© March 2023| IJIRT | Volume 9 Issue 10 | ISSN: 2349-6002

Energy Consumption Forecasting using Machine Learning

Neenu Johnson[1], Vibha Rajput[2], Shreya Cholera[3]
[1]*Asst. Prof., Department of IT, Thakur College of Science and Commerce*
[2,3]*Department of Data Science, Thakur College of Science and Commerce*

Abstract-Forecasting electricity demand and consumption accurately is critical to the optimal and cost effective operation system, providing a competitive advantage to companies. In working with seasonal data and external variables, the traditional time-series forecasting methods cannot be applied to electricity consumption data. In energy planning for a generating company, accurate power forecasting for the electrical consumption prediction, as a technique, to understand and predict the market electricity demand is of paramount importance.

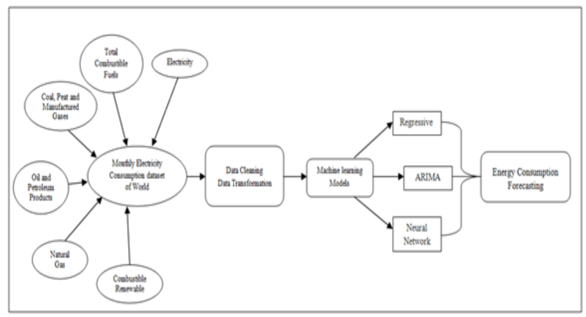
Their power production can be adjusted accordingly in a deregulated market. As data type is seasonal, Seasonal Autoregressive Integrated Moving Averages with exogenous repressors (SARIMAX), and Decision Tree, Random forest is used to explicitly deal with seasonality as a class of time-series forecasting models.

The main purpose of this project is to perform exploratory data analysis of the Spain power, then use different forecasting models to once-daily predict the next 24 hours of energy demand and daily peak demand. To split the electricity consumption data into training and test sets.

The obtained results showed that the machine learning algorithms proposed in the recent literature outperformed the tested algorithms. Models are evaluated using root mean squared error (RMSE) to be directly comparable to energy readings in the data. RMSE has calculated two ways. First to represent the error of predicting each hour at a time. Second to represent the models' overall performance. The results show that electricity demand can be modeled using machine learning algorithms, deploying renewable energy, planning for high/low load days, and reducing wastage from polluting on reserve standby generation, detecting abnormalities in consumption trends, and quantifying energy and cost-saving measures.

INTRODUCTION

In artificial intelligence (AI) in general and machine learning (ML) techniques in specific terms as well as a growing trove of publicly available energy consumption data have been proposed for accurate power forecasting and optimal decision making in energy planning. It enables generating companies to manage energy demand effectively to cost.



Processes for Forecasting Energy Consumption

Energy consumption is on the increase and has a significant impact on the environment. Current predictions show that the growing trend of CO2 emissions which is often held responsible for most of the Earth's progressive warming will continue therefore, international political, economic, and environmental research has focused on energy consumption reduction and energy efficiency improvement to cope with the problem of global warming and over-exploitation of natural resources.

LITERATURE REVIEWS

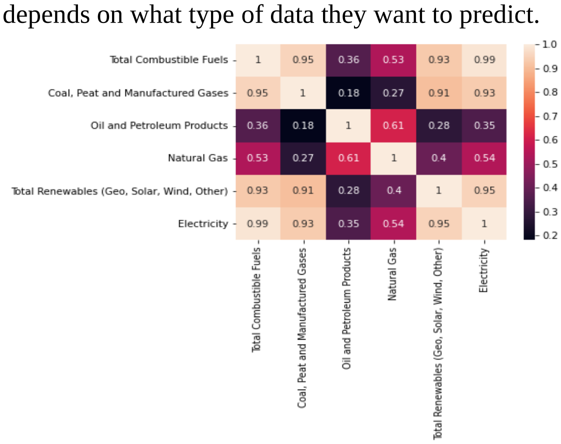
Energy is essential to the functioning of all activities of nation-states, be they developed or developing. As such, a number of energy consumption forecasting models have been developed using economic, social, geographic, and demographic factors. Energy demand models can be classified in several ways such as static versus dynamic, Univariate versus multivariate, techniques ranging from time series to hybrid models. In utilized a univariate ARIMA (Auto Regressive Integrated Moving Average) model to predict patterns in energy supply and demand in the northern region of

Spain of Asturias. Ceylan and Ozturk used the GNP of Turkey, its population and import, export figures as a basis for two forms of the GAEDM model to calculate the energy demand. Crompton and Wu attempted at predicting the energy consumption of China via a Bayesian vector-based autoregression method. The results showed low growth, predicting a slowing down in its growth, which opened the discussion on its potential. Mohamed and Bodger used the GDP, cost of electricity, and population via a multi-linear regression model to predict the power consumption of New Zealand.

METHODS

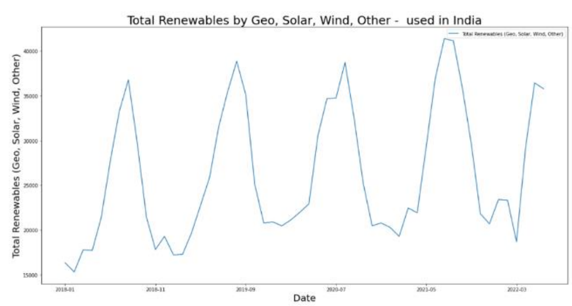
Python is an extremely popular choice in the field of machine learning and AI development. Its short and simple syntax make it extremely easy to learn. Python is a general-purpose language, which means it is designed to be used in a range of applications, including data science, software and web development, automation, and generally getting stuff done.

Machine learning (ML) is a type of artificial intelligence (AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning algorithms use historical data as input to predict new output values. Recommendation engines are a common use case for machine learning. Classical machine learning is often categorized by how an algorithm learns to become more accurate in its predictions. There are four basic approaches: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. The type of algorithm data scientists chooses to use depends on what type of data they want to predict.



To create and assess all of our models, we use a series of helper functions that perform the following functions.

1. Train test split: We separate our data so that the last 12 months of each Country are part of the test set and the rest of the data is used to train our model
2. Scale the data: using a min-max scale, we will scale the data so that all of our variables fall within the range of -1 to 1
3. Reverse scaling: After running our models, we will use this helper function to reverse the scaling of step 2
4. Create a predictions data frame: generate a data frame that includes the actual sales captured in our test set and the predicted results from our model so that we can quantify our success
5. Score the models: this helper function will save the root mean squared error (RMSE) and mean absolute error (MAE) of our predictions to compare performance of our five models
6. Regressive models: For our regressive models, we can use the fit-predict structure of the scikit-learn library. We therefore can set up a base modeling structure that we will call for each model.
7. The ARIMA model looks slightly different than the models above. We use the stats models SARIMAX package to train the model and generate dynamic predictions.

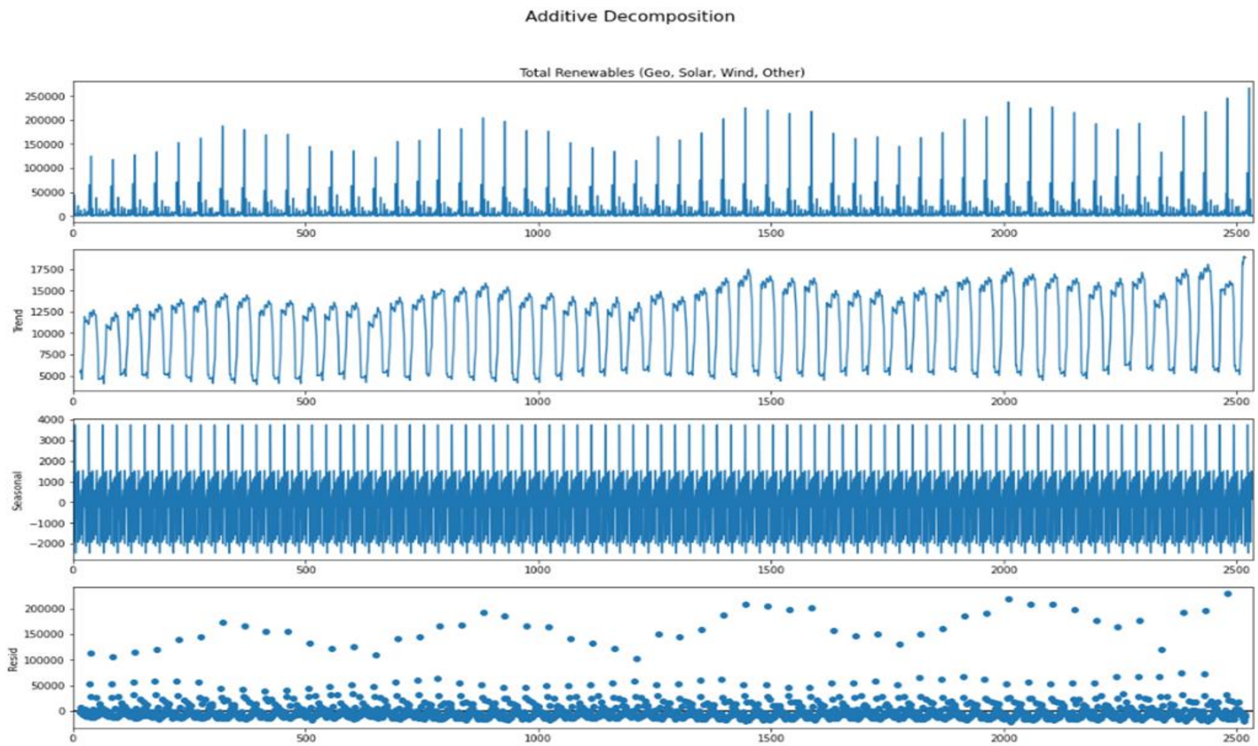


Algorithms:

1. SARIMAX:

Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors, or SARIMAX, is an extension of the ARIMA class of models. Intuitively, ARIMA models compose 2 parts: the autoregressive term (AR) and the moving-average term (MA). The former views the value at one time just as a weighted sum of past values. The latter model that same value also as a weighted sum but of past residuals (confer. *time series decomposition*). There is also an integrated term (I) to difference the time series. SARIMA is the predecessor of SARIMAX. One shorthand notation for SARIMA models is:

$$\text{SARIMA } (p,d,q) \times (P,D,Q,S)$$



2. Linear Regression:
Regression is a technique for investigating the relationship between independent variables or features and a dependent variable or outcome. It's used as a method for predictive modeling in machine learning, in which an algorithm is used to predict continuous outcomes. Regression is used to identify patterns and relationships within a dataset, which can then be applied to new and unseen data. This makes regression a key element of machine learning in finance and is often leveraged to help forecast portfolio performance or stock costs and trends. Models can be trained to understand the relationship between a variety of diverse features and a desired outcome.

3. Logistic Regression:
Logistic regression is used when the dependent variable can have one of two values, such as true or false, or success or failure. Logistic regression models can be used to predict the probability of a dependent variable occurring. Generally, the output values must be binary. A sigmoid curve can be used to map the relationship between the dependent variable and independent variables.

4. Decision Tree:
Probably the most intuitive approach is to consider the observed time-series as a function of time itself, i.e.

$$y_t = f(t) + \epsilon$$

With some stochastic additive error term. In an earlier article, I have already made some remarks on why regression against time itself is problematic.

5. Random forest:
It is widely used for classification and regression predictive modeling problems with structured (tabular) data sets, e.g. data as it looks in a spreadsheet or database table.
Random Forest can also be used for **time series forecasting**, although it requires that the time series dataset be transformed into a supervised learning problem first. It also requires the use of a specialized technique for evaluating the model called walk-forward validation, as evaluating the model using k-

fold cross validation would result in optimistically biased results.

FUTURE WORK

The forecasting can be enhanced further to add following functionality:

- The main future work's focus will be on updating the datasets from time to time to produce accurate predictions and the process can be automated.
- We can provide customer a genuine-time market of energy consumption.
- We can also provide better understanding by visualization of graph by each country.

CONCLUSION

In this project, the monthly electricity load consumption is used to forecast future load electricity demands. As such, traditional techniques may not be able to forecast future values accurately. The monthly electricity load values between 01/01/2018 to 30/06/2022, are reported in World energy dataset. We summarized the importance of demand forecasting and related literature. To explore the dataset's characteristics, we started with exploratory data analysis, providing descriptive information. In the data cleaning process, we will replace null values with mean values, extracted redundant attributes, and aggregated monthly load values monthly level to see the trend and seasonality functions more clearly.

REFERENCE

[1] Mohammadigohari, Mahdi, "Energy consumption forecasting using machine learning" (2021). Thesis. Rochester Institute of Technology. Accessed from https://scholarworks.rit.edu/theses/10989

[2] Mosavi, A.; Bahmani, A. Energy Consumption Prediction Using Machine Learning; A Review. Preprints 2019, 2019030131. https://doi.org/10.20944/preprints201903.0131.v1

[3] Montoya, Francisco G., Maria G. Montoya, Julio Gomez, Francisco Manzano-Agugliaro, and Enrique Alameda-Hernandez. "The research on energy in Spain: A scientometric approach." *Renewable and Sustainable Energy Reviews* 29 (2014): 173-183. https://www.sciencedirect.com/science/article/pii/S1364032113006345

[4] Olu-Ajayi, Razak, Hafiz Alaka, Ismail Sulaimon, Funlade Sunmola, and Saheed Ajayi. "Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques." *Journal of Building Engineering* 45 (2022): 103406. https://www.sciencedirect.com/science/article/abs/pii/S235271022101264X

[5] Cheng, Hui-fang, Jian-li Zhao, and Hai-hua Qing. "Energy Consumption Development Based on Gene Expression Programming model in Hebei Province." *International Journal of Electronics Communication and Computer Engineering* 6, no. 5 (2015): 577. https://ijecce.org/administrator/components/com_jresearch/files/publications/IJECCE_3570_Final.pdf

[6] Fathi, Soheil, Ravi Srinivasan, Andriel Fenner, and Sahand Fathi. "Machine learning applications in urban building energy performance forecasting: A systematic review." *Renewable and Sustainable Energy Reviews* 133 (2020): 110287. https://www.sciencedirect.com/science/article/pii/S136403212030575X

[7] Chou, Jui-Sheng, and Dinh-Nhat Truong. "Multistep energy consumption forecasting by metaheuristic optimization of time-series analysis and machine learning." *International Journal of Energy Research* 45, no. 3 (2021): 4581-4612. https://onlinelibrary.wiley.com/doi/abs/10.1002/er.6125

[8] Bourhnane, Safae, Mohamed Riduan Abid, Rachid Lghoul, Khalid Zine-Dine, Najib Elkamoun, and Driss Benhaddou. "Machine learning for energy consumption prediction and scheduling in smart buildings." *SN Applied Sciences* 2 (2020): 1-10. https://rdcu.be/c6ziU