International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Energy Pooling Market Forecating Using Machine Learning

Y. Shanmukha Manohara Reddy[1], B. Srilatha[2], Ch. Yaswanth[3], P. Dilip[4]

[1,2,3,4]Student, Raghu Engineering College

Abstract

In competitive electricity markets, accurate price forecasting is required to both power producers and consumers for planning their bidding strategies in order to maximize their own benefits. Price classification is an alternative approach to fore casting where the exact values of future prices are not mandatory. Presently, two efficient algorithms are proposed for both short term price forecasting (STPF) and classification (STPC) purposes. The algorithms include various methodologies like wavelet trans form (WT), fuzzy adaptive particle swarm optimization (FA-PSO) and feed forward neural networks (FFNN). WT is utilized to con vert the pathetic price series to an inviolable price series without losing the originality in the signal. Standard PSO (C) is implemented to tune the fixed architecture FFNN weights and biases. In the present nonlinear problem, linear variation of inertia weight does not resemble exact search process. Hence, dynamic inertia weight is accomplished by implementing the fuzzy systems in the GRADIENTBOOSTINGREGRESSOR approach. The hybrid methodology is implemented on Spanish electricity markets for the year 2002. To validate, three types of price classes and historical price series that are utilized by many researches as input features, are considered. Various statistical indicators are evaluated to compare and validate the proposed approaches with the past approaches available in the literature survey.

Index Terms: Forecasting, Fuzzy Systems, Neural Networks (NN), Particle Swarm Optimization (PSO), Wavelet Transform (WT).

INTRODUCTION

In the electric power sector, deregulatory policies are crucial, with the primary goal being to boost efficiency through competition. There are two main ways to trade electricity in this new environment: pool-based markets and bilateral contracts. Power providers offer the generating volumes and corresponding prices in pool markets, while consuming corporations give the demand requirements and corresponding prices. To clear the biddings, the independent system operator employs a market clearing algorithm based on single round disposals.[1]. If both the power producer and consumer businesses have a good forecast of a day ahead market clearing prices (MCPs) and volumes, they can build a plan to maximise their respective utilities utilising the electricity transacted in pool-based markets.[2].

Many researchers and academics are involved in the development of load and pricing forecasting systems. Recent literatures show significant advancements in load forecasting systems, with mean absolute percentage error (MAPE) less than 3%[3],[4],[5]. However, the price predicting algorithms that are being used are still in their early phases of maturation. Midterm price predictions (MTPF) and short-term price forecasting (STPF) are two forms of forecasts that are described as several days ahead prediction and a day ahead prediction, respectively.[6] Many researchers have devised numerous algorithms for MTPF.[7]For example, Xing et al.pro posed several models based on multiple support vector machine (SVM)[8]. (ARMAX)[9]and SVM and ARMAX. In order to effectively estimate the future MCP, the MTPF model must be designed with sufficient flexibility to handle the sampled data during the training phase. Over the years, various researchers have worked to address STPF-related concerns.

STPF focuses primarily on time series, econometric, and artificial intelligence (AI) models. Future price series are intrinsically nonlinear and may not be effectively captured by time series models due to their linear properties. [10]. In general, MTPF model must have well-built flexibility to handle the sampled data in the training phase in order to forecast the future MCP accurately. Over the years, several researchers have attempted the issues related STPF. In STPF, models based on time series, econometrics, and artificial intelligence (AI) are mainly focused. Future price series are inherently nonlinear and may not be captured accurately by the time series models due to their linear characteristics[11]. The market price growth and strategic behavior of the market participants have been studied in econometric based models. It is based on equilibrium and simulation analysis, which are based on economic theories such as classical Nash equilibrium .It is well suited for medium and long term time periods but may not be suitable for short term due to the exponential rising level of computational complexity[12].Support vector regression (SVR) and empirical mode decomposition (EMD) depend on input data mining that extracts the nonlinear data patterns. A step ahead price forecasting using feed forward neural network (FFNN) and autoregressive fractionally integrated moving average (ARFIMA) models was proposed in[13]. In general, AI techniques are classified into hard and soft computing techniques that are employed for predicting the STPF. The hard computing techniques like, WT-ARIMA models[14]. hybrid models based on WT, ARIMA and radial basis function neural networks (RBFN)[15]. integrated model based on WT, ARIMA, LSSVM and PSO[16]. and Nonparametric regression estimation methods are used to forecast future price series[17]. Computational costs for the hard computing techniques are very high and, also need more information regarding the past series, which are the main drawbacks[18].

Soft computing techniques include algorithms like neuralnet works (NN)[19] fuzzy neural networks (FNN)[20] hybrid intelligent method based on the WT and a hybrid of NN and fuzzy logic (WNF) technique and combination of WT, PSO and adaptive-network-based fuzzy inference system (ANFIS) (WPA) techniques[21]

The historical price series that are available in public domain are used for input–output mapping; thus, modelling of system is not necessary. Hence, if precise input data are considered, the soft computing techniques could perform more accurately than the hard computing techniques, which will overcome main drawbacks of hard computing techniques[11].

The historical prices and load values mostly influence the future electricity prices more than many other factors. How ever, the ANN models have received the largest share of attention compared to the other models. ANN models have shown improvement in forecast accuracy related to the other well-specified models. Feature preprocessing technique in fore casting model also influences the forecasting accuracy significantly. Especially, an ANN combined with pre-processed input feature data will achieve better prediction accuracy[22] Recent literature reveals that the combination of wavelets with ANN have resulted in good prediction results in various fields like high impedance faults detection[23]water quality prediction[24] detecting various winding faults in windmill generators[25]. Dynamic natured NN are frequently utilized in very large-scale integrated circuits. Also, very recent research shows the application $L\infty$ analysis for the single and inter connected NN with time varying delay [26]and[27]It is also apprehended from the past literature that the method ologies are developed to predict the precise value of electricity prices of the future hours using the present or past price series[28]

In many of the real world applications, the exact value of prices is not required but the class or level of the price is required. This converts the price forecasting problem to a classification problem in which the class of future prices of interest[22]. In many of the real world applications, the exact value of prices is not required but the class or level of the price is required. This converts the price forecasting problem to a classification problem in which the class of future prices of interest of electricity prices using FFNN ,generalized regression neuralnet works (GRNN), and cascade forward neural network (CFNN)[11].and discrete cosine transformed (DCT)-based NN[22]. Also, the stability analysis for the NN for both forecasting and classification purposes is proposed in[29] and [30].

In this, several required conditions are examined for both asymptotically and robustly stochastically stable systems. In light of this, this work offers a hybrid methodology consisting of WT, fuzzy adaptive particle swarm optimisation (FA-PSO), and FFNN for forecasting and classification in Spanish power markets. WT is used to penalise irregular input data to FFNNs, while the GRADIENT BOOSTING REGRESSOR method is used to tune FFNN weights and biases. Furthermore, the dynamic character of inertia weight is substituted in place of conventional inertia weight in the GRADIENT BOOSTING REGRESSOR algorithm to provide an exact nonlinear search. space, which is a significant new contribution to the subject. Furthermore, this document is structured as follows. Section II describes the mathematical modelling of several approaches used in price forecasting and classification. Section III outlines the procedures involved in implementing the proposed approaches, whereas Section IV describes the control settings for the proposed approaches. Section V illustrates the input data relevant to the current challenge. Section VI displays the results and the debates that accompany them

Methodology Gradient Boost Regressor Histogram-based gradient boosting classification machine (HGBCM)

HGBCM is an integration of a gradient boosting machine and a histogram-based algorithm for the construction of a classification tree structure. Gradient boosting refers to an ensemble of machine learning models relying on decision trees (classification trees for pattern categorization problems and regression trees for function approximation problems) as base learners[31][32][33][34] A subsequent tree in an ensemble is dependent on its predecessors (refer to Fig. (1) It is because the latter tree aims at reducing misclassified cases committed by the former. By this mechanism, the loss function which indicates the misclassification rate of the whole ensemble is gradually decreased during the model training phase. Notably, gradient boosting achieves a global convergence by following the direction of the negative gradient[35][33][36].After the training phase, a powerful committee based on relatively a weak base classifier can be formed[37]

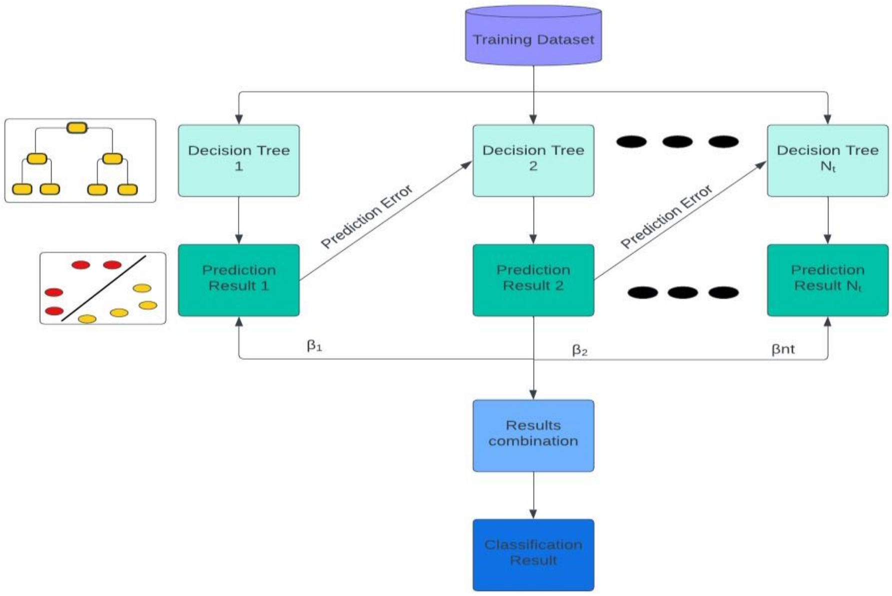
Finally, the prediction outcomes of novel data samples are obtained via a summation of the results computed from all individual decision trees.

Let $\{X; \, , \, t; \, \}e^{-1}$ represent the collected dataset. Categorical cross entropy LO is the loss function appropriate for multi-label classification. Let $h(x)$denote a base learner where $x;$ is the i^{th} data instance and $t;$ is the ground true label associated with $x;$.The initial prediction result of the model is given by[33]

$F_0(X) = argmin \, L^{Nv}{}_{i=0}{}^{=1} L(t; \, , J)$

where Nv denotes the number of data instances; $/J$ is the parameter of initial model weight.

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com



$CN-1$

$L = \sum t_i \log(y_i)$ (1)

$K-0$ where CN denotes the number of class labels; t_i, is the ground truth label. y, is the predicted output At iteration $m(m = 0,1, \ldots M$ with M being the maximum number of iterations), the gradient direction of classification error can be obtained as follows:

$$y_i^m = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$

(2)

A decision tree is employed to fit the data in the training set. Based on the least square algorithm, the parameter a of the model is attained and the base model h() is fitted as follows

$$a_m = argmin \sum_{i=0}^{N_D-1} [y_i^m - \beta \times h(x_i, a_m)]$$

(3)

The model weight parameter is updated as follows:

$$\beta_m = \operatorname{argmin} \sum_{i=0}^{N_D-1} L(y_i, F_{m-1}(x) + \beta h(x_j, a_m))$$ (4)

The whole committee is updated as follows:

$$F_m(x) = F_{m-1}(x) + \beta_m h(x_i, a_m)$$ (5)

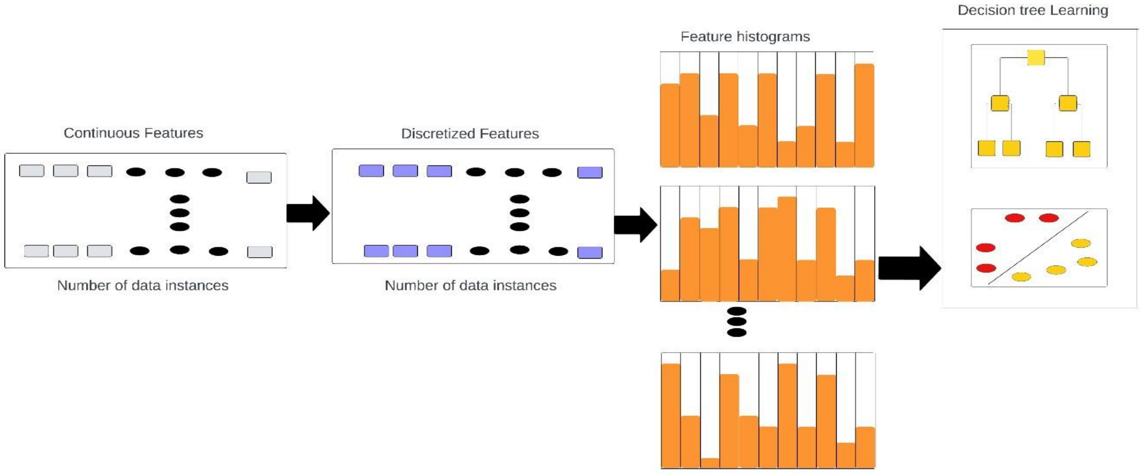
Additionally, the histogram-based algorithm is an effective method for gradient boosting based model training. With this algorithm, the ranges of continuous features used by decision trees are discretized into small bins. As suggested by Hossain and Deb [38], the number of bins can be set to be 255. These bins are subsequently employed to construct histograms representing the distribution of features values, Basic statistics, including the number of data instances and the sum of gradients in each bin, can be computed. Based on these indices, the optimal split points used for training the base learners can be determined. The histogram-based algorithm can significantly reduce the computational cost because the training phase of a decision tree does not require the scanning of the whole ranges of features for split point assessment[39]. Furthermore, the histogram-based algorithm can also help to achieve better generalization because the learning phase is less susceptible to noise[40]. Due to such advantages in both computation efficiency and learning performance, the histogram-based algorithm is used in this study for the construction of the classification tree structure.

Support Vector Machine

The SVM, which is a supervised machine learning algorithm was introduced by Vapnik in 1995 to solve classification and regression problems[41].The SVM is one of the supervised learning models that investigate data and identifies data samples used for classification. An SVM training algorithm constructs a model that labels a new data set into one class. Assuming that road accident prediction is a regression problem X, fatal accidents(x_1),severe accident(x_2),minor accidents(x_3) and property damage accidents (x_4)are considered as input (predictors) for SVM. This was done in order to understand how the response X depends simultaneously on the predictors in our context assumed to be(x_1), (x_2), (x_3)and(x_4). Given the training data consisting of the input matrix $X = [x_1,x_2, ... , x_n]$ and an output vector $Y = [y_1,y_2, ... , y_n]$the SVM construct an optimized linear regression through mapping the input vector x. SVM in machine learning approaches includes a set of learning methods and shows better results [42]and [43]. SVM performs linear and nonlinear classification. SVM supports linear and nonlinear regression applications [44].

Random Forests

The RF, first introduced in is also a powerful decision tree ensemble that can be used for various pattern classification tasks in civil engineering[45],[46]. The RF model employs a set of classification trees as base learners. This method relies on bagging (also called bootstrap aggregating) and random subspace sampling to construct a committee. Accordingly, the final class label of each data instance is determined via a majority vote[47]. Let $\{X, D$ denote a set of training data where $X = x_0,x_1,. . . x_{n-1}$and $T = t_0,t_1,. . . t_{n-1}$Let $h(x)$ represent a classification tree. For each individual tree $h(x)$, the model selects a random sample with replacement of the collected training data and uses such sampled data to train $h(x)$. This procedure aims at achieving better model performance because it is able to reduce the model variance without inflating the model bias. Besides sample bagging, RF also employs the mechanism of feature bagging. This means that a subset of features is used to train $h(x)$. This process is for reducing the correlation of the base learners in the whole committee. Usually, for pattern classification tasks, The number of _features selected by an individual $h(x)$ is $\sqrt{D}$where D is the total number of available features[37].



RESULTS AND DISCUSSIONS

Table-(1)

Various data sets used for training and testing the Indian electricity markets:

TRAINING	13/9/2023 TO 5/12/2023
TESTING	6/12/2023TO 12/12/2023

Price Forecasting for Indian Electricity Markets.

The proposed approaches are implemented on Indian electricity market to forecast the electricity prices for the year 2023.

The price forecasting results that are obtained by using deep learning and tabulated in table II. The results of interest are bold faced from the results, it is cleared that gradient boost regressor is good enough to predict the Indian electricity markets for the considered week.

Table-(2)

Forecasting results of Indian electricity markets using training on RMSE and testing on RMSE

Method	Training on RMSE	Testing ON RMSE
Baseline	0.08	0.07
Linear Regression	0.26	0.23
Lasso	0.08	0.07
Ridge	0.26	0.23
Elastic Net	0.03	0.07
Random Forest Regressor	0.05	0.07
Gradient Boosting Regressor	0.01	0.07
Extra Trees Regressor	0.08	0.08
Linear Svr	0.08	0.08

The RMSE ranges between 0.08 and 0.26 which show a middling prediction accuracy of the RMSE approach. It is observed that the results obtained using Deep learning are more accurate with each other than the previous references with least errors seen in the graph 1.

Root mean square error (RMSE) is a standard way to measure the error of a model in predicting quantitative data.

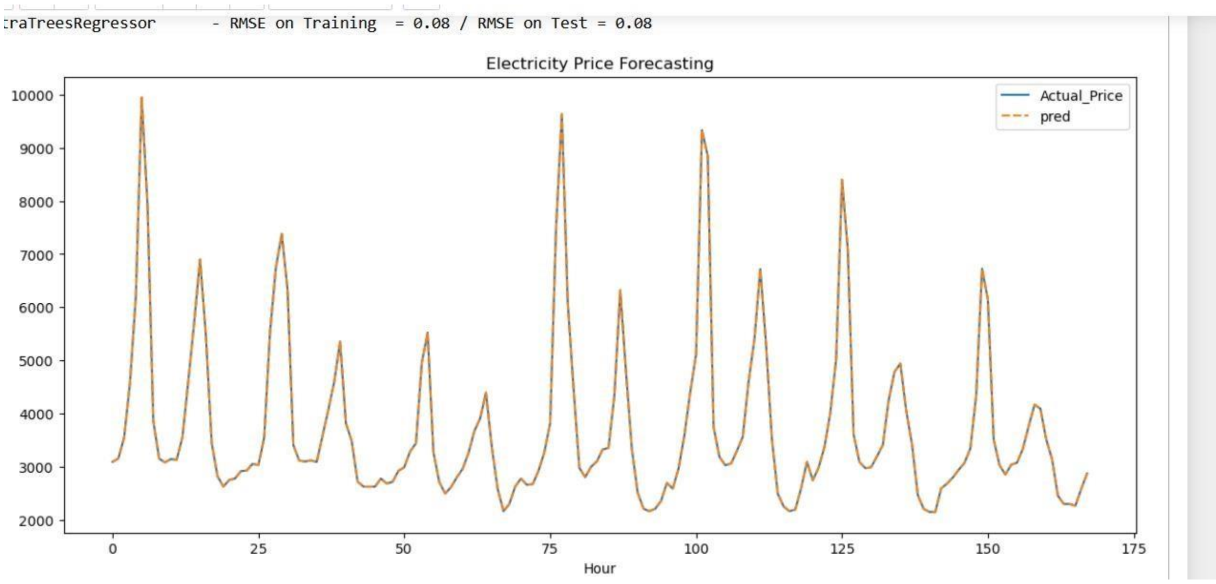
However to achieve more refined result with greater accuracy, gradient boost regressor method is used in the deep learning which helps us to decrease the bias error.

Gradient Boosting for regression. This estimator builds an additive model in a forward stage-wise fashion, it allows for the optimization of arbitrary differentiable loss functions. In each stage a regression tree is fit on the negative gradient of the given loss function.

Here on the table the training time has 84 days with baseline of 0.08 RMSE and linear regression of 0.26 RMSE and gradient boost regression of 0.01 RMSE and linear Svr of 0.08 RMSE and some more values less than “1”.

After that on the testing time has 7 days and the test data RMSE’s are with baseline of 0.07 RMSE, linear regression of 0.23 RMSE and gradient boost regression of 0.07 RMSE and linear Svr of 0.08 RMSE and the values are nearly equal to the 84 days of training time.

The RSME obtained for 84 days (i.e:2,016 hours) is equal to the RSME of testing time of 7 days (i.e;168 hours) where the errors are less than one so the deep learning method is considered as the best method to forecast the market pooling than FFNN and MAPE.



REFERENCES

1. M. Prabavathi and R. Gnanadass, "Energy bidding strategies for restructured electricity market," *International Journal of Electrical Power & Energy Systems*, vol. 64, pp. 956–966, Jan. 2015, doi: 10.1016/J.IJEPES.2014.08.018.
2. F. J. Nogales, J. Contreras, A. J. Conejo, and R. Espínola, "Forecasting Next-Day Electricity Prices by Time Series Models," 2002.
3. C. Unsihuay-Vila, A. C. Zambroni de Souza, J. W. Marangon-Lima, and P. P. Balestrassi, "Electricity demand and spot price forecasting using evolutionary computation combined with chaotic nonlinear dynamic model," *International Journal of Electrical Power and Energy Systems*, vol. 32, no. 2, pp. 108–116, Feb. 2010, doi: 10.1016/j.ijepes.2009.06.018.
4. R. A. Hooshmand, H. Amooshahi, and M. Parastegari, "A hybrid intelligent algorithm based short-term load forecasting approach," *International Journal of Electrical Power and Energy Systems*, vol. 45, no. 1, pp. 313–324, Feb. 2013, doi: 10.1016/j.ijepes.2012.09.002.
5. S. Kelo and S. Dudul, "A wavelet Elman neural network for short-term electrical load prediction under the influence of temperature," *International Journal of Electrical Power and Energy Systems*, vol. 43, no. 1, pp. 1063–1071, Dec. 2012, doi: 10.1016/j.ijepes.2012.06.009.
6. S. K. Aggarwal, L. M. Saini, and A. Kumar, "Electricity price forecasting in deregulated markets: A review and evaluation," *International Journal of Electrical Power and Energy Systems*, vol. 31, no. 1, pp. 13–22, Jan. 2009, doi: 10.1016/j.ijepes.2008.09.003.
7. P. Kou, D. Liang, L. Gao, and J. Lou, "Probabilistic electricity price forecasting with variational heteroscedastic Gaussian process and active learning," *Energy Convers Manag*, vol. 89, pp. 298–308, Jan. 2015, doi: 10.1016/J.ENCONMAN.2014.10.003.
8. X. Yan and N. A. Chowdhury, "Mid-term electricity market clearing price forecasting: A multiple SVM approach," *International Journal of Electrical Power & Energy Systems*, vol. 58, pp. 206–214, Jun.

2014, doi: 10.1016/J.IJEPES.2014.01.023.
9. X. Yan and N. A. Chowdhury, "Mid-term electricity market clearing price forecasting: A hybrid LSSVM and ARMAX approach," *International Journal of Electrical Power & Energy Systems*, vol. 53, no. 1, pp. 20–26, Dec. 2013, doi: 10.1016/J.IJEPES.2013.04.006.
10. X. Yan and N. A. Chowdhury, "Mid-term electricity market clearing price forecasting utilizing hybrid support vector machine and auto-regressive moving average with external input," *International Journal of Electrical Power & Energy Systems*, vol. 63, pp. 64–70, Dec. 2014, doi: 10.1016/J.IJEPES.2014.05.037.
11. S. Anbazhagan and N. Kumarappan, "A neural network approach to day-ahead deregulated electricity market prices classification," *Electric Power Systems Research*, vol. 86, pp. 140– 150, May 2012, doi: 10.1016/j.epsr.2011.12.011.
12. K. He, L. Yu, and L. Tang, "Electricity price forecasting with a BED (Bivariate EMD Denoising) methodology," *Energy*, vol. 91, pp. 601–609, Nov. 2015, doi: 10.1016/J.ENERGY.2015.08.021.
13. N. Chaâbane, "A hybrid ARFIMA and neural network model for electricity price prediction," *International Journal of Electrical Power & Energy Systems*, vol. 55, pp. 187–194, Feb. 2014, doi: 10.1016/J.IJEPES.2013.09.004.
14. Z. Tan, J. Zhang, J. Wang, and J. Xu, "Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models," *Appl Energy*, vol. 87, no. 11, pp. 3606– 3610, Nov. 2010, doi: 10.1016/J.APENERGY.2010.05.012.
15. M. Shafie-Khah, M. P. Moghaddam, and M. K. Sheikh-El-Eslami, "Price forecasting of dayahead electricity markets using a hybrid forecast method," *Energy Convers Manag*, vol. 52, no. 5, pp. 2165–2169, May 2011, doi: 10.1016/J.ENCONMAN.2010.10.047.
16. J. Zhang, Z. Tan, and S. Yang, "Day-ahead electricity price forecasting by a new hybrid method," *Comput Ind Eng*, vol. 63, no. 3, pp. 695–701, Nov. 2012, doi: 10.1016/J.CIE.2012.03.016.
17. J. M. Vilar, R. Cao, and G. Aneiros, "Forecasting next-day electricity demand and price using nonparametric functional methods," *International Journal of Electrical Power & Energy Systems*, vol. 39, no. 1, pp. 48–55, Jul. 2012, doi: 10.1016/J.IJEPES.2012.01.004.
18. G. C. Liao, "Hybrid Improved Differential Evolution and Wavelet Neural Network with load forecasting problem of air conditioning," *International Journal of Electrical Power & Energy Systems*, vol. 61, pp. 673–682, Oct. 2014, doi: 10.1016/J.IJEPES.2014.04.014.
19. J. P. S. Catalão, S. J. P. S. Mariano, V. M. F. Mendes, and L. A. F. M. Ferreira, "Short-term electricity prices forecasting in a competitive market: A neural network approach," *Electric Power Systems Research*, vol. 77, no. 10, pp. 1297–1304, Aug. 2007, doi: 10.1016/J.EPSR.2006.09.022.
20. J. P. S. Catalão, H. M. I. Pousinho, and V. M. F. Mendes, "Short-term electricity prices forecasting in a competitive market by a hybrid intelligent approach," *Energy Convers Manag*, vol. 52, no. 2, pp. 1061–1065, Feb. 2011, doi: 10.1016/J.ENCONMAN.2010.08.035.
21. J. P. S. Catalão, H. M. I. Pousinho, and V. M. F. Mendes, "Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 137–144, Feb. 2011, doi: 10.1109/TPWRS.2010.2049385.
22. S. Anbazhagan and N. Kumarappan, "Day-ahead deregulated electricity market price forecasting using neural network input featured by DCT," *Energy Convers Manag*, vol. 78, pp. 711–719, Feb. 2014, doi: 10.1016/J.ENCONMAN.2013.11.031.

23. I. Baqui, I. Zamora, J. Mazón, and G. Buigues, "High impedance fault detection methodology using wavelet transform and artificial neural networks," *Electric Power Systems Research*, vol. 81, no. 7, pp. 1325–1333, Jul. 2011, doi: 10.1016/J.EPSR.2011.01.022.
24. M. J. Alizadeh and M. R. Kavianpour, "Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean," *Mar Pollut Bull*, vol. 98, no. 1–2, pp. 171–178, Sep. 2015, doi: 10.1016/J.MARPOLBUL.2015.06.052.
25. Z. E. Gketsis, M. E. Zervakis, and G. Stavrakakis, "Detection and classification of winding faults in windmill generators using Wavelet Transform and ANN," *Electric Power Systems Research*, vol. 79, no. 11, pp. 1483–1494, Nov. 2009, doi: 10.1016/J.EPSR.2009.05.001.
26. C. K. Ahn, P. Shi, and L. Wu, "Receding Horizon Stabilization and Disturbance Attenuation for Neural Networks with Time-Varying Delay," *IEEE Trans Cybern*, vol. 45, no. 12, pp. 2680– 2692, Dec. 2015, doi: 10.1109/TCYB.2014.2381604.
27. C. K. Ahn, P. Shi, R. K. Agarwal, and J. Xu, "L∞ performance of single and interconnected neural networks with time-varying delay," *Inf Sci (N Y)*, vol. 346–347, pp. 412–423, Jun. 2016, doi: 10.1016/J.INS.2016.02.004.
28. S. Anbazhagan and N. Kumarappan, "Day-ahead deregulated electricity market price forecasting using recurrent neural network," *IEEE Syst J*, vol. 7, no. 4, pp. 866–872, 2013, doi: 10.1109/JSYST.2012.2225733.
29. C. K. Ahn, L. Wu, and P. Shi, "Stochastic stability analysis for 2-D Roesser systems with multiplicative noise," *Automatica*, vol. 69, pp. 356–363, Jul. 2016, doi: 10.1016/J.AUTOMATICA.2016.03.006.
30. C. K. Ahn, P. Shi, and M. V. Basin, "Two-dimensional dissipative control and filtering for roesser model," *IEEE Trans Automat Contr*, vol. 60, no. 7, pp. 1745–1759, Jul. 2015, doi:
31. 10.1109/TAC.2015.2398887.
32. C. Bentéjac and A. Csörg˝ O B Gonzalo Martínez-Muñoz, "A Comparative Analysis of XGBoost," 2019.
33. W. Liang, S. Luo, G. Zhao, and H. Wu, "Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms," *Mathematics*, vol. 8, no. 5, May 2020, doi: 10.3390/MATH8050765.
34. H. Rao *et al.*, "Feature selection based on artificial bee colony and gradient boosting decision tree," *Appl Soft Comput*, vol. 74, pp. 634–642, Jan. 2019, doi: 10.1016/J.ASOC.2018.10.036. [34] R. E. Schapire, "The Strength of Weak Learnability," 1990.
35. L. Devos, W. Meert, and J. Davis, "Fast Gradient Boosting Decision Trees with Bit-Level Data Structures." [Online]. Available: https://github.com/laudv/bitboost E. Tuv, A. Borisov, G. Runger, K. Torkkola, I. Guyon, and A. R. Saffari, "Feature Selection with Ensembles, Artificial Variables, and Redundancy Elimination," 2009.
36. T. Hastie, R. Tibshirani, and J. Friedman, "Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction."
37. S. Md. M. Hossain and K. Deb, "Plant Leaf Disease Recognition Using Histogram Based Gradient Boosting Classifier," 2021, pp. 530–545. doi: 10.1007/978-3-030-68154-8_47.
38. G. Ke *et al.*, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." [Online]. Available: https://github.com/Microsoft/LightGBM.
39. P. Perner, "Decision Tree Induction Methods and Their Application to Big Data", doi: 10.1007/978-3-319-09177-8-3.

40. C. Cortes, V. Vapnik, and L. Saitta, “Support-Vector Networks Editor,” Kluwer Academic Publishers, 1995.
41. V. Kecman, “Support Vector Machines – An Introduction,” 2005, pp. 1–47. doi: 10.1007/10984697_1.
42. Y. Lin and R. Li, “Real-time traffic accidents post-impact prediction: Based on crowdsourcing data,” *Accid Anal Prev*, vol. 145, p. 105696, Sep. 2020, doi: 10.1016/J.AAP.2020.105696.
43. V.Vapnik, “The nature of statistical learning theory,” *V.Vapnik* , 2013.
44. P. jo Chun, I. Ujike, K. Mishima, M. Kusumoto, and S. Okazaki, “Random forest-based evaluation technique for internal damage in reinforced concrete featuring multiple nondestructive testing results,” *Constr Build Mater*, vol. 253, p. 119238, Aug. 2020, doi: 10.1016/J.CONBUILDMAT.2020.119238.
45. N. D. Hoang and Q. L. Nguyen, “A novel method for asphalt pavement crack classification based on image processing and machine learning,” *Eng Comput*, vol. 35, no. 2, pp. 487–498, Apr. 2019, doi: 10.1007/s00366-018-0611-9.
46. V. H. Dang, T. B. Dieu, X. L. Tran, and N. D. Hoang, “Enhancing the accuracy of rainfall-induced landslide prediction along mountain roads with a GIS-based random forest classifier,” *Bulletin of Engineering Geology and the Environment*, vol. 78, no. 4, pp. 2835–2849, Jun. 2019, doi: 10.1007/s10064-018-1273-y.