PREDICTION OF ELECTRICITY BILL USING SML TECHNIQUE

MsAishwarya Franklin S[1], Biruntha J[2], Nivetha S[3], Kowsalya S[4]

[1]Assistant Professor, CSE, Agni College of Technology, Chennai, Tamilnadu, India.

[2]Student, B.E (CSE), Agni College of Technology, Chennai, Tamilnadu, India.

[3]Student, B.E (CSE), Agni College of Technology, Chennai, Tamilnadu, India.

[4]Student, B.E (CSE), Agni College of Technology, Chennai, Tamilnadu, India.

Abstract

One of the most crucial procedures is predicting the electricity bill. This is a highly tricky situation with no guarantees. To get around this, we can use machine learning approaches. As a result, anticipating power bills has become a hot study issue. The goal is to predict results with the highest possible accuracy using machine learning techniques. A example dataset will be made available for use. We must analyze the data in this dataset. Supervised Machine Learning Methods are used to analyse the data (SMIT). Data cleaning, data preparation, and data visualization will all be collected using this data analysis. To provide a machine learning-based strategy for accurately estimating the value of the Electricity Price Index by comparing supervise classification machine learning algorithms and predicting outcomes in the form of electricity price increase or stable state. In addition, the performance of several machine learning methods will be compared and discussed. Dataset with evaluation classification report, confusion matrix, and data prioritisation, and the results show that the proposed method is effective. machine learning algorithm technique can be compared to the best accuracy MAE,MSE,R2, and the result shows that the effectiveness of the proposed machine learning algorithm technique can be compared to the best accuracy MAE,MSE,R2.

1. INTRODUCTION:

1.1 DATA SCIENCE:

Data science is the study of extracting useful insights from data by integrating topic knowledge, computer skills, and a grasp of math and statistics statistics.Data science is a field that combines mathematics, business knowledge, tools, algorithms, and machine learning methodologies to help in decision making. the discovery of hidden insights or patterns in raw data that may be used to make important business decisions.

1.2 ARTIFICIAL INTELLIGENCE:

Artificial intelligence is the imitation of human intellectual processes by computers, particularly computer systems.AI applications include expert systems, natural language processing, speech recognition, and machine vision.

- Learning processes: This section of AI programming is focused with obtaining data and generating rules for translating it into useful knowledge. Algorithms are rules that instruct computer equipment how to do a certain activity in a step-by-step way.
- Processes of reasoning: This part of AI programming focuses on selecting the best algorithm to achieve a given result.
- Processes of self-correction: This element of AI programming is intended to constantly fine-tune algorithms in order to produce the most accurate results feasible.

Machine learning is used to forecast the future based on past data. Machine learning (ML) is a kind of artificial intelligence (AI) that allows computers to learn without having to be explicitly programmed.In the training and prediction phase, specialized algorithms are utilized. It feeds training data to an algorithm, which utilizes the training data to make predictions on fresh test data. Machine learning may be divided into three areas. There are three types of learning: supervised learning, unsupervised learning, and reinforcement learning.

- The supervised learning algorithm is provided the input data as well as the relevant labelling to learn data, which must be labelled by a person beforehand.
- Unsupervised learning does not use labels. It supplied the learning algorithm. This algorithm must find out how to cluster the supplied data.
- Reinforcement learning is a type of learning that interacts with its environment in a dynamic way and receives both positive and negative feedback in order to improve its performance.

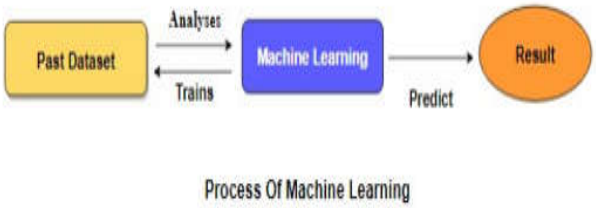

Figure 1: Process of Machine Learning

2. EXISTING SYSTEM:

Because they cannot access confidential power system models and operating information, electricity market participants rely on data-driven methodologies that use public market data to anticipate locational marginal pricing (LMPs) and determine optimal bidding tactics. In this study, system-wide heterogeneous public market data is arranged into a three-dimensional (3D) tensor, NN, that can store their spatiotemporal correlations. Learning the spatio-temporal correlations encoded in the historical market data tensor to predict real-time locational marginal prices (RTLMPs). To increase forecast accuracy, an autoregressive moving average (ARMA) calibration method is used. The suggested method may learn spatio-temporal connections among RTLMPs and conduct accurate RTLMP prediction, according to Case research based on Mid-continent Independent System Operator (MISO) and Southwest Power Pool public market data(SPP).Another relevant area of research is predicting the energy price component of LMP by evaluating the entire supply-demand balance. Other spatial learning algorithms can forecast the price component of congestion and the price component of loss. These three independent pricing components can be used to assemble the final spatio-temporal correlated nodal LMPs.

2.1 DRAWBACKS OF EXISTING SYSTEM:

- They do not use a machine learning classification technique to classify electricity prices, and they do not provide any accuracy statistics.
- As a result, it will be unable to analyse the consistency of power price projection data and produce more accurate prediction results.

3. PROPOSED SYSTEM:

3.1 Electricity Prediction Exploratory Data Analysis

Multiple datasets from various sources would be integrated to make a generalized dataset, and then various machine learning methods would be used to extract patterns and obtain the most accurate findings possible.

3.2 Data wrangling:

The data will be loaded into this area of the report, and it will be checked for cleanliness before being trimmed and cleaned for analysis. Ensure that the document steps are followed carefully and that cleaning decisions are justified.

3.3 Data Collection:

The data set for predicting provided data is divided into two parts: training and testing. In most cases, 7:3 ratios are used to divide the Training and Test sets.The Data Model, which was generated using machine learning methods, is applied to the Training set, and the Test set is predicted based on the test result accuracy.

3.4 Building the Regression Model:

The decision tree algorithm prediction model for estimating the price of electricity is effective for the following reasons: In a classification problem, it produces superior outcomes. It excels in removing outliers, irrelevant variables, and a mix of continuous, categorical, and discrete data during preprocessing.It generates out-of-bag estimate error, which has been shown to be unbiased in numerous experiments and is quite simple to tweak.

3.5 ADVANTAGES OF PROPOSED SYSTEM:

These publications look into the applicability of machine learning approaches for predicting electricity prices in real-world situations.

Finally, it discusses potential future research topics, obstacles, and requirements.

4. REVIEW OF LITERATURE SURVEY

Title : Electricity price forecasting in deregulated markets: A review and evaluation

Author: Sanjeev Kumar Aggarwal, Lalit Mohan Saini, Ashwani Kumar

Year :2008

This study examines the various approaches for projecting power prices. The following price predicting methods have been discussed: I stochastic time series, (ii) causal models, and (iii) models based on artificial intelligence. The quantitative analysis of various authors' work has been offered based on (a) prediction time horizon, (b) input variables, (c) output variables, (d) results, (e) data points utilized for analysis, (f) preprocessing technique used, and (g) model architecture. For convenience of comparison, the findings have been given in the form of tables. A classification of several price-influencing factors utilized by various researchers has been completed and made available for reference.It is also possible to analyse the applicability of various models to diverse electrical markets.

Title : An Integrated Machine Learning Model for Day-Ahead Electricity Price Forecasting

Author: Shu Fan, James R. Liao, Kazuhiro Kaneko, Member, IEEE, and Luonan Chen, Senior Member, IEEE

Year : 2006

This research offers a novel model for forecasting short-term power prices based on the combination of two machine learning technologies: Bayesian Clustering by Dynamics (BCD) and Support Vector Machine (SVM). An integrated architecture is used in the suggested forecasting system. To begin, an unsupervised BCD classifier is used to cluster the input data set into multiple subgroups. The training data of each subgroup is then supervised fitted using groups of 24 SVMs for the next day's power price profile. To illustrate its efficacy, the suggested model was trained and tested using historical energy prices from the New England electrical market.

Title : A Review for Electricity Price Forecasting Techniques in Electricity Markets

Author: Ankur Jain, Ankit Tuli, Misha Kakkar

Year : 2013

Electricity Price forecasting is becoming an essential and critical problem in every country. This document highlights the different suggested price forecasting models and approaches. Price forecasting accuracy has a significant influence on the revenues of transmission companies, distributors, and suppliers, to name a few. It provides a thorough examination of several techniques to price forecasting. ARIMA (Auto Regressive Integrated Moving Average), LSSVM (Least Square Support Vector Machine), LLWNN (Local Linear Wavelet Neural Network), ANN (Artificial Neural Network), and other techniques and methodologies are contrasted and their weaknesses and strengths are assessed. Because both consumers and producers rely on price forecasting information for their bidding tactics, an efficient solution is necessary.

Title : Price Forecasting for the Balancing Energy Market Using Machine-Learning Regression

Author: Alexandre Lucas, Konstantinos Pegios, Evangelos Kotsakis and Dan Clarke

Year :2020

With the expansion of aggregators and the general liberalization of European power markets in recent years, the relevance of price forecasting has grown. Market players must choose between bidding in a lower-priced (day-ahead) market with morevolume and aiming for a market with a smaller volume but perhaps higher benefits (balance energy market).Companies attempt to foresee revenue or price extremes in order to manage risk and opportunity while allocating their assets optimally. Electricity markets are said to be based on quasi-deterministic principles rather than speculation in general. which explains the urge to anticipate the price based on factors that can define the market's result. Many studies handle this issue statistically or by doing multiple-variable regressions, although they frequently focus solely on time series analysis. For the first time in 2019, the Loss of Load Probability (LOLP) was made available in the United Kingdom. Taking use of this potential, this study focuses on five LOLP variables with varying time-ahead In order to explain the pricing behaviour of a multi-variable regression, new quasi-deterministic factors (e.g., estimates) and further quasi-deterministic factors These include contributions from base production, system load, solar and wind generation, seasonality, day-ahead pricing, and imbalance volume. To evaluate

performance, three machine-learning methods were used: Gradient Boosting (GB), Random Forest (RF), and XGBoost. Because of its better performance, XGBoost was picked to carry out the real-time prediction phase. The model produces a Mean Absolute Error (MAE) of 7.89 £/MWh, an R2 score of 76.8 percent, and a Mean Squared Error (MSE) of 124.74. The Net Imbalance Volume and the LOLP are the variables that contribute the most to the model. (aggregated), the month, and the De-rated margins (aggregated), with feature significance weights of 28.6 percent, 27.5 percent, and 27.5 percent, 14.0 percent, and 8.9 percent, respectively.

Title :Energy Markets Forecasting. From Inferential Statistics to Machine Learning: The German Case

Author: Emma Viviani, Luca Di Persio and Matthias Ehrhardt

Year : 2021

In this paper, we study a probabilistic strategy for predicting power prices that outperforms previous ones. We begin by analyzing statistical techniques for point forecasting in terms of efficiency, accuracy, and dependability, and then use Neural Network methodologies to develop a hybrid model for probabilistic type forecasting. We demonstrate that our solution meets the highest efficiency and precision standards by evaluating its output using German power pricing data.

5. SCOPE OF THE PROJECT:

The main goal is to discover Electricity Bill Prediction, which is a standard text regression issue that may be solved using a machine learning method. We must deploy the Flask framework when we have got the accuracy.

6. SYSTEM ARCHITECTURE:

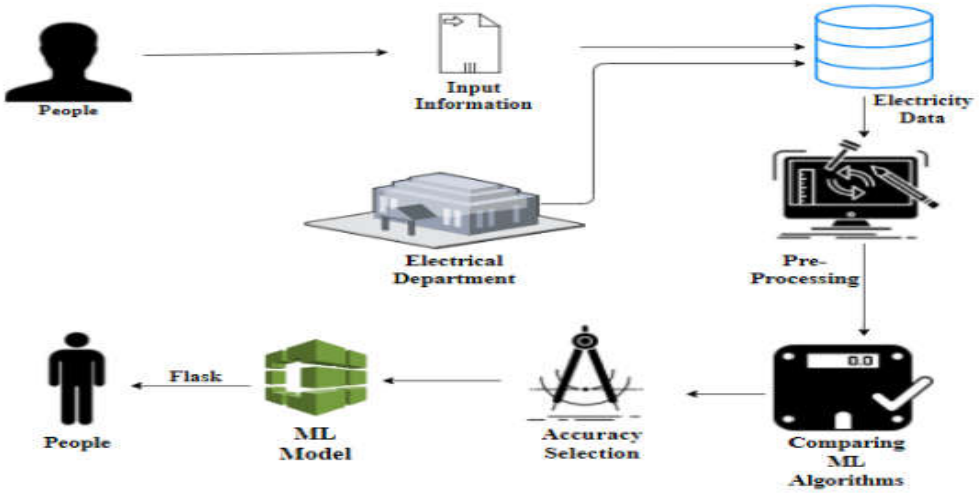

Figure 2: System Architecture

7. WORK FLOW DIAGRAM:

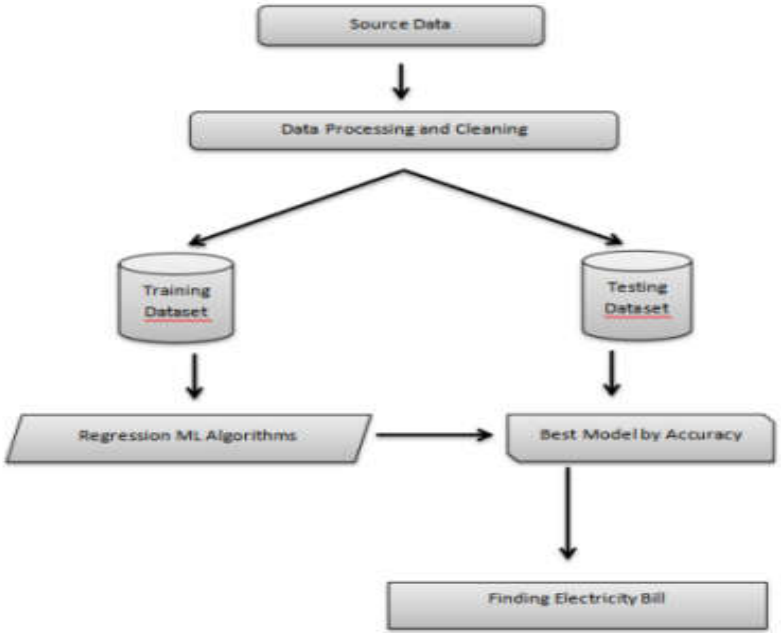

Figure 3 : Work Flow Diagram

8. MODULES:

- Data preprocessing
- Data analysis of visualization
- Comparing Algorithm with prediction in the form of best accuracy result
- Deployment using Flask

MODULE DESCRIPTION:

8.1 MODULE 1

DATA PREPROCESSING:

Loading the supplied dataset and importing library packages. Identifying variables depending on the form and kind of the data, as well as analysing missing and duplicate values. A validation dataset is a sample of data kept back after training your model that is used to measure model skill when tweaking models and processes for making the greatest use of validation and test datasets while evaluating your models. Data cleaning/preparation is accomplished by renaming the provided dataset and removing the To analyse uni-variate, bi-variate, and multi-variate processes, among other things, columns are usedstuff. The methods and techniques for cleaning data will differ depending on

the dataset. The fundamental purpose of data cleaning is to find and fix mistakes and abnormalities so that data may be used for analytics and decision-making.

8.2 MODULE 2

DATA ANALYSIS OF VISUALISATION:

In applied statistics and machine learning,data visualization is a crucial ability. Statistics is concerned with quantitative data descriptions and estimations. Data visualisation is a set of tools that may help you gain a qualitative understanding of data. This might be useful for spotting patterns, faulty data, outliers, and other things when exploring and getting to know a dataset. Data visualizations can be used to express and demonstrate crucial relationships in plots and charts that Measurements of connection or relevance with a small issue are more visceral and relevant to stakeholders. Data visualization and exploratory data analysis are fields in and of themselves, and it will For more information, it is suggested that you read some of the works listed at the end. Data that is not presented in a visual style, such as charts and graphs, may not make sense. In both applied statistics and applied machine learning, the ability to swiftly visualise data samples and other objects is a critical skill. It will show you how to use the many sorts of plots available when visualizing data in Python to better understand your own data.

How to graph time series data using line plots and how to graph data with bar charts categorical numbers, as well as

How to use histograms and box plots to describe data distributions.

8.3 MODULE 3

COMPARING ALGORITHM WITH PREDICTION IN THE FORM OF BEST ACCURACY RESULT:

Support Vector Regression:

Support Vector Machines (SVM) are commonly utilized in machine learning for classification and regression tasks. The regression problem is a variant of the classification problem in which the model produces a continuous-valued output rather than a finite-valued result. A regression model, in other words, estimates a continuous-valued multivariate function. By recasting binary classification issues as convex optimization problems, SVMs are utilised to solve them.The optimization task is to find the biggest margin separating the hyperplane while correctly classifying as many training points as possible.SVMs describe this ideal hyperplane using support vectors. The SVM's sparse solution and high generalization make it well suited to regression issues.The introduction of a -insensitive zone around the function, known as the ε -tube, allows SVM to be generalized to SVR. This tube rewrites the optimization problem to find the tube that best approximates the continuous-valued function while balancing model complexity and prediction error. SVR is defined as an optimization problem by first constructing a convex ε -insensitive loss should be minimised, and then finding the flattest tube that contains the bulk of the function training cases.As a result, the loss function is combined with the geometrical parameters of the tube to produce a multi-objective function. The convex optimization, which has a unique solution, is then solved using appropriate numerical optimization methods.

Decision Tree Regression:

Regression using a Decision Tree. In the shape of a tree structure, a decision tree constructs regression or classification models. It gradually cuts down a dataset into smaller and smaller sections while also developing an associated decision tree. The end result is a tree containing leaf nodes and decision nodes.In the shape of a tree structure, a decision tree constructs regression or classification models. It gradually cuts down a dataset into smaller and smaller sections while also developing an associated decision tree.

The ultimate result is a tree with decision nodes and leaf nodes. A decision node (such as Outlook) can contain two or more branches (such as Outlook). Sunny, Overcast, and Rainy), each of which represents a value for the attribute being checked. A choice on the numerical aim is represented as a leaf node (e.g., Hours Played). The best predictor is represented by the root node, which is the topmost decision node in a tree. Both category and numerical data may be handled by decision trees.

RANDOM FOREST REGRESSION:

Random Forest Regression is a supervised learning approach for regression that use the ensemble learning method. A random forest is a meta estimator that uses averaging to improve the predictability and control of results Fitting a number of classification decision trees on various sub-samples of the dataset to achieve over-fitting. The sub-sample size is controlled by the max samples parameter if Bootstrap = True (default); otherwise, the whole dataset is used to generate each tree.

LINEAR REGRESSION:

One of the most basic and widely used Machine Learning methods is linear regression. It's a statistical technique for performing predictive analysis. Sales, salary, age, product price, and other continuous/real or quantitative parameters are all predicted using linear regression.The term comes from the fact that the linear regression approach displays a linear relationship between a dependent (y) variable and one or more independent (y) variables.Because linear regression reveals a linear connection, it determines how the value of the dependent variable changes as the value of the independent variable changes. In the linear regression model, the relationship between the variables is represented by a slanted straight line.

LASSO REGRESSION:

"LASSO" stands for "Least absolute Shrinkage and Selection Operator." It's a statistical formula for regularizing data models and selecting features. Lasso regression is a regularisation technique. It is chosen over regression techniques for a more accurate forecast. Shrinkage is used in this model. In shrinkage, data values are shrunk towards a central point known as the mean. Simple, sparse models are encouraged by the lasso approach (i.e., models with fewer parameters). This form of regression is ideal for models with a lot of multicollinearities or when You want to automate parts of the model selection process, such variable selection and parameter selection removal.

RIDGE REGRESSION:

Ridge regression is a model tuning technique that may be used to analyse data with multicollinearity. This method is used to produce L2 regularisation. When multicollinearity is a concern, least-squares is unbiased, and variances are large, the predicted values are far from the actual values. Ridge regression is a method for removing multicollinearity from data models. Ridge regression is the most suited approach when the number of observations is less than the number of predictor variables.

8.4 DEPLOYMENT:

FLASK:

Framework Flask is a web framework written in the Python programming language. Flask is a library and collection of code that may be used to create websites without having to start from scratch. The Model View Controller (MVC) technique is still not used by Framework frame work. Flask-RESTful is a Flask plugin that adds further features for creating REST APIs. The time it takes to design an API will never disappoint you. Flask-Restful is a simple abstraction that integrates with current ORMs and frameworks. Flask-RESTful promotes excellent practices while requiring less setup.

Flask Restful is a Flask plugin for creating REST APIs in Python using Flask as the backend.It encourages good habits and is simple to apply. Flask Restful is a Flask plugin that allows you to create REST APIs in Python with Flask as the backend. It promotes excellent practices and is simple to implement. Restful is simple to learn if you're already familiar with flask. Flask is a Python web framework that includes capabilities for constructing online applications, such as handling HTTP requests and displaying templates, as well as the ability to add to this application to create an API.

9. RESULT ANALYSIS:

After the completion of data cleaning, data preprocessing and data analysis there will be a comparison between those mentioned algorithms to find the best accuracy level. When comparing algorithms, the one that provides the highest level of accuracy is chosen as the working algorithm.

In the output there will be a webpage to know the details about the customer’s house type, number of rooms in the house, number of people in the house, area in which the house is present etc.Then the sample data are compared and then the best accuracy level will be found.

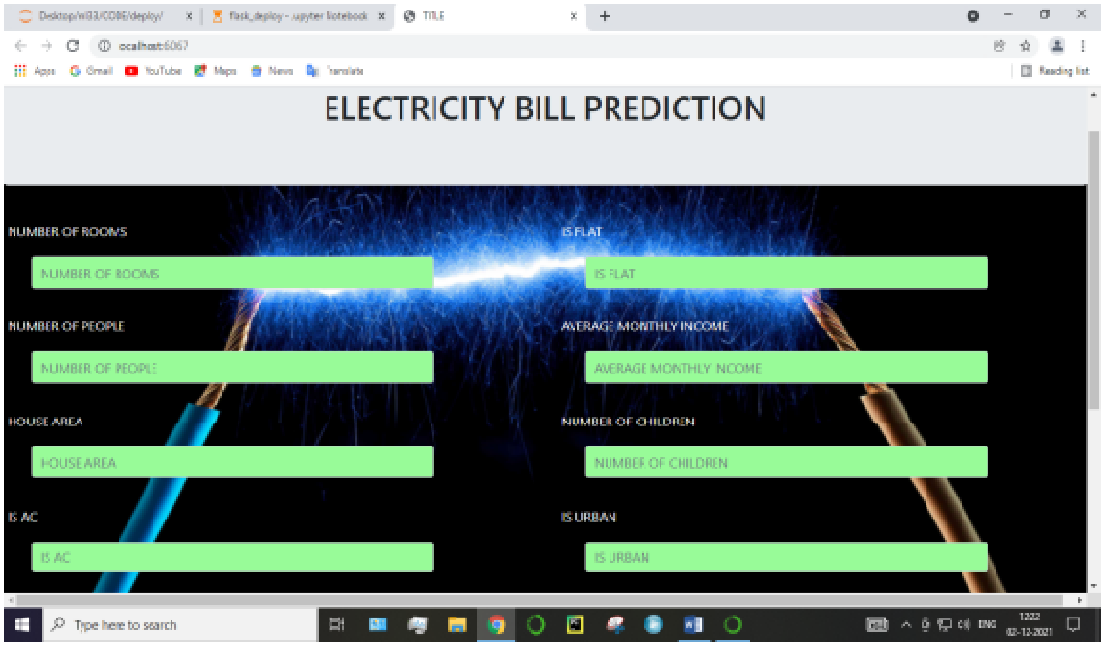

Figure 4 : Final output

10.CONCLUSION:

The analytical procedure began with data cleaning and processing, followed by missing value analysis, exploratory analysis, and lastly model creation and assessment. The highest accuracy score on a public test set will be determined. This application can assist you in locating the Electricity Bill.

11.FUTURE WORK:

The forecast of electricity bills will be linked to an AI algorithm.

To automate this procedure by displaying the prediction result in a cloud-based web or desktop application.

To optimize the task for implementation in an AI context.

REFERENCES:

[1] ISO New England, "Market rule 1," 2019. [Online]. Available: www.iso-ne.com/participate/rules-procedures/tariff/market-rule-1

[2] PJM Interconnection, "Operating agreement of pjm interconnection, l.l.c." 2011. [Online]. Available: pjm.com/directory/merged-tariffs/oa.pdf

[3] F. C. Schweppe, *Spot pricing of electricity / by Fred C. Schweppe ... [et al.]*. Kluwer Academic Boston, 1988.

[4] V. Kekatos, G. B. Giannakis, and R. Baldick, "Online energy price ma- trix factorization for power grid topology tracking," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1239–1248, 2016.

[5] Q. Zhou, L. Tesfatsion, and C. Liu, "Short-term congestion forecasting in wholesale power markets," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2185–2196, 2011.

[6] G. Hamoud and I. Bradley, "Assessment of transmission congestion cost and locational marginal pricing in a competitive electricity market," *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 769–775, May 2004.

[7] W. Deng, Y. Ji, and L. Tong, "Probabilistic forecasting and simulation of electricity markets via online dictionary learning," 2016.

[8] J. F. Toubeau, T. Morstyn, J. Bottieau, K. Zheng, D. Apostolopoulou, Z. De Gre`ve, Y. Wang, and F. Valle´e, "Capturing spatio-temporal dependencies in the probabilistic forecasting of distribution locational marginal prices," *IEEE Transactions on Smart Grid*, pp. 1–1, 2020.

[9] X. Geng and L. Xie, "A data-driven approach to identifying system pattern regions in market operations," in *2015 IEEE Power Energy Society General Meeting*, July 2015, pp. 1–5.

[10] Y. Ji, R. J. Thomas, and L. Tong, "Probabilistic forecasting of real-time lmp and network congestion," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 831–841, March 2017.

[11] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," 2018.

[12] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015.

[13] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," 2015.

[14] E. Denton, S. Chintala, A. Szlam, and R. Fergus, "Deep generative image models using a laplacian pyramid of adversarial networks," 2015.

[15] X. Zhou, Z. Pan, G. Hu, S. Tang, and C. Zhao, "Stock market prediction on high-frequency data using generative adversarial nets," *Mathematical Problems in Engineering*, vol. 2018, pp. 1–11, 04 2018.

[16] M. Claesen and B. D. Moor, "Hyperparameter search in machine learning," 2015.

[17] W. Polasek, "Time series analysis and its applications: With r examples, third edition by robert h. shumway, david s. stoffer," *International Statistical Review*, vol. 81, no. 2, pp. 323–325, 2013.

[18] MATLAB, "Estimate parameters of armax model," 2019. [Online]. Available:https://www.mathworks.com/help/ident/ref/armax.html

[19] "Tensorflow," 2019. [Online]. Available:https://www.tensorflow.org/

[20] Y. Chen, "A new methodology of spatial cross-correlation analysis," *PLOS ONE*, vol. 10, no. 5, p. e0126158, May 2015. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0126158

"Miso market data," 2020. [Online]. Avail- able: https://www.misoenergy.org/markets-and-operations/real-time– market-data/market-reports/