IBM NAAN MUDHALVAN

ELECTRICITY PRICES

PREDICTION

|  |  |
| --- | --- |
| DOMAIN | APPLIED DATA SCIENCE |
| PROJECT TOPIC | ELECTRICITY PRICES PREDICTION |
| TEAM MEMBERS&  REGISTER NUMBER | ARAVINDHATCHAN S - 420421104007  ANBUMANI P - 420421104006  VIGNESHWAR T - 420421104086  KOLLURU LAKSHMI MANI - 42042110437 |

**Introduction:**

* Predicting electricity prices is a complex task that involves numerous factors, including supply and demand dynamics, weather conditions, generation sources, infrastructure constraints, and market regulations. Accurate predictions are essential for energy market participants to make informed decisions and manage costs effectively. Here's an overview of the key steps and approaches involved in electricity price prediction:

**Problem Statement:**

* Create a predictive model that utilizes historical electricity prices and relevant factors to forecast future electricity prices, assisting energy providers and consumers in making informed decisions regarding consumption and investment.
* The objective is to create a tool that assists both energy providers and consumers in making informed decisions regarding consumption and investment by predicting future electricity prices.

**Design Thinking:**

1. **Data Source:**

* Utilize a dataset containing historical electricity prices and relevant factors like date, demand, supply, weather conditions, and economic indicators.

1. **Data Preprocessing**:

* Clean and preprocess the data, handle missing values, and convert categorical features into numerical representations.

1. **Feature Engineering**:

* Create additional features that could enhance the predictive power of the model, such as time-based features and lagged variables.

1. **Model Selection**:

* Choose suitable time series forecasting algorithms (e.g., ARIMA, LSTM) for predicting future electricity prices.

1. **Model Training**:

* Train the selected model using the preprocessed data.

1. **Evaluation**:

* Evaluate the model's performance using appropriate time series forecasting metrics (e.g., Mean Absolute Error, Root Mean Squared Error).

**Phases of Devolopment:**

1. **Data Collection**:
   * Gather historical data on electricity prices, typically at various time intervals (e.g., hourly, daily) and for specific geographical regions.
   * Collect relevant auxiliary data, such as weather information (temperature, humidity, wind speed), fuel prices, power generation data, and market fundamentals (demand forecasts, generation capacities).
   * The Data for the problem statement is collected from [**https://www.kaggle.com/datasets/chakradharmattapalli/electricity-price-prediction**](https://www.kaggle.com/datasets/chakradharmattapalli/electricity-price-prediction)
2. **Data Preprocessing**:
   * Clean and preprocess the data to handle missing values, outliers, and inconsistencies.
   * Normalize or scale the data to ensure that different features have a comparable impact on the prediction model.
   * Convert timestamps into appropriate formats for time-series analysis.
3. **Feature Engineering**:
   * Identify relevant features that may influence electricity prices, such as time of day, day of the week, holidays, and market-specific indicators.
   * Incorporate lagged values of electricity prices and auxiliary variables to capture temporal dependencies.
4. **Model Selection**:
   * Choose an appropriate predictive modeling technique. Common approaches include time series analysis, machine learning, and statistical models.
   * Popular models for electricity price prediction include autoregressive integrated moving average (ARIMA), seasonal decomposition of time series (STL), long short-term memory (LSTM) neural networks, and regression models.
5. **Training and Validation**:
   * Split the historical data into training and validation sets for model development and evaluation.
   * Use appropriate evaluation metrics, such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE), to assess the model's performance.
6. **Model Tuning**:
   * Fine-tune model hyperparameters to improve prediction accuracy. Grid search or random search can help find the best hyperparameter settings.
7. **Prediction Horizon**:
   * Determine the time horizon for predictions (e.g., hourly, daily, weekly) based on the specific needs of stakeholders.
8. **Incorporate External Factors**:
   * Include external factors that can impact electricity prices, such as changes in energy policies, infrastructure upgrades, or extreme weather events.
9. **Real-Time Updates**:
   * For operational use, implement a system that continuously updates the prediction model with new data to adapt to changing market conditions.
10. **Model Deployment:**
    * Deploy the trained model in a production environment where it can generate real-time or future electricity price forecasts.
    * Make predictions available to relevant stakeholders through dashboards, APIs, or other communication channels.
11. **Monitoring and Evaluation**:
    * Continuously monitor the model's performance in real-world conditions and retrain it periodically to maintain accuracy.
    * Consider incorporating feedback loops to improve model performance based on observed errors.
12. **Interpretability and Communication**:
    * Ensure that the results are interpretable and provide insights into the factors driving electricity price fluctuations.
    * Communicate predictions and uncertainties effectively to decision-makers and end-users.
13. **Compliance and Regulation**:
    * Ensure that the prediction model complies with regulatory requirements and market rules, especially in regulated energy markets.
14. **Risk Management**:
    * Assess the financial and operational risks associated with electricity price predictions and implement risk mitigation strategies as needed.

**Steps To Be Followed For The Analysis:**

**STEP 1** - Collect the dataset of ELECTRICITY PRICES PREDICTION. We have collected it from

[**https://www.kaggle.com/datasets/chakradharmattapalli/electricity-price-prediction**](https://www.kaggle.com/datasets/chakradharmattapalli/electricity-price-prediction)

**STEP 2** - Preprocess the data and transform it according to the analysis

**STEP 3** - Remove the outliers, null values and other error data

**STEP 4** - Fit the preprocessed data into a model for predictions

**STEP 5** - Find the prediction score using r2\_score, accuracy\_score

**STEP 6** - Use the preprocessed data for visualizations and other summarization of data given

**STEP 7** - Derive the insights from the visualizations made and make it as a report.

**About The Dataset:**

[**https://www.kaggle.com/datasets/chakradharmattapalli/electricity-price-prediction**](https://www.kaggle.com/datasets/chakradharmattapalli/electricity-price-prediction)

* The price of electricity depends on many factors. Predicting the price of electricity helps many businesses understand how much electricity they have to pay each year. The Electricity Price Prediction task is based on a case study where you need to predict the daily price of electricity based on the daily consumption of heavy machinery used by businesses.
* You do not know the actual cost of the electricity consumed by the machines throughout the day, but the organization has provided you with historical data of the price of the electricity consumed by the machines. Below is the information of the data we have for the task of forecasting electricity prices.

COLOUMNS THAT ARE USED:

* **DateTime:** Date and time of the record
* **Holiday:** contains the name of the holiday if the day is a national holiday
* **HolidayFlag:** contains 1 if it’s a bank holiday otherwise 0
* **DayOfWeek:** contains values between 0-6 where 0 is Monday
* **WeekOfYear:** week of the year
* **Day:** Day of the date
* **Month:** Month of the date
* **Year:** Year of the date
* **PeriodOfDay:** half-hour period of the day
* **ForcastWindProduction:** forecasted wind production
* **SystemLoadEA :**forecasted national load
* **SMPEA:** forecasted price
* **ORKTemperature:** actual temperature measured
* **ORKWindspeed:** actual windspeed measured
* **CO2Intensity:** actual C02 intensity for the electricity produced
* **ActualWindProduction:** actual wind energy production
* **SystemLoadEP2:** actual national system load
* **SMPEP2:** the actual price of the electricity consumed (labels or values to be predicted)

Gh

Predicting electricity prices involves analyzing various factors and data to estimate future price trends accurately. Here are some essential features and explanations of the factors that can influence electricity prices:

1**. Supply and Demand:** The fundamental factor in electricity price prediction is the balance between supply and demand. When demand exceeds supply, prices tend to rise, and when supply exceeds demand, prices may fall. Factors influencing these include:

* **Weather Conditions:** Extreme weather conditions can increase electricity demand (e.g., for heating or cooling) and affect supply (e.g., by disrupting power generation from renewable sources).
* **Seasonal Variations**: Electricity demand can vary significantly by season, with peak demand often occurring in the summer and winter.
* **Economic Conditions**: Economic growth or recession can impact industrial and commercial electricity consumption.

2. **Energy Source Mix:** The mix of energy sources used for power generation can influence electricity prices. Different sources have varying production costs and supply reliability:

* **Renewable Energy**: An increase in renewable energy sources (e.g., wind and solar) can lower electricity prices due to their low operating costs when available.
* **Fossil Fuels**: The prices of fossil fuels, like natural gas and coal, can significantly impact electricity prices, as they are used in many power plants.

3. **Generation and Transmission Costs**: The cost of generating electricity and transmitting it through the grid affects prices. Factors include:

* **Fuel Prices**: Fluctuations in fuel prices, such as natural gas or coal, directly affect the operating costs of power plants.
* **Infrastructure Investment**: Costs associated with building and maintaining power generation and transmission infrastructure can influence prices.

4. **Regulatory Policies:** Government regulations and policies can have a substantial impact on electricity prices:

* **Energy Market Structure:** Deregulated markets tend to have more volatile prices, as they are influenced by market forces, while regulated markets may have more stable prices but with potentially less flexibility.
* **Environmental Regulation**: Environmental regulations can increase the cost of power generation, especially for older, less efficient power plants.

5. **Energy Imports and Exports:** Electricity markets can be interconnected, allowing for the import and export of electricity. This can affect prices if neighboring regions have different supply and demand dynamics.

6. **Electricity Storage**: The availability and cost of energy storage technologies can influence prices by enabling more efficient use of intermittent renewable energy sources.

7.**Consumer Behavior**: Consumer choices, such as energy conservation and the adoption of energy-efficient technologies, can impact demand and, consequently, prices.

8. **Market Data:** Historical market data, such as past prices and trading volumes, can be used for time-series analysis and modeling to predict future prices.

9. **Market Sentiment and News**: External events and news, like geopolitical developments or natural disasters, can create uncertainty and affect investor sentiment, leading to price fluctuations.

10. **Renewable Energy Forecasting:** Accurate predictions of renewable energy generation, such as wind and solar, are crucial for anticipating supply variations and their impact on prices.

To predict electricity prices, data from these features should be collected, analyzed, and used to develop forecasting models. Common techniques include time-series analysis, machine learning, and econometric models. Accurate and timely data, along with robust modeling methods, are essential for making reliable predictions in the electricity market.

**DATA PREPROCESSING:**

* **Data Cleaning:**
* **Handling Missing Data**
* **Outlier Detection and Treatment**
* **Data Transformation:**
* **Normalization and Scaling**
* **Encoding Categorical Variables**
* **Feature Engineering**
* **Logarithm or Power Transformation**
* **Data Reduction:**
* **Dimensionality Reduction**
* **Sampling**
* **Data Splitting**
* **Handling Time Series Data**
* **Dealing with Imbalanced Data**
* **Text Data Preprocessing**
* **Data Scaling and Standardization**
* **Handling Time-Dependent Data**
* **Handling Highly Correlated Features**
* **Data Formatting.**

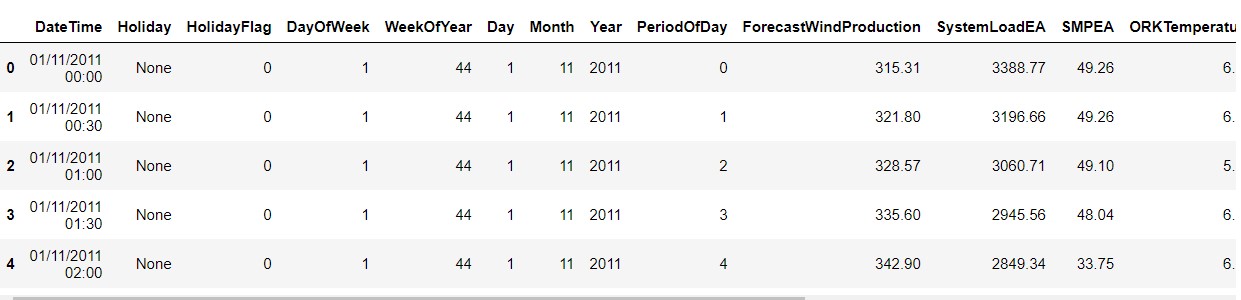
**IMPORT LIBRARIES:**

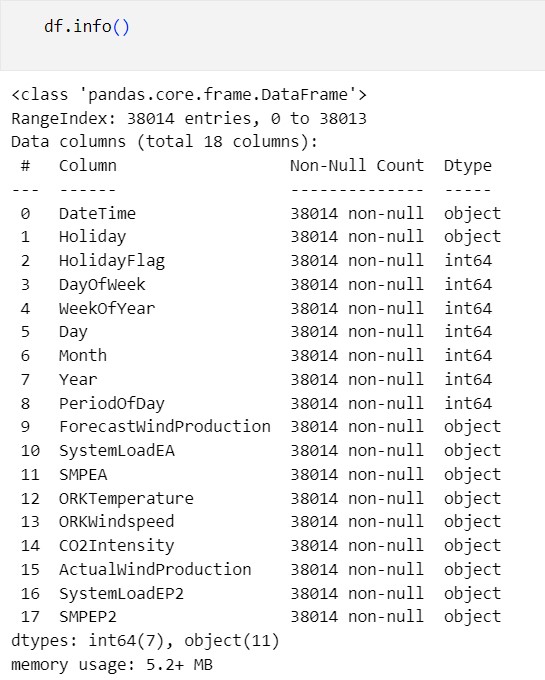
* Numpy
* Pandas
* Matplotlib
* Seaborn

**READING FILE:**

|  |
| --- |
| df=pd.read\_csv("Electricity.csv", low\_memory=False)  df.head() |

**OUTPUT:**



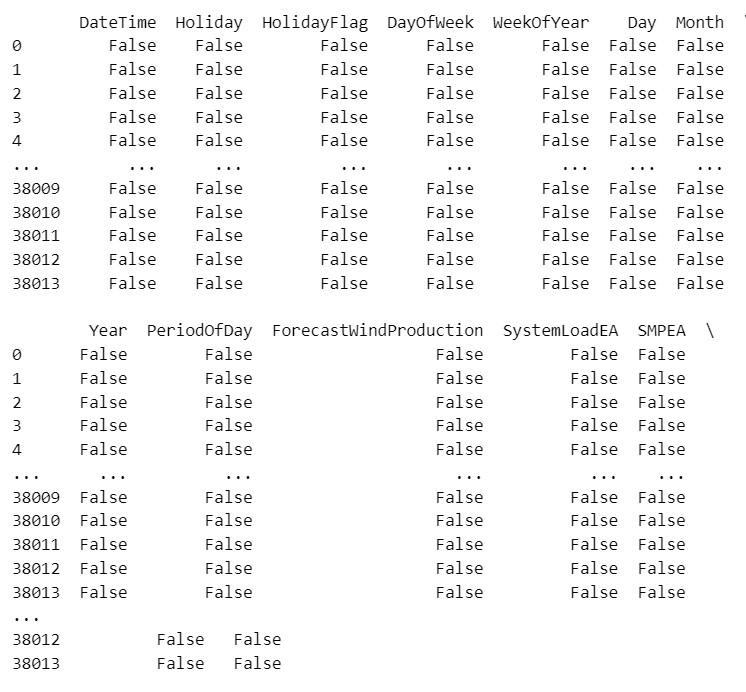




**HANDLING DATASET:**

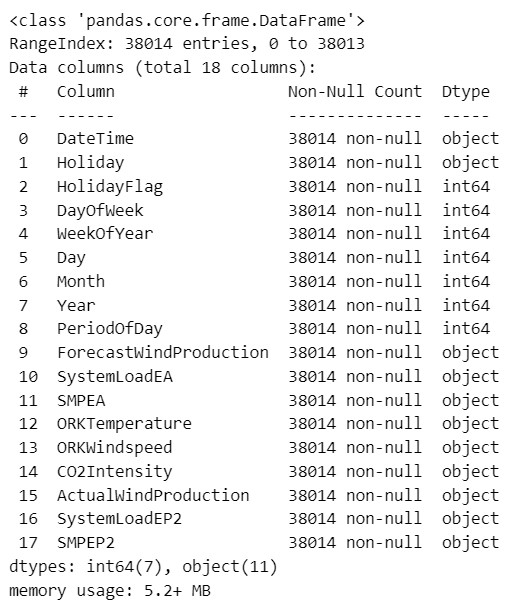
|  |
| --- |
| ##Missing Data  import pandas as pd  # Load your data into a DataFrame  df = pd.read\_csv("Electricity.csv")  # Check for missing data  missing\_data = df.isnull()  print(missing\_data)  # Remove rows with missing data  df.dropna(axis=0, inplace=True)  # Remove columns with missing data  df.dropna(axis=1, inplace=True) |

**OUTPUT:**

****

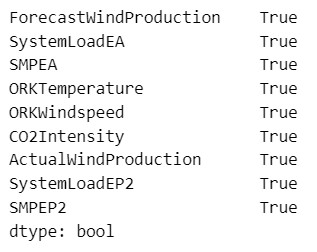
|  |
| --- |
| df.info() |

**OUTPUT:**

****

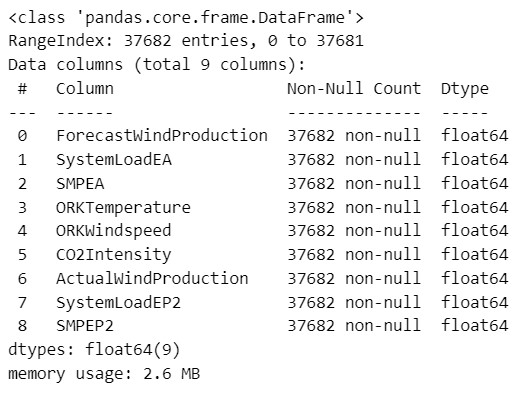
|  |
| --- |
| data.isin(['?']).any() |

**OUTPUT:**

****

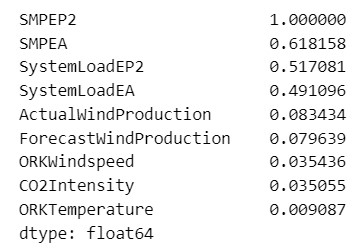
|  |
| --- |
| data=data.apply(pd.to\_numeric)  data=data.reset\_index()  data.drop('index', axis=1, inplace=True)  data.info() |

**OUTPUT:**



|  |
| --- |
| data.corrwith(data['SMPEP2']).abs().sort\_values(ascending=False) |

**OUTPUT:**



**MACHINE LEARNING ALGORITHM:**

Machine learning algorithms are computational methods that enable computers to automatically learn and make predictions or decisions based on data. These algorithms are at the core of many artificial intelligence and data analysis applications. Here are some of the most commonly used machine learning algorithms, categorized by their primary use cases.

TYPES:

1. **Supervised Learning**:
   * **Linear Regression**: Used for regression tasks, it models the relationship between a dependent variable and one or more independent variables.
   * **Logistic Regression**: Primarily used for binary classification problems, it models the probability of an instance belonging to a particular class.
   * **Decision Trees**: A tree-like structure used for classification and regression tasks, making decisions by splitting data based on feature conditions.
   * **Random Forest**: An ensemble method that combines multiple decision trees to improve predictive accuracy and reduce overfitting.
   * **Support Vector Machines (SVM)**: Effective for both classification and regression, SVM finds the optimal hyperplane that best separates data into classes.
   * **K-Nearest Neighbors (K-NN)**: A non-parametric algorithm that classifies data points based on the majority class of their k-nearest neighbors.
   * **Naive Bayes**: A probabilistic algorithm commonly used for text classification and spam filtering.
   * **Gradient Boosting**: A powerful ensemble method that combines weak learners to create a strong predictive model.
2. **Unsupervised Learning**:
   * **K-Means Clustering**: Used for grouping similar data points into clusters based on distance metrics.
   * **Hierarchical Clustering**: Builds a hierarchy of clusters, which can be visualized as a tree-like structure (dendrogram).
   * **Principal Component Analysis (PCA)**: A dimensionality reduction technique that identifies the most important features in the data.
   * **Autoencoders**: Neural networks used for unsupervised feature learning, often employed in anomaly detection and data compression.
   * **DBSCAN**: A density-based clustering algorithm that can find irregularly shaped clusters in data.
   * **t-Distributed Stochastic Neighbor Embedding (t-SNE)**: Used for data visualization and reducing high-dimensional data to lower dimensions while preserving similarities.
3. **Semi-Supervised and Self-Supervised Learning**:
   * **Self-Organizing Maps (SOM)**: Unsupervised learning method for dimensionality reduction and visualization.
   * **Label Propagation**: A semi-supervised learning algorithm that propagates labels from labeled data to unlabeled data based on similarity.
   * **Word2Vec**: A self-supervised learning approach used to learn word embeddings from large text corpora.
4. **Reinforcement Learning**:
   * **Q-Learning**: Used for solving Markov decision processes, such as game playing and robotic control.
   * **Deep Q-Networks (DQN)**: Combines Q-learning with deep neural networks to handle complex tasks, like playing video games.
   * **Policy Gradient Methods**: Directly optimize the policy of an agent to maximize the expected reward.
   * **Proximal Policy Optimization (PPO)**: An actor-critic method that is well-suited for continuous action spaces.

POPULAR MACHINE LEARNING ALGORITHM:

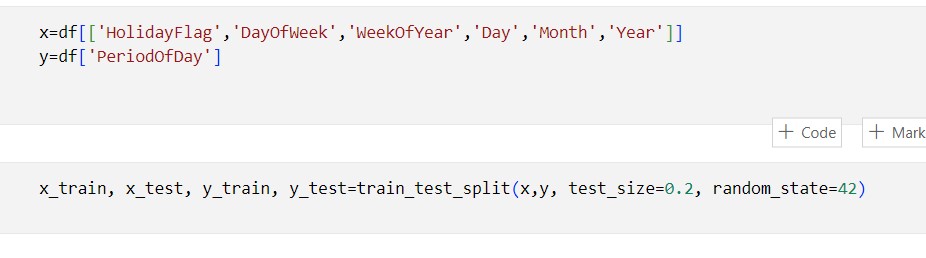
1. Linear Regression Algorithm
2. Logistic Rgression algorithm
3. Decision tree
4. SVM
5. Naïve bayes
6. KNN
7. K-Means Algorithm
8. Random Forest
9. Apriori
10. PCA

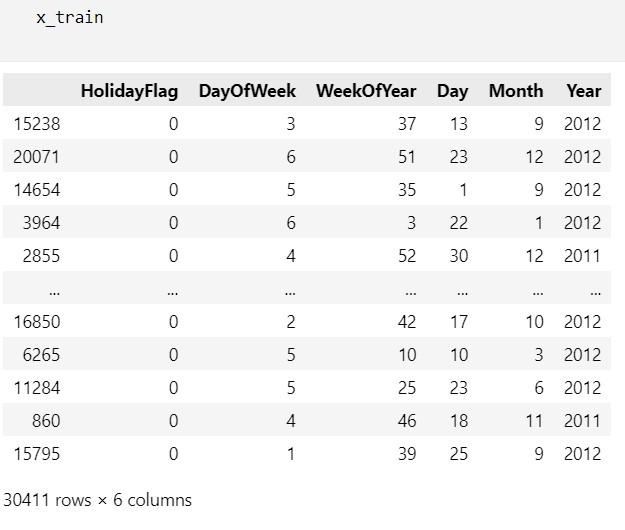
**MODEL BUILDING:**

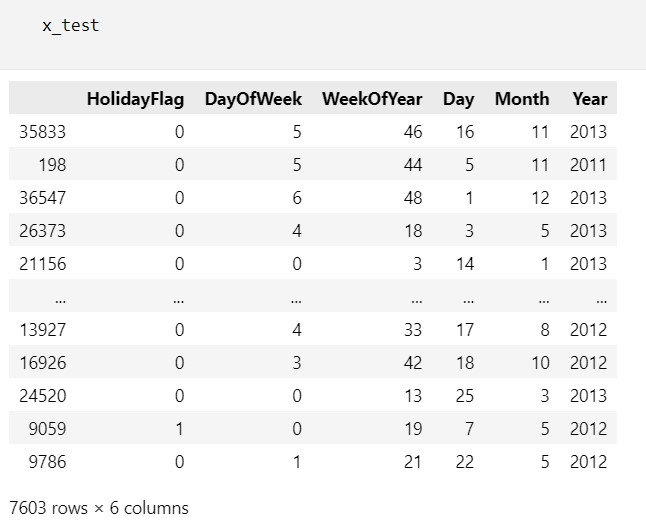
Building a model for a dataset involves a series of steps and considerations. Here's a general outline of the process:

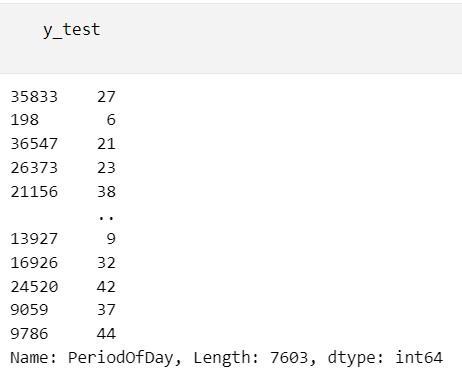
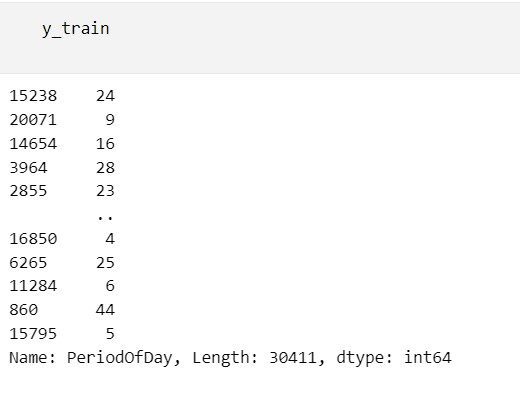
* **Understand the Problem**
* **Data Collection**
* **Data Preprocessing**
* **Feature Engineering**
* **Model selection**
* **Model Training**
* **Model Evaluation**
* **Model Testing**
* **Model Interpretation**
* **Model Deployement**

**PROGRAM:**





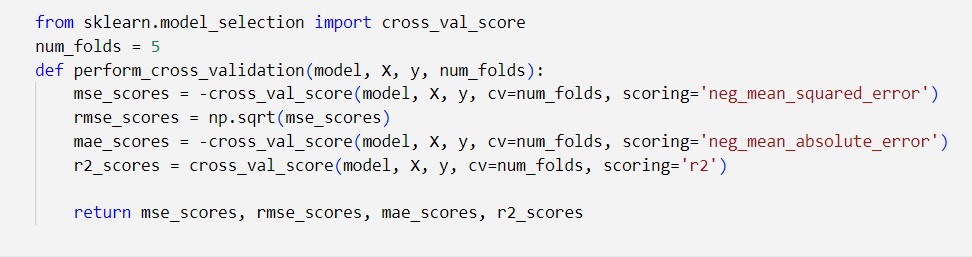




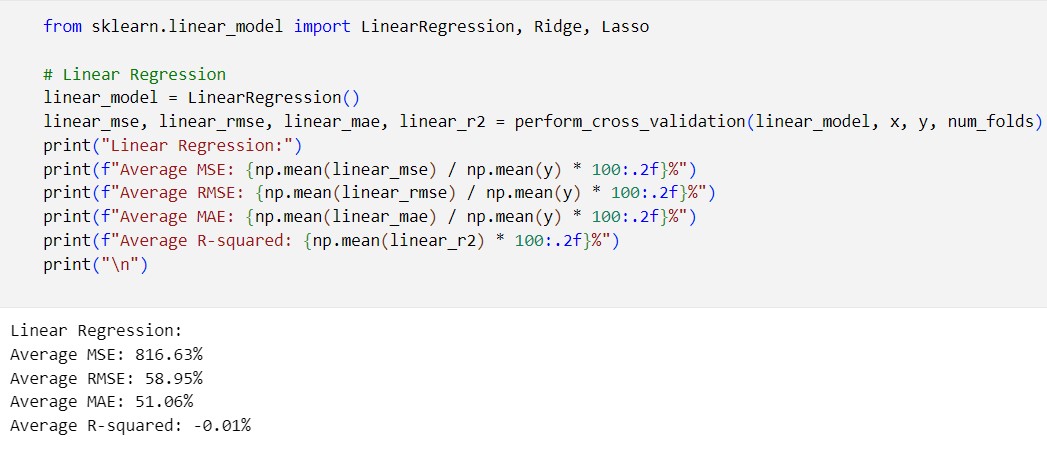
**MODEL EVALUATION:**

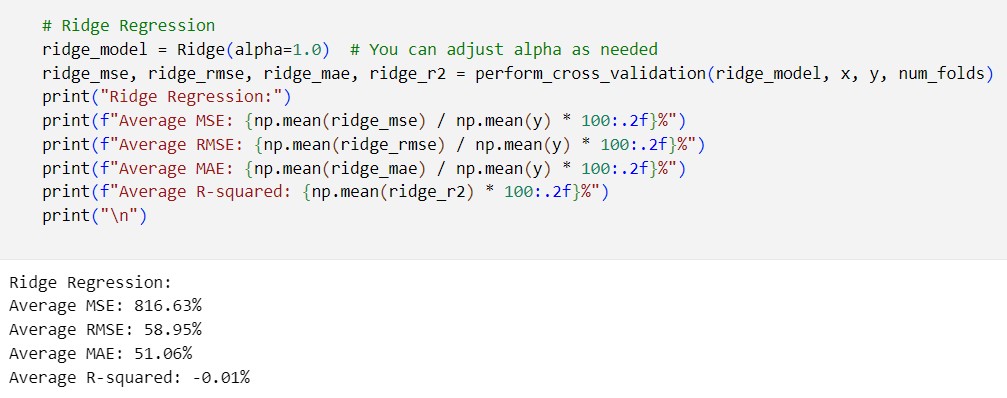
* Model evaluation is a critical step in the machine learning and data analysis process. It involves assessing how well a trained model performs on a given dataset. The goal of model evaluation is to determine the model's effectiveness, generalization capability, and suitability for a specific task. Here are some common techniques and metrics used for model evaluation.
* **Splitting the Data**
* **Training the Model**
* **Model Evaluation Metrics**
* **Classification Problems**
* Accuracy
* Precision, Recall, F1-Score
* ROC AUC
* Confusion Matrix
* **Regression Problems**
* Mean Absolute Error (MAE)
* Mean Squared Error (MSE)
* Root Mean Squared Error (RMSE)
* R-squared (R2)
* **Error Analysis**
* **Deployment and Monitoring**

**PROGRAM:**

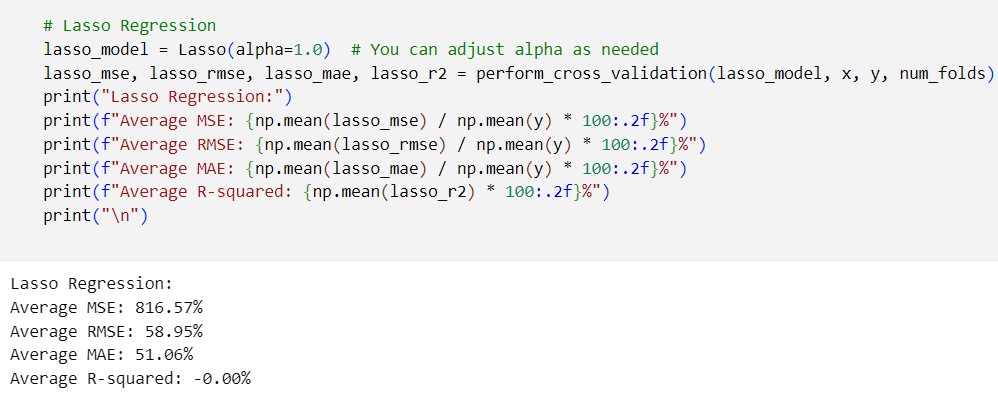


##LINE REGRESSION##

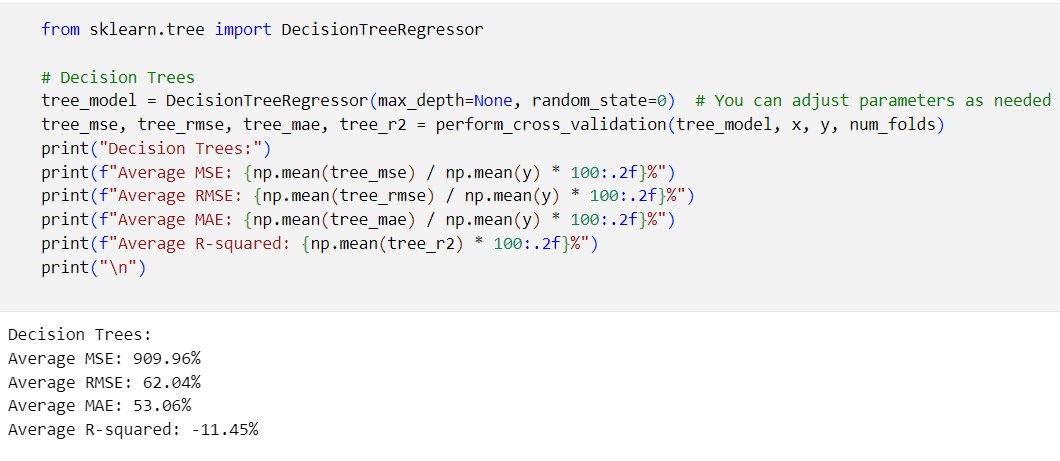
##RIDGE REGRESSION##



##LASSO REGRESSION##



##DECISION TREE##



##RANDOM FOREST##

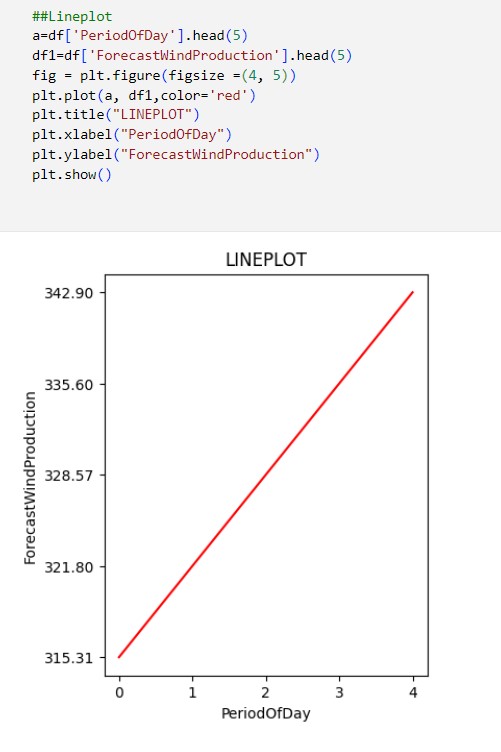


VISUALIZATION:

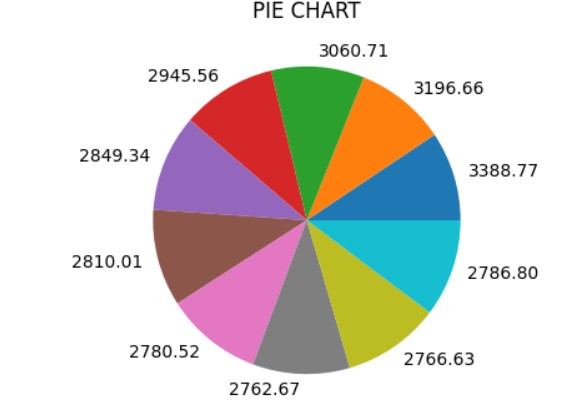
* Data visualization is a powerful way to represent and communicate information from data through visual elements like charts, graphs, and maps. Effective data visualization can make complex data more understandable and can help identify patterns, trends, and insights that might be hidden in raw data. Here are some key concepts and best practices for data visualization.

**PROGRAM:**

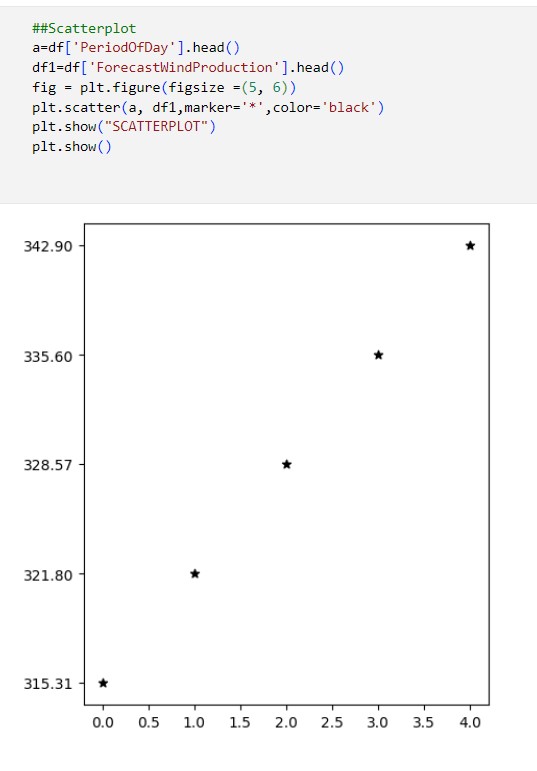
##LINE PLOT##



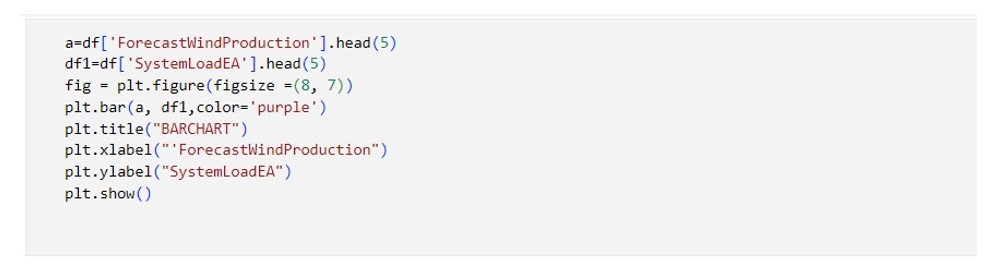
##PIE CHART##



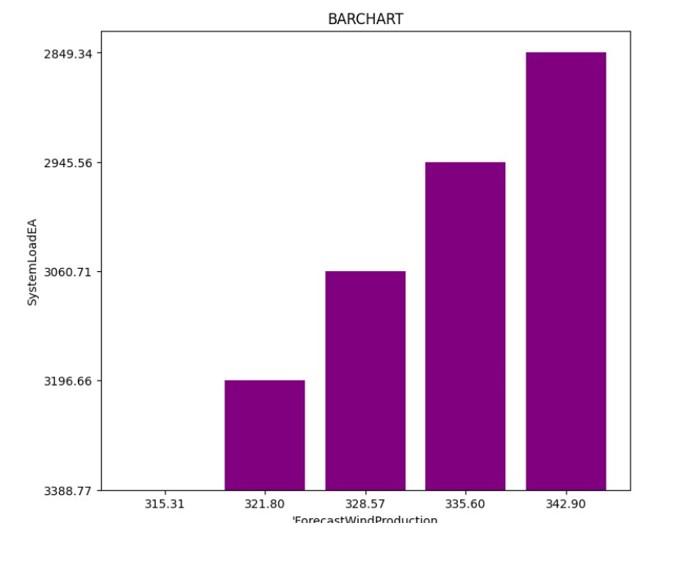
##SCATTER PLOT##



##BAR CHART##



OUTPUT:

****

**INSIGHTS GAINED:**

Predicting electricity prices can yield valuable insights that benefit both consumers and electricity providers. Here are some key insights that can be gained from electricity price prediction:

1. **Optimal Consumption Patterns:**
   * Consumers can adjust their electricity consumption patterns to take advantage of lower-priced periods and reduce costs. For instance, running high-energy appliances during off-peak hours can lead to significant cost savings.
2. **Renewable Energy Integration:**
   * Understanding the relationship between electricity prices and renewable energy generation (e.g., wind and solar) can help in optimizing the use of green energy sources when prices are low.
3. **Demand Response Programs:**
   * Electricity providers can design demand response programs that incentivize consumers to reduce consumption during peak hours, which can help avoid electricity shortages and reduce the need for expensive peak power generation.
4. **Resource Allocation:**
   * Predictions can guide resource allocation decisions, such as investing in additional power generation capacity during periods of expected high demand.
5. **Risk Management:**
   * Electricity market participants, including traders and utilities, can use price predictions for risk management and financial planning. This includes hedging strategies to protect against price volatility.
6. **Grid Management:**
   * Electricity grid operators can better manage grid stability and reliability by predicting price fluctuations. They can prepare for sudden spikes in demand and allocate resources more effectively.
7. **Market Competition:**
   * Electricity price predictions can help market participants anticipate price movements, enhancing competition and efficiency in deregulated energy markets.
8. **Cost Reduction for Consumers:**
   * Consumers can make informed decisions about their electricity plans, potentially choosing lower-cost providers or pricing plans, leading to cost savings.
9. **Environmental Impact:**
   * By predicting low-priced periods with a high share of renewable energy, consumers and providers can reduce carbon emissions by aligning energy consumption with green energy availability.
10. **Investment Decisions:**
    * Investors in the energy sector can use price predictions to guide decisions on infrastructure investments, such as building new power plants or upgrading existing ones.
11. **Emergency Preparedness:**
    * Price predictions can help utilities and government agencies prepare for emergencies and natural disasters by anticipating potential electricity supply and demand disruptions.
12. **Electric Vehicle Charging Optimization:**
    * Electric vehicle owners can use price predictions to determine the most cost-effective times to charge their vehicles, reducing the overall cost of ownership.
13. **Smart Grids:**
    * Predicting electricity prices is a key component of smart grid systems, enabling them to optimize energy distribution and improve efficiency.
14. **Policy Decisions:**
    * Governments and regulatory bodies can use electricity price predictions to formulate energy policies that promote economic growth and environmental sustainability.
15. **Education and Awareness:**
    * Public awareness of electricity price fluctuations can lead to more informed decisions about energy consumption and conservation.

**BENEFITS:**

1. Cost Savings for Consumers
2. Improved Budgeting
3. Energy Efficiency
4. Demand Response
5. Optimal Resource Allocation
6. Renewable Energy Integration
7. Electric Vehicle Charging Optimization

**CONCLUSION:**

* In conclusion, electricity price prediction plays a pivotal role in the energy sector, offering a multitude of benefits to consumers, providers, and governments. By harnessing the power of data analysis and machine learning, stakeholders can gain valuable insights that lead to more informed and cost-effective decisions. These insights enable consumers to optimize their energy consumption, providers to manage resources and demand efficiently, and policymakers to shape sustainable energy policies.
* Electricity price prediction not only helps reduce costs for consumers but also fosters a culture of energy efficiency and environmental responsibility. It empowers the energy industry to adapt to dynamic market conditions, accommodate renewable energy sources, and enhance grid reliability through smart grid technology.
* As we move toward a more sustainable energy future, electricity price prediction remains an indispensable tool for driving progress, reducing waste, and ensuring a reliable and affordable energy supply for all.
* The Dataset has been Model Building,Model Evaluation and Visualization in the phase executed successfully.