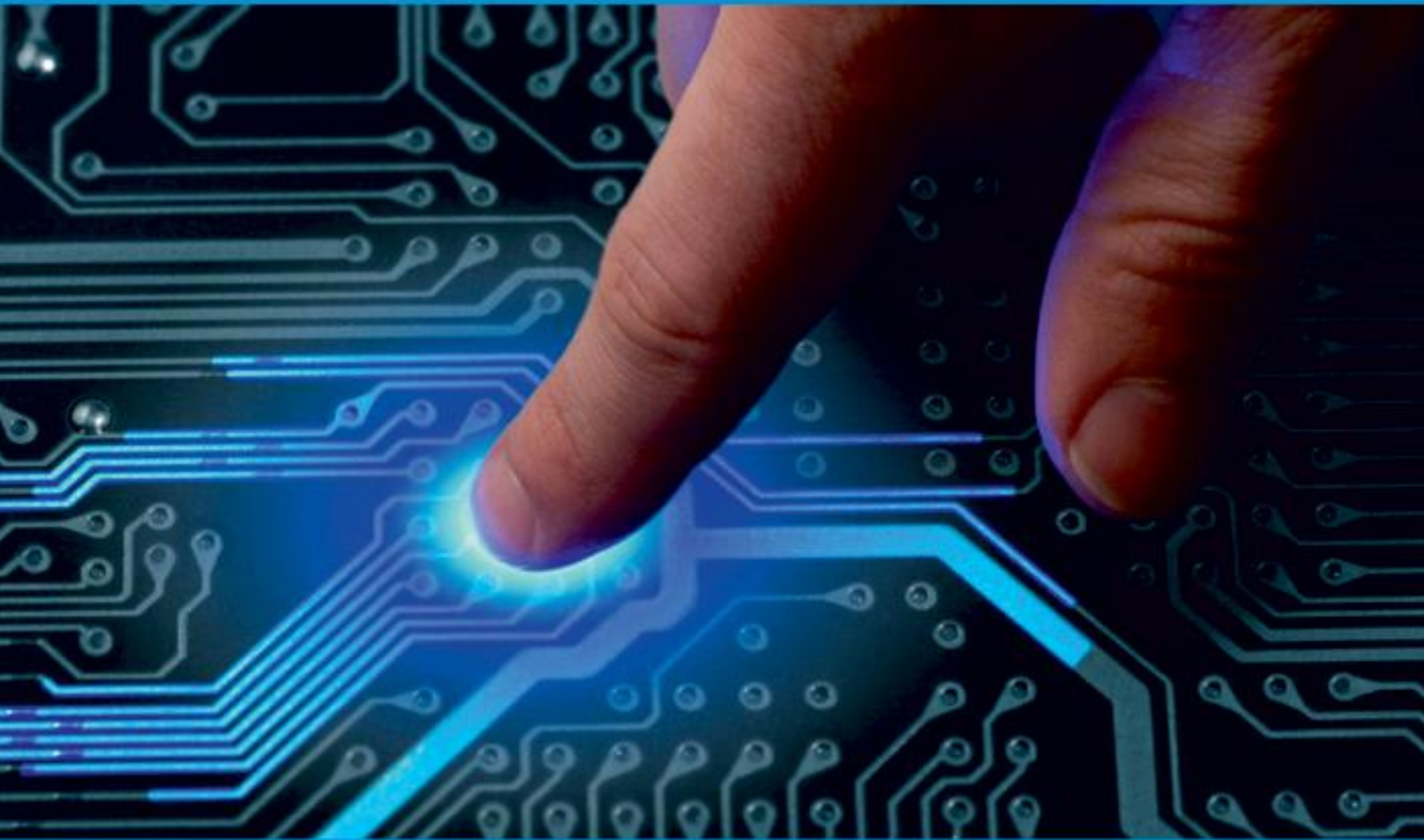
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205161 |

A Performance Comparison of Machine Learning Algorithm for Load Forecasting in Smart Grid

Jay There, Prof Pushpa Tandekar, Prof Lowlesh Yadav

Department of Computer Science and Engineering, Shri Sai College of Engineering and Technology, Chandrapur, India

ABSTRACT: Load forecasting plays a crucial role in the efficient management and planning of electricity distribution in smart grids. Machine learning algorithms have shown promising results in load forecasting, enabling accurate predictions and aiding decision-making processes. This paper presents a comprehensive performance comparison of various machine learning algorithms for load forecasting in smart grids. The study evaluates and compares the performance of multiple machine learning algorithms, including but not limited to Support Vector Machines (SVM), Random Forests (RF), Artificial Neural Networks (ANN), and Gradient Boosting Machines (GBM). Real-world load data from a smart grid system is used as the dataset for training and testing the algorithms.

The performance of each algorithm is evaluated based on several metrics, such as mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). Additionally, the computational complexity and training time of each algorithm are considered to assess their suitability for real-time load forecasting applications.

KEYWORDS: Object detection; YOLO v5; Open CV; image processing, Neural Network

I. INTRODUCTION

The smart grid concept is introduced to accelerate operational efficiency and enhance the reliability and sustainability of the power supply. The load forecasting technique involves estimating future loads using historical and present data. In a smart grid, the forecasting of loads is done by considering the power consumption by users and the power produced by all types of generations (renewable and non-renewable) with the help of smart energy meters. In this study, we consider a range of machine learning algorithms commonly used in load forecasting tasks. These include Support Vector Machines (SVM), Random Forests (RF), Artificial Neural Networks (ANN), and Gradient Boosting Machines (GBM), among others. Each algorithm has its own characteristics and underlying principles, which can impact their performance in load forecasting. Supervised vs. unsupervised algorithms.

The word Machine Learning **was first coined by Arthur Samuel in 1959. The definition of machine learning can be defined as that machine learning** gives computers the ability to learn without being explicitly programmed**. Also in 1997, Tom Mitchell defined machine learning that** "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E". Machine learning is considered to be the most interesting field of computer science.

Working of Machine Learning

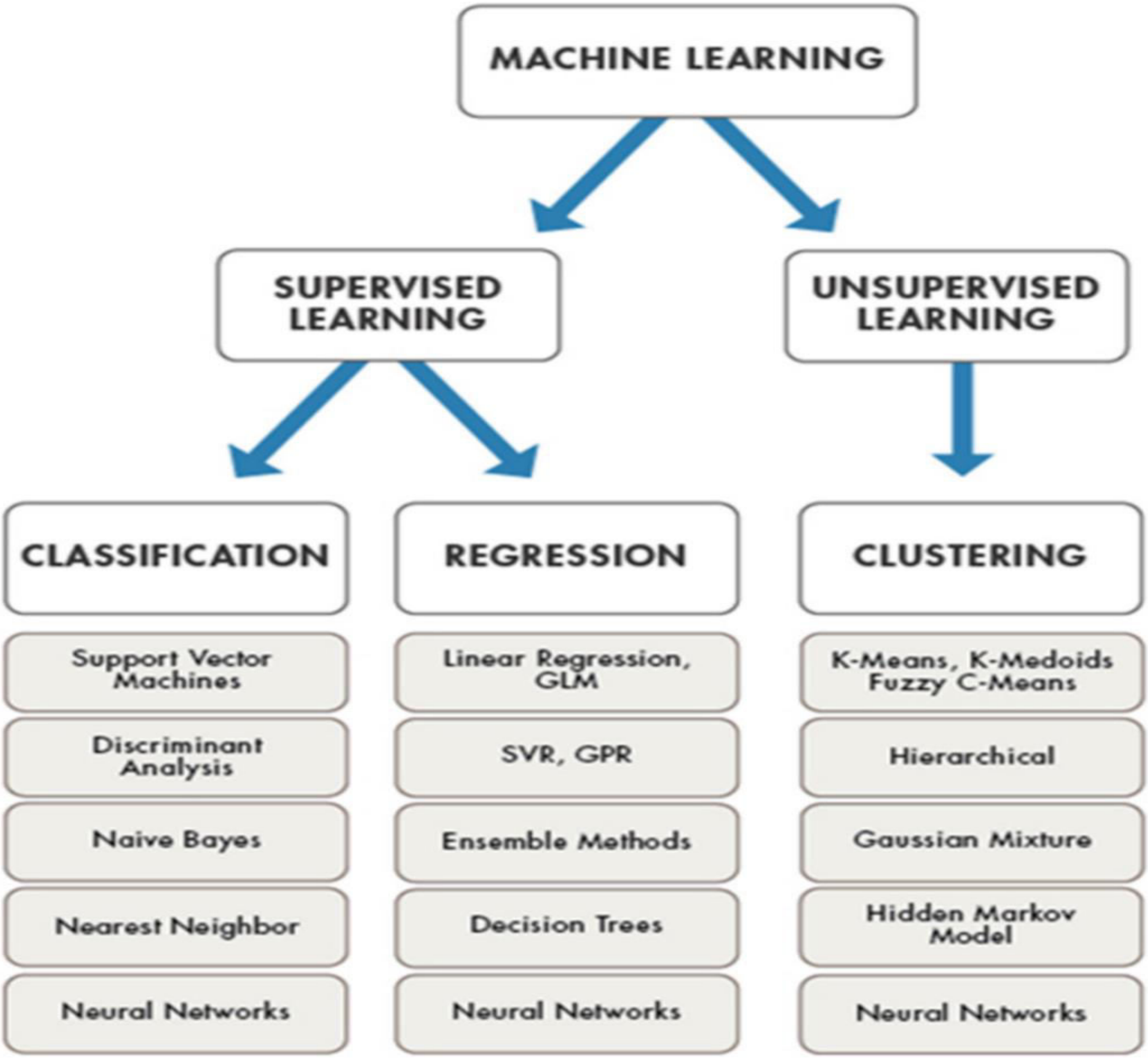
1. Clean the data obtained from the dataset
2. Select a proper algorithm for building a prediction model
3. Train your model to understand the pattern of project
4. Predict your results with higher accuracy

Most ML algorithms are broadly categorized as being either supervised or unsupervised. The fundamental difference between supervised and unsupervised learning algorithms is how they deal with data. Two other categories are semi-supervised and reinforcement algorithms.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205161 |



Supervised algorithms

These algorithms deal with clearly labeled data, with direct oversight by a data scientist. They have both input data and desired output data provided for them through labeling.

Supervised algorithms typically serve two purposes: classification or regression. In classification problems, an algorithm can accurately assign different data into specific categories – such as *dogs* and *cats* -- which becomes feasible with labeled data.

There are many real-world use cases for supervised algorithms, including healthcare and medical diagnoses, as well as image recognition. In both cases, classification of data is needed.

In regression problems, an algorithm is used to predict the probability of an event taking place – known as the *dependent variable* -- based on prior insights and observations from training data -- the independent variables. A use case for regression algorithms might include time series forecasting used in sales.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205161 |

Unsupervised data

Unsupervised algorithms deal with unclassified and unlabeled data. As a result, they operate differently from supervised algorithms. For example, clustering algorithms are a type of unsupervised algorithm used to group unsorted data according to similarities and differences, given the lack of labels.

Unsupervised algorithms can also be used to identify associations, or interesting connections and relationships, among elements in a data set. For example, these algorithms can infer that one group of individuals who buy a certain product also buy certain other products.

Semi-supervised algorithms

However, many machine learning techniques can be more accurately described as semi-supervised, where both labeled and unlabeled data are used.

Reinforcement algorithms

Reinforcement algorithms – which use reinforcement learning techniques-- are considered a fourth category. They're unique approach is based on rewarding desired behaviors and punishing undesired ones to direct the entity being trained using rewards and penalties.

Types of machine learning algorithms

There are several types of machine learning algorithms, including the following:

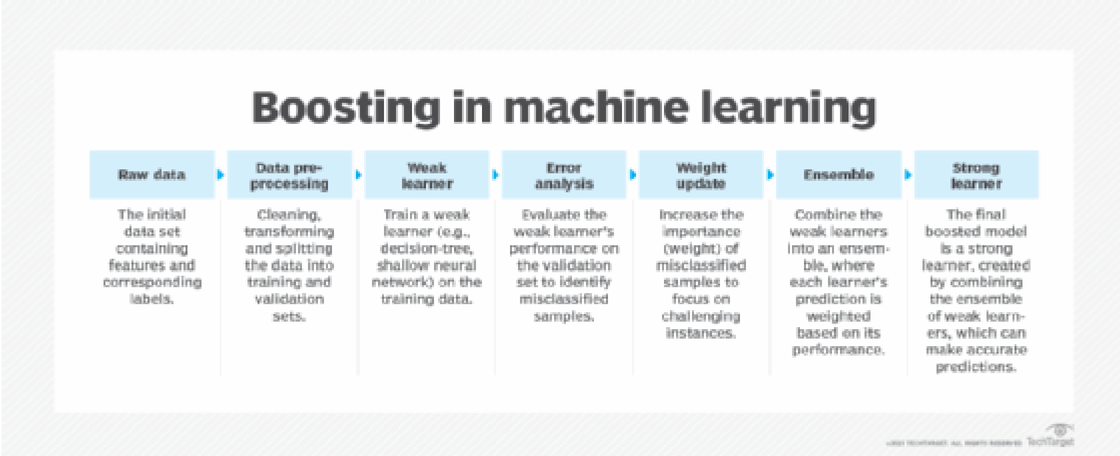
- **Linear regression.** A linear regression algorithm is a supervised algorithm used to predict continuous numerical values that fluctuate or change over time. It can learn to accurately predict variables like age or sales numbers over a period of time.
- **Logistic regression.** In predictive analytics, a machine learning algorithm is typically part of a predictive modeling that uses previous insights and observations to predict the probability of future events. Logistic regressions are also supervised algorithms that focus on binary classifications as outcomes, such as "yes" or "no."
- **Decision tree.** This is a supervised learning algorithm used for both classification and regression problems. Decision trees divide data sets into different subsets using a series of questions or conditions that determine which subset each data element belongs in. When mapped out, data appears to be divided into branches, hence the use of the word *tree*.
- **Support vector machine.** SVMs are used for classification, regression and anomaly detection in data. An SVM is best applied to binary classifications, where elements from a data set are classified into two distinct groups.
- **Naïve Bayes.** This algorithm performs classifications and makes predictions. However, it's one of the simplest supervised learning algorithms and assumes that all features in the input data are independent of one another; one data point won't affect another when making predictions.
- **Random forest.** These algorithms combine multiple unrelated decision trees of data, organizing and labeling data using regression and classification methods.
- **K-means.** This unsupervised learning algorithm identifies groups of data within unlabeled data sets. It groups the unlabeled data into different clusters; it's one of the most popular clustering algorithms.
- **K-nearest neighbors.** KNNs classify data elements through proximity or similarity. An existing data group that most closely resembles a new data element is the one that element will be grouped with.
- **Artificial neural networks.** ANNs, or simply neural networks, are groups of algorithms that recognize patterns in input data using building blocks called *neurons*. These neurons loosely resemble neurons in the human brain. They're trained and modified over time through supervised training methods.
- **Dimensionality reduction.** When a data set has a high number of features, it's said to have high dimensionality. Dimensionality reduction refers to stripping down the number of features so that only the most meaningful insights or information remain. An example of this method is principal component analysis.
- **Gradient boosting.** This optimization algorithm reduces a neural network's cost function, which is a measure of the size of the error the network produces when its actual output deviates from its intended output.
- **AdaBoost.** Also called *adaptive boosting*, this supervised learning technique boosts the performance of an underperforming ML classification or regression algorithm by combining it with weaker ones to form a stronger algorithm that produces fewer errors.

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205161 |



II. METHODOLOGY

Step 1: Data collection

The first step in the machine learning process is data collection. Data is the lifeblood of machine learning - the quality and quantity of your data can directly impact your model's performance. Data can be collected from various sources such as databases, text files, images, audio files, or even scraped from the web.

Once collected, the data needs to be prepared for machine learning. This process involves organizing the data in a suitable format, such as a CSV file or a database, and ensuring that the data is relevant to the problem you're trying to solve.

Step 2: Data preprocessing

Data preprocessing is a crucial step in the machine learning process. It involves cleaning the data (removing duplicates, correcting errors), handling missing data (either by removing it or filling it in), and normalizing the data (scaling the data to a standard format).

Preprocessing improves the quality of your data and ensures that your machine learning model can interpret it correctly. This step can significantly improve the accuracy of your model. Our course, Preprocessing for Machine Learning in Python, explores how to get your cleaned data ready for modeling.

Step 3: Choosing the right model

Once the data is prepared, the next step is to choose a machine learning model. There are many types of models to choose from, including linear regression, decision trees, and neural networks. The choice of model depends on the nature of your data and the problem you're trying to solve.

Factors to consider when choosing a model include the size and type of your data, the complexity of the problem, and the computational resources available.

Step 4: Training the model

After choosing a model, the next step is to train it using the prepared data. Training involves feeding the data into the model and allowing it to adjust its internal parameters to better predict the output.During training, it's important to avoid overfitting (where the model performs well on the training data but poorly on new data) and underfitting (where the model performs poorly on both the training data and new data).

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205161 |

Step 5: Evaluating the model

Once the model is trained, it's important to evaluate its performance before deploying it. This involves testing the model on new data it hasn't seen during training.

Common metrics for evaluating a model's performance include accuracy (for classification problems), precision and recall (for binary classification problems), and mean squared error (for regression problems).

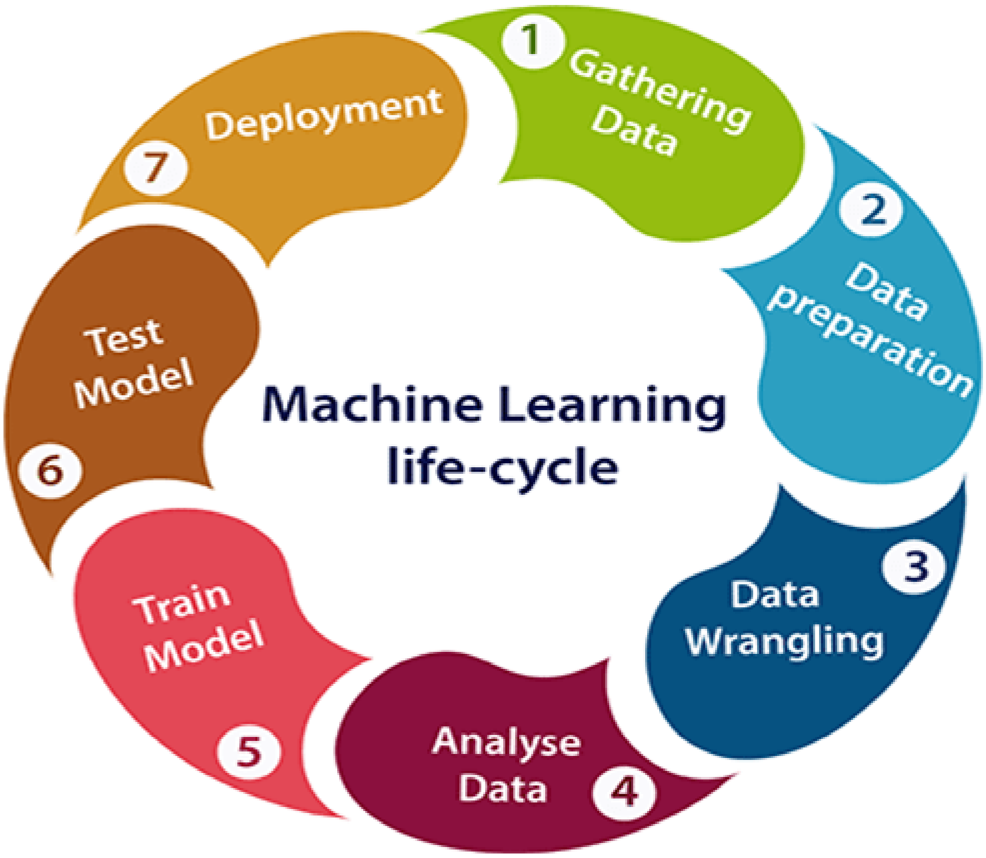
Step 6: Hyperparameter tuning and optimization

After evaluating the model, you may need to adjust its hyperparameters to improve its performance. This process is known as parameter tuning or hyperparameter optimization.

Techniques for hyperparameter tuning include grid search (where you try out different combinations of parameters) and cross validation (where you divide your data into subsets and train your model on each subset to ensure it performs well on different data).

Step 7: Predictions and deployment

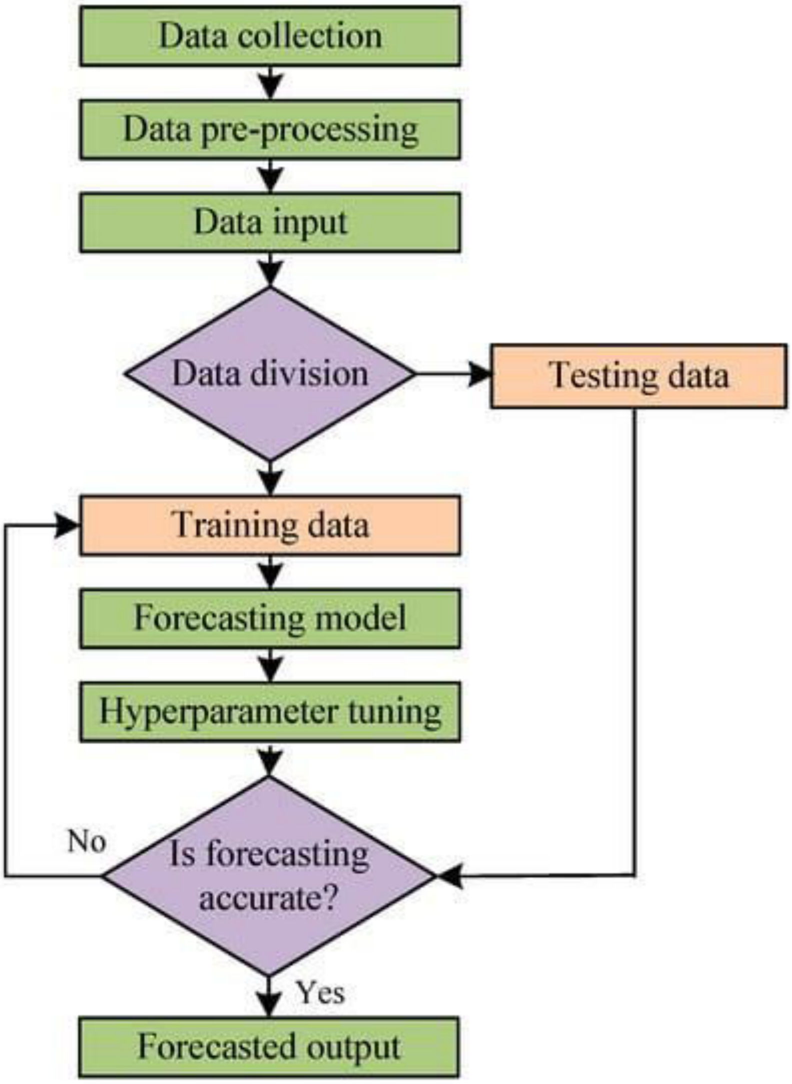
Once the model is trained and optimized, it's ready to make predictions on new data. This process involves feeding new data into the model and using the model's output for decision-making or further analysis.
Deploying the model involves integrating it into a production environment where it can process real-world data and provide real-time insights. This process is often known as MLOps.



| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205161 |



III. CONCLUSION AND FUTURE SCOPE

Forecasting for all types of loads has been widely discussed, including very short-term, short-term, medium-term, and long-term. A vast majority of forecasting models use smart meters to gather electricity load consumption data. In smart grids, these smart meters can be installed in a variety of locations depending on the data requirements. Researchers have shown interest in forecasting individual consumer loads and then summing them up for regional forecasts based on smart meter information. Many researchers directly use smart meter data from smart meters installed at the distribution end to forecast load region-by-region.

Generally, the forecast models are of two types: parametric and non-parametric. The parametric methods work with linear data forms, whereas non-parametric methods use non-linear data and are based on artificial intelligence. It is also observed that, apart from historical load data from smart meters, many techniques consider various inputs to their models, including weather data (temperature, humidity, wind speed, solar radiations, etc.) and time horizons. To determine the accuracy of the system, key performance indicators are used for each model.

Considering the present scenario of increased electricity demand, smart meters are a must for the smooth operation and handling of future smart grid. With such a facility, load forecasting should also be easier and accurate, to handle the future energy market. In addition, it will be possible to upload large and authentic datasets generated by smart meters directly to a server so that they can be analyzed more precisely.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205161 |

REFERENCES

[1] B. Zhao, L. Zeng, B. Li, Y. Sun, Z. Wang, M. Shahzad, and P. Xi, ''Collaborative control of thermostatically controlled appliances for balancing renewable generation in smart grid,'' IEEJ Trans. Electr. Electron. Eng., vol. 15, no. 3, pp. 460–468, Mar. 2020.
[2] K. Berk, A. Hoffmann, and A. Müller, ''Probabilistic forecasting of industrial electricity load with regime switching behavior,'' Int. J. Forecasting, vol. 34, no. 2, pp. 147–162, Apr. 2018.
[3] J. R. Cancelo, A. Espasa, and R. Grafe, ''Forecasting the electricity load from one day to one week ahead for the Spanish system operator,'' Int. J. Forecasting, vol. 24, no. 4, pp. 588–602, 2008.
[4] M. Djukanovic, S. Ruzic, B. Babic, D. J. Sobajic, and Y. H. Pao, ''A neural net based short term load forecasting using moving window procedure,'' Int. J. Electr. Power Energy Syst., vol. 17, no. 6, pp. 391–397, Dec. 1995.
[5] L. J. Soares and M. C. Medeiros, ''Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data,'' Int. J. Forecasting, vol. 24, no. 4, pp. 630–644, Oct. 2008.
[6] Lowlesh Yadav and Asha Ambhaikar, "IOHT based Tele-Healthcare Support System for Feasibility and performance analysis," Journal of Electrical Systems, vol. 20, no. 3s, pp. 844–850, Apr. 2024, doi: 10.52783/jes.1382.
[7] L. Yadav and A. Ambhaikar, "Feasibility and Deployment Challenges of Data Analysis in Tele-Healthcare System," 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), Raipur, India, 2023, pp. 1-5, doi: 10.1109/ICAIIHI57871.2023.10489389.
[8] L. Yadav and A. Ambhaikar, "Approach Towards Development of Portable Multi-Model Tele-Healthcare System," 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), Raipur, India, 2023, pp. 1-6, doi: 10.1109/ICAIIHI57871.2023.10489468.
[9] Lowlesh Yadav and Asha Ambhaikar, Exploring Portable Multi-Modal Telehealth Solutions: A Development Approach. International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), vol. 11, no. 10, pp. 873–879, Mar. 2024.11(10), 873–879, DOI: 10.13140/RG.2.2.15400.99846.
[10] Lowlesh Yadav, Predictive Acknowledgement using TRE System to reduce cost and Bandwidth, March 2019. International Journal of Research in Electronics and Computer Engineering (IJRECE), VOL. 7 ISSUE 1 (JANUARY-MARCH 2019) ISSN: 2393-9028 (PRINT) | ISSN: 2348-2281 (ONLINE).