. Though it can be expected that the price of power is to some extent dependent upon the availability of wind energy, the correlation coefficient between the two does not support a meaningful link. Therefore, it is crucial to evaluate the impact that wind energy has on the price of electricity.

## 4. Results and Discussion

Regression models were utilized to assess the effect on electricity price variation by imposing wind energy generation constraints in the market to meet the demand. Specifically, the DT, RF, LR, and LASSO models were run on numerical data related to Austria's electricity market utilizing a machine equipped with 6 GB, Double Data RAM3, a 1.6 GHz Intel Core i7 processor, and the Jupyter notebook development environment. Results of electricity price forecasting influenced by wind energy generation are detailed in the following section.

The same dataset was employed for training and evaluating models for electricity price prediction. Wind energy was excluded from the training and testing of the models, and the electricity price was then forecasted. After this, the same procedure was repeated while incorporating wind energy as an input parameter. Extensive hyperparameter tuning was conducted, and the highest performance values for this task have been reported. The efficacy of the model was gauged by utilizing forecasting metrics, and the numerical data were compared. MAPE, RMSE, and MAE are important metrics for evaluating the accuracy of forecasting models and the performance of a forecasting system. They are commonly used to gauge the accuracy of a forecasting model and provide a benchmark for comparing the performance of different forecasting systems. MAPE, RMSE, and MAE measure how closely the forecast values are to the actual values and provide an overall assessment of the accuracy of the system. This allows us to select the most accurate forecasting system, and identify areas for improvement. The metrics used to determine the models' effectiveness were the RMSE, MAE, and MAPE as expressed in Equations (5)–(7). Here, n was utilized to denote the total number of samples, and the actual and predicted prices for the jth instance were represented by $Y_{A,j}$ and $Y_{P,j}$ in Equations (5)–(7), respectively. Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are the metrics used to assess the performance of the models. In actuality, it is generally accepted that the better the model, the lower its RMSE number should be. The model is regarded as performing well without being overfitted if the RMSE values of the training and testing samples are within a narrow range. We think that the statistic of key focus should be RMSE, out of the two analyzed metrics. It uses a quadratic scoring rule, which squares the errors before averaging them. This enables RMSE to assign large mistakes a relatively high weight. Lower values of RMSE would therefore suggest a low error rate in practice.

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|Y_{A,j} - Y_{P,j}\right| \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(Y_{A,j} - Y_{P,j}\right)^{2}} \tag{6}$$

$$MAPE = \frac{100}{n}\sum_{j=1}^{n}\left|\frac{Y_{A,j} - Y_{P,j}}{Y_{A,j}}\right| \tag{7}$$

The results obtained from validating the models, namely the RF, LASSO, DT, and LR models, are illustrated in Figure 5 with three lines of distinct colors (green, blue, and orange). The DT model has shown slightly superior results in comparison with others (LR, RF, and LASSO) because it is close to the actual price line. Additionally, the MAPE for the DT model is lowest (5.80) with wind energy, and without wind energy error is minimal (6.01). These results are further corroborated by Table 7 which compares the evaluation metrics for the proposed regression models used in the analysis of the proposed task.

**Table 7.** Forecasting Matrices summary.

| Forecasting Matrices | Proposed Models | | | |
|---|---|---|---|---|
| | **DT** | **RF** | **LASSO** | **LR** |
| MAE without considering Wind | 2.08 | 4.81 | 8.89 | 10.82 |
| MAE considering wind | 2.20 | 3.50 | 10.82 | 12.01 |
| RMSE without considering wind | 6.19 | 5.76 | 15.51 | 15.69 |
| RMSE considering wind | 2.08 | 4.06 | 14.62 | 14.93 |
| MAPE without considering wind | 6.01 | 12.01 | 23.10 | 25.30 |
| MAPE considering wind | 5.80 | 10.91 | 22.20 | 24.50 |

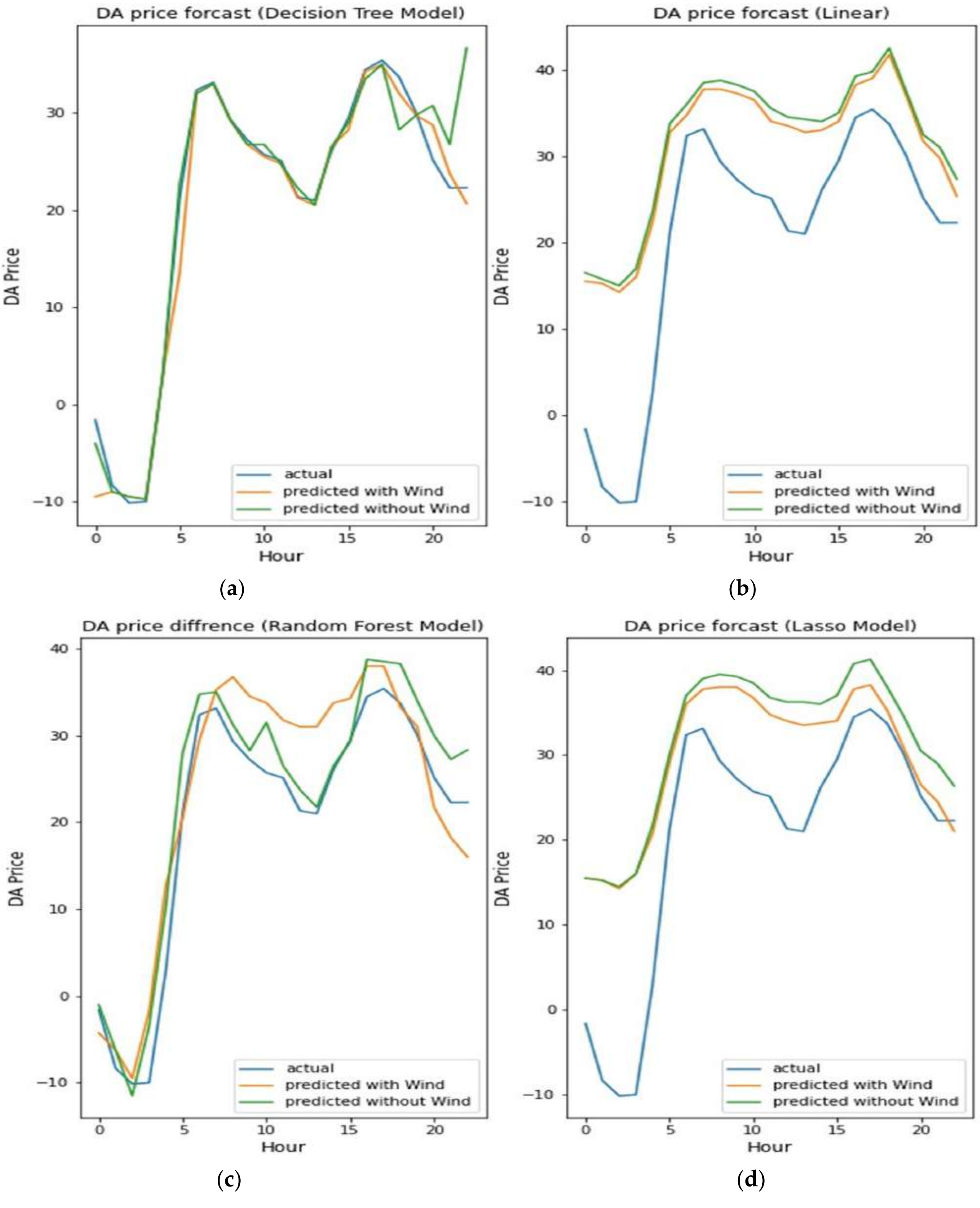


**Figure 5.** Forecasted and actual price with and without consideration of wind energy using (**a**) DT model, (**b**) LR model, (**c**) RF model, and (**d**) LASSO Model.

## 5. Conclusions and Future Scope

This study looked into aspects of the effects of wind energy penetration on electricity prices by proposing DT, RF, LR, and LASSO models. DT showed superiority over the other models due to its ability to capture complex mappings, variability, and complexity.

Analysis of the data revealed a reduction in electricity prices as an outcome of the introduction of wind energy sources, as well as an augmented degree of volatility due to the unpredictable nature of the energy source. This increases uncertainty for coal-based power producers. To better integrate wind energy into the grid, advanced forecasting is necessary for a reliable and effective power supply. Electricity price forecasting has great potential for the future, especially considering the increasing penetration of wind energy into the grid. Wind energy is highly variable and intermittent, making demand forecasting and resource planning more difficult. However, with the development of advanced forecasting algorithms, electricity price forecasting can be improved to better respond to the changes in wind energy production. In addition, with the integration of machine learning and big data analytics, electricity price forecasting can be used to anticipate the impact of renewable energy sources on the electricity market and develop strategies to minimize price volatility. This will enable utilities to better understand and manage the risks associated with renewable energy sources, and more efficiently plan for the future. Accurate prognostication of rates in renewable energy-integrated markets is a paramount worry due to the fluctuation in production. Hybrid techniques may enhance the precision of forecasting, which would be beneficial for operators in scheduling resources and exploiting RE capacity to the fullest.

**Author Contributions:** Conceptualization, N.K., M.M.T., S.G., M.A.A., H.M. and A.A.; Methodology, N.K., M.M.T., S.G., M.A.A., H.M. and A.A.; Software, N.K., S.G., M.A.A., H.M. and A.A.; Validation, N.K., S.G., M.A.A. and H.M.; Formal analysis, N.K., M.M.T., S.G., M.A.A. and H.M.; Investigation, N.K., S.G., M.A.A. and H.M.; Resources, N.K., S.G., M.A.A. and H.M.; Data curation, N.K., S.G. and M.A.A.; Writing—original draft, N.K. and H.M.; Writing—review & editing, M.M.T. and A.A.; Visualization, A.A.; Supervision, M.M.T. and A.A.; Project administration, M.M.T., M.A.A., H.M. and A.A.; Funding acquisition, M.A.A. and H.M. All authors contributed equally to the preparation of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** King Saud University, Riyadh, Saudi Arabia, project number RSP2023R278.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** King Saud University and MDPI Research Data Policies.

**Acknowledgments:** The authors extend their appreciation to the Researchers Supporting Project at King Saud University, Riyadh, Saudi Arabia, for funding this research work through the project number RSP2023R278. The authors extend their appreciation to the Intelligent Prognostic Private Limited India for providing the research facility.

**Conflicts of Interest:** We certify that there is no conflict of interest with any financial/research/academic organization regarding the content/research work discussed in the manuscript.

## References

1. Ziel, F. Modeling the impact of wind and solar power forecasting errors on intraday electricity prices. In Proceedings of the 2017 14th International Conference on the European Energy Market (EEM), Dresden, Germany, 6–9 June 2017; pp. 1–5. [CrossRef]
2. International Renewable Energy Agency. Available online: https://www.irena.org/wind (accessed on 1 January 2023).
3. Global wind Energy Council. Available online: https://www.worldenergy.org/ (accessed on 22 January 2023).
4. Zamani-Dehkordi, P.; Rakai, L.; Zareipour, H. Estimating the Price Impact of Proposed Wind Farms in Competitive Electricity Markets. *IEEE Trans. Sustain. Energy* **2017**, *8*, 291–303. [CrossRef]
5. Woo, C.; Zarnikau, J.; Kadish, J.; Horowitz, I.; Wang, J.; Olson, A. The Impact of Wind Generation on Wholesale Electricity Prices in the Hydro-Rich Pacific Northwest. *IEEE Trans. Power Syst.* **2013**, *28*, 4245–4253. [CrossRef]
6. Nuño, E.; Pereira, A.J.C.; Ferreira, C.M.M. Impact of variable renewable energy in the Iberian Electricity Market. In Proceedings of the 2015 50th International Universities Power Engineering Conference (UPEC), Stoke on Trent, UK, 1–4 September 2015; pp. 1–6. [CrossRef]
7. Gill, F.M. Impacts and best practices of large-scale wind power integration into electricity markets—Some Australian perspectives. In Proceedings of the 2012 IEEE Power and Energy Society General Meeting, San Diego, CA, USA, 22–26 July 2012; pp. 1–6. [CrossRef]
8. Kou, L.; Li, Y.; Zhang, F.; Gong, X.; Hu, Y.; Yuan, Q.; Ke, W. Review on Monitoring, Operation and Maintenance of Smart Offshore Wind Farms. *Sensors* **2022**, *22*, 2822. [CrossRef] [PubMed]

9. Carlo, B.M.-A.; Greg, B.; Bri-Mathias, H. The impact of wind power on electricity prices. *Renew. Energy* **2016**, *94*, 474–487. [CrossRef]
10. Available online: https://www.wind-energy-the-facts.org/the-impact-of-wind-power-on-the-power-market-dk-case.html (accessed on 25 December 2022).
11. Meneguzzo, F.; Ciriminna, R.; Pagliaro, M. The remarkable impact of renewable energy generation in Sicily onto electricity price formation in Italy. *Energy Sci. Eng.* **2016**, *4*, 194–204. [CrossRef]
12. Pereira, J.P.; Rodrigues, P.M.M. The impact of wind generation on the mean and volatility of electricity prices in Portugal. In Proceedings of the 2015 12th International Conference on the European Energy Market (EEM), Lisbon, Portugal, 19–22 May 2015; pp. 1–5. [CrossRef]
13. Green, R.; Vasilakos, N. Market behaviour with large amounts of intermittent generation. *Energy Policy* **2010**, *38*, 3211–3220. [CrossRef]
14. Masoumzadeh, A.; Nekouei, E.; Alpcan, T. Wind Versus Storage Allocation for Price Management in Wholesale Electricity Markets. *IEEE Trans. Sustain. Energy* **2020**, *11*, 817–827. [CrossRef]
15. Maciejowska, K. Assessing the impact of renewable energy sources on the electricity price level and variability—A quantile regression approach. *Energy Econ.* **2020**, *85*, 104532. [CrossRef]
16. Vu, D.H.; Muttaqi, K.M.; Agalgaonkar, A.P.; Bouzerdoum, A. Short-Term Forecasting of Electricity Spot Prices Containing Random Spikes Using a Time-Varying Autoregressive Model Combined With Kernel Regression. *IEEE Trans. Ind. Inform.* **2019**, *15*, 5378–5388. [CrossRef]
17. Saini, L.M.; Aggarwal, S.K.; Kumar, A. Parameter optimisation using genetic algorithm for support vector machine-based price-forecasting model in National electricity market. *IET Gener. Transm. Distrib.* **2010**, *4*, 36–49. [CrossRef]
18. Acaroğlu, H.; García Márquez, F.P. Comprehensive Review on Electricity Market Price and Load Forecasting Based on Wind Energy. *Energies* **2021**, *14*, 7473. [CrossRef]
19. Aineto, D.; Iranzo-Sánchez, J.; Lemus-Zúñiga, L.G.; Onaindia, E.; Urchueguía, J.F. On the Influence of Renewable Energy Sources in Electricity Price Forecasting in the Iberian Market. *Energies* **2019**, *12*, 2082. [CrossRef]
20. Owolabi, O.O.; Schafer, T.L.J.; Smits, G.E.; Sengupta, S.; Ryan, S.E.; Wang, L.; Matteson, D.S.; Sherman, M.G.; Sunter, D.A. Role of Variable Renewable Energy Penetration on Electricity Price and its Volatility across Independent System Operators in the United States. *Data Sci. Sci.* **2023**, *2*, 2158145. [CrossRef]
21. Kumar, N.; Tripathi, M.M. Investigation on Effect of Solar Energy Generation on Electricity Price Forecasting. *J. Intell. Fuzzy Syst.* **2022**, *42*, 1185–1197. [CrossRef]
22. Oosthuizen, A.M.; Inglesi-Lotz, R.; Thopil, G.A. The relationship between renewable energy and retail electricity prices: Panel evidence from OECD countries. *Energy* **2022**, *238 Pt B*, 121790. [CrossRef]
23. Meng, A.; Wang, P.; Zhai, G.; Zeng, C.; Chen, S.; Yang, X.; Yin, H. Electricity price forecasting with high penetration of renewable energy using attention-based LSTM network trained by crisscross optimization. *Energy* **2022**, *254 Pt A*, 124212. [CrossRef]
24. Yang, H.; Schell, K.R. Real-time electricity price forecasting of wind farms with deep neural network transfer learning and hybrid datasets. *Appl. Energy* **2021**, *299*, 117242. [CrossRef]
25. Trujillo-Baute, E.; del Río, P.; Mir-Artigues, P. Analysing the impact of renewable energy regulation on retail electricity prices. *Energy Policy* **2018**, *114*, 153–164. [CrossRef]
26. Talari, S.; Shafie-khah, M.; Osório, G.J.; Wang, F.; Heidari, A.; Catalão, J.P.S. Price Forecasting of Electricity Markets in the Presence of a High Penetration of Wind Power Generators. *Sustainability* **2017**, *9*, 2065. [CrossRef]
27. Ballester, C.; Furió, D. Effects of renewables on the stylized facts of electricity prices. *Renew. Sustain. Energy Rev.* **2015**, *52*, 1596–1609. [CrossRef]
28. Shcherbakova, A.; Kleit, A.; Blumsack, S.; Cho, J.; Lee, W. Effect of increased wind penetration on system prices in Korea's electricity markets. *Wind Energy* **2014**, *17*, 1469–1482. [CrossRef]
29. Moreno, B.; López, A.J.; García-Álvarez, M.T. The electricity prices in the European Union. The role of renewable energies and regulatory electric market reforms. *Energy* **2012**, *48*, 307–313. [CrossRef]
30. Höltinger, S.; Salak, B.; Schauppenlehner, T.; Scherhaufer, P.; Schmidt, J. Austria's wind energy potential—A participatory modeling approach to assess socio-political and market acceptance. *Energy Policy* **2016**, *98*, 49–61. [CrossRef]
31. Available online: https://data.open-power-system-data.org/time_series/2019-05-15 (accessed on 5 December 2022).
32. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
33. Yang, F. An Extended Idea about Decision Trees. In Proceedings of the 2019 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 5–7 December 2019; pp. 349–354. [CrossRef]
34. Alkawaz, A.N.; Abdellatif, A.; Kanesan, J.; Khairuddin, A.S.M.; Gheni, H.M. Day-Ahead Electricity Price Forecasting Based on Hybrid Regression Model. *IEEE Access* **2022**, *10*, 108021–108033. [CrossRef]
35. Wilson, B.; Dhas, J.P.M.; Sreedharan, R.M.; Krish, R.P. Ensemble learning-based classification on local patches from magnetic resonance images to detect iron depositions in the brain. *Int. J. Bio-Inspired Comput.* **2021**, *17*, 260–266. [CrossRef]
36. Panigrahi, S.S.; Mantri, J.K. A text based Decision Tree model for stock market forecasting. In Proceedings of the 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), Noida, Delhi, India, 8–10 October 2015; pp. 405–411. [CrossRef]
37. Yan, X.; Su, X.G. *Linear Regression Analysis*; World Scientific Publishing: Singapore, 2009.

38. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*, 5th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2012.
39. Sangani, D.; Erickson, K.; Hasan, M.A. Predicting Zillow Estimation Error Using Linear Regression and Gradient Boosting. In Proceedings of the 2017 IEEE 14th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Orlando, FL, USA, 22–25 October 2017; pp. 530–534. [CrossRef]
40. Arribas-Gil; Bertin, K.; Meza, C.; Rivoirard, V. Lasso-type Estimators for Semiparametric Nonlinear Mixed-Effects Models Estimation. *Stat. Comput.* **2014**, *24*, 443–460. [CrossRef]
41. Boulesteix, A.-L.; De Bin, R.; Jiang, X.; Fuchs, M. IPF-LASSO: Integrative-Penalized Regression with Penalty Factors for Prediction Based on Multi- Omics Data. *Comput. Math. Methods Med.* **2017**, *2017*, 7691937. [CrossRef]
42. Fragkioudaki, A.; Marinakis, A.; Cherkaoui, R. Forecasting price spikes in European day-ahead electricity markets using decision trees. In Proceedings of the 2015 12th International Conference on the European Energy Market (EEM), Lisbon, Portugal, 19–22 May 2015; pp. 1–5. [CrossRef]