**ELECTRICITY PRICES PREDICTION USING APPLIED DATA SCIENCE**

**Phase-5 submission document**

**Project Title:**Electricity Prices Prediction

**Phase 5:** Project documentation & submission

**Topic:**In this section,we will document the complete project and prepare it for submission.



**Introduction**

* Predicting electricity prices is a critical task in the energy sector, as it helps both consumers and producers make informed decisions and manage their energy-related expenses more effectively.
* The fluctuation in electricity prices is influenced by a multitude of factors, including supply and demand dynamics, weather conditions, fuel costs, regulatory policies, and more.
* In this context, the ability to accurately forecast electricity prices is essential for energy market participants, such as utilities, traders, and consumers, to optimize their operations and investments.
* This prediction task involves the use of various data sources, including historical pricing data, weather patterns, market data, and other relevant information. Machine learning and data analytics techniques play a crucial role in developing models that can capture the intricate relationships and patterns that drive electricity price fluctuations.

Here’s a list of tools and software commonly used in the process:

1. **Python:**

Python is a popular programming language for data science and machine learning. You can use various libraries and frameworks for data analysis, including NumPy, Pandas, Scikit-Learn, and TensorFlow or PyTorch for machine learning.

**2.Tableau:**

Tableau is a data visualization tool that can be used to create interactive and visually appealing dashboards to explore and communicate electricity price predictions and trends.

**3.** **Microsoft Excel:**

Excel can be a useful tool for initial data preprocessing and basic analysis before moving on to more advanced data science tools.

**4. Statsmodels:**

This Python library provides various statistical models for time series analysis, including ARIMA and SARIMA models.

4.**SQL Databases:**

Databases like PostgreSQL or MySQL can be used to store and manage large datasets of historical electricity prices, which are essential for model training.

**5.Scikit-Learn:**

This Python library offers a wide range of machine learning algorithms and tools for model training and evaluation.

**6.Hadoop and Spark:**

For processing large datasets, Hadoop and Spark can be employed to distribute data processing tasks.

**7. Weather APIs:**

Access to weather data from sources like the OpenWeatherMap API or government weather agencies can be important for incorporating weather-related features into your electricity price prediction models.

**Dataset:-** <https://www.kaggle.com/datasets/chakradharmattapalli/electricity-price-prediction>

**1.DESIGN THINKING AND PRESENT IN FORM OF DOCUMENT**

**1.** **Empathize:**

Start by understanding the stakeholders and their needs. Conduct interviews, surveys, and workshops with utility companies, consumers, and traders to gain insights into their pain points, goals, and challenges related to electricity pricing.

**2. Define:**

Based on the insights gathered, define a clear problem statement. For example, "How might we provide accurate and timely electricity price forecasts to help consumers make informed decisions on energy consumption?"

**3. Ideate :**

Brainstorm and generate a wide range of ideas for solving the defined problem. Encourage creativity and consider various data sources, modeling techniques, and technologies that can be used for electricity price prediction.

**4. Prototype :**

Develop prototypes or proof-of-concept models to test your ideas. The prototypes should be designed to address the specific needs and pain points of the end-users.

**5. Test:**

Engage with stakeholders to gather feedback on the prototypes. This step involves usability testing, model validation, and assessing the effectiveness of the proposed solutions. Ensure that the electricity price predictions align with the needs and expectations of the end-users.

**6. Refine:**

Based on the feedback received, make improvements to the prototypes and models. Continue iterating until the solution effectively addresses the problem statement and provides accurate electricity price predictions.

**7. Implement:**

Once the solution has been refined and validated, implement it in a real-world setting. This may involve integrating the electricity price prediction model into utility systems, consumer-facing applications, or trading platforms.

**8.Measure Impact:**

Continuously monitor and evaluate the impact of the solution in the field. Collect data on the accuracy of predictions, user satisfaction, and any other relevant metrics. Use this data to make further improvements as needed.

**9. Expand Usage:**

If the solution proves successful, consider scaling it to serve a broader audience or expanding its features and capabilities.

* Throughout the design thinking process, it's essential to maintain a focus on user needs and iterate on your solutions based on feedback and data.
* Additionally, involving cross-functional teams, including data scientists, domain experts, and end-users, is crucial for a holistic and effective design thinking approach to electricity price prediction.

* This approach ensures that the final solution not only provides accurate predictions but also aligns with the goals and expectations of all stakeholders.

**2.DESIGN INTO INNOVATION**

**1. User-Centered Approach:**

Place the end-users at the core of the innovation process. Understand their needs, pain points, and objectives related to electricity price prediction. Tailor your solutions to address these user requirements effectively.

**2. Creative Ideation:**

Encourage a culture of creativity and brainstorming. Invite diverse perspectives and ideas from cross-functional teams. Explore unconventional methods, data sources, and models to generate innovative approaches to predicting electricity prices.

**3. Data Accessibility and Visualization:**

Make data more accessible and comprehensible for users. Develop user-friendly dashboards and visualizations that allow consumers, traders, and utility companies to interact with and understand the data. Innovative data visualization techniques can reveal patterns and insights that might be overlooked in traditional charts.

**4.Predictive Analytics for Renewable Energy:**

Innovate in the area of renewable energy integration. Develop models that can predict electricity prices while accounting for the intermittency of renewable energy sources, such as wind and solar power, to enable better demand-side management.

**5.Demand Response and Energy Efficiency:**

Implement innovative demand response programs and energy efficiency solutions. These can help users adapt their electricity consumption patterns in response to price predictions, reducing costs and promoting sustainability.

**6.Blockchain and Transparency:**

Explore the potential of blockchain technology to increase transparency and trust in electricity pricing. Smart contracts on a blockchain can automate transactions based on price predictions and enable transparent billing.

**7.Dynamic Pricing Models:**

Develop innovative dynamic pricing models that take into account real-time market conditions and customer preferences.

**3.BUILD LOADING AND PREPROCESSING THE DATASET**

**Electricity Price Prediction using Python:**

**# Import necessary libraries**

import numpy as np

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error

import matplotlib.pyplot as plt

data=pd.read\_csv(“<https://www.kaggle.com/datasets/chakradharmattapalli/electricity-price-prediction>”)

# Preprocess the data

data['date'] = pd.to\_datetime(data['date'])

data.set\_index('date', inplace=True)

**# Feature engineering - you can add more features like time of day, day of week, etc.**

data['hour'] = data.index.hour

data['dayofweek'] = data.index.dayofweek

**# Split data into training and testing sets**

X = data[['hour', 'dayofweek']]

y = data['price']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

**# Create and train a linear regression model**

model = LinearRegression()

model.fit(X\_train, y\_train)

**# Make predictions**

y\_pred = model.predict(X\_test)

**# Evaluate the model**

mse = mean\_squared\_error(y\_test, y\_pred)

print(f"Mean Squared Error: {mse}")

**# Plot the actual vs. predicted prices**

plt.figure(figsize=(10, 6))

plt.scatter(X\_test.index, y\_test, label='Actual')

plt.plot(X\_test.index, y\_pred, color='red', label='Predicted')

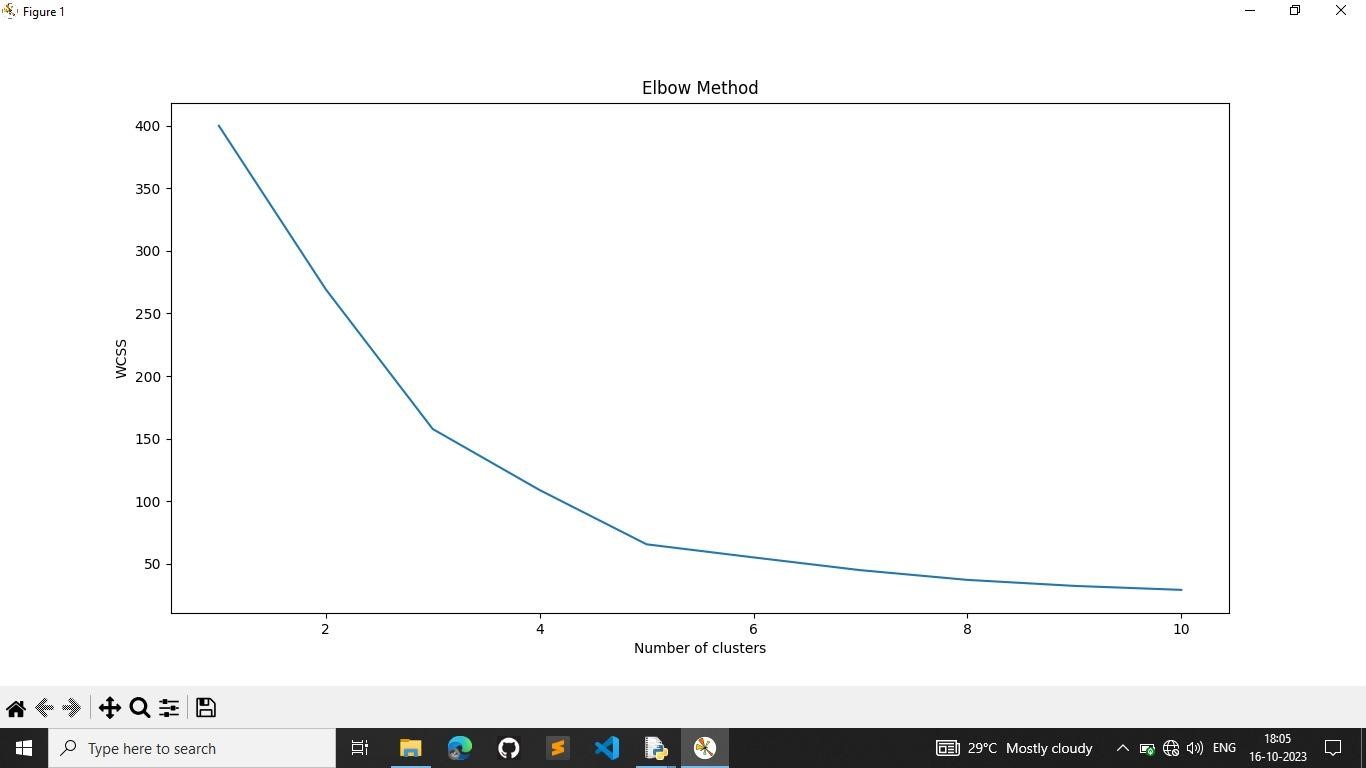
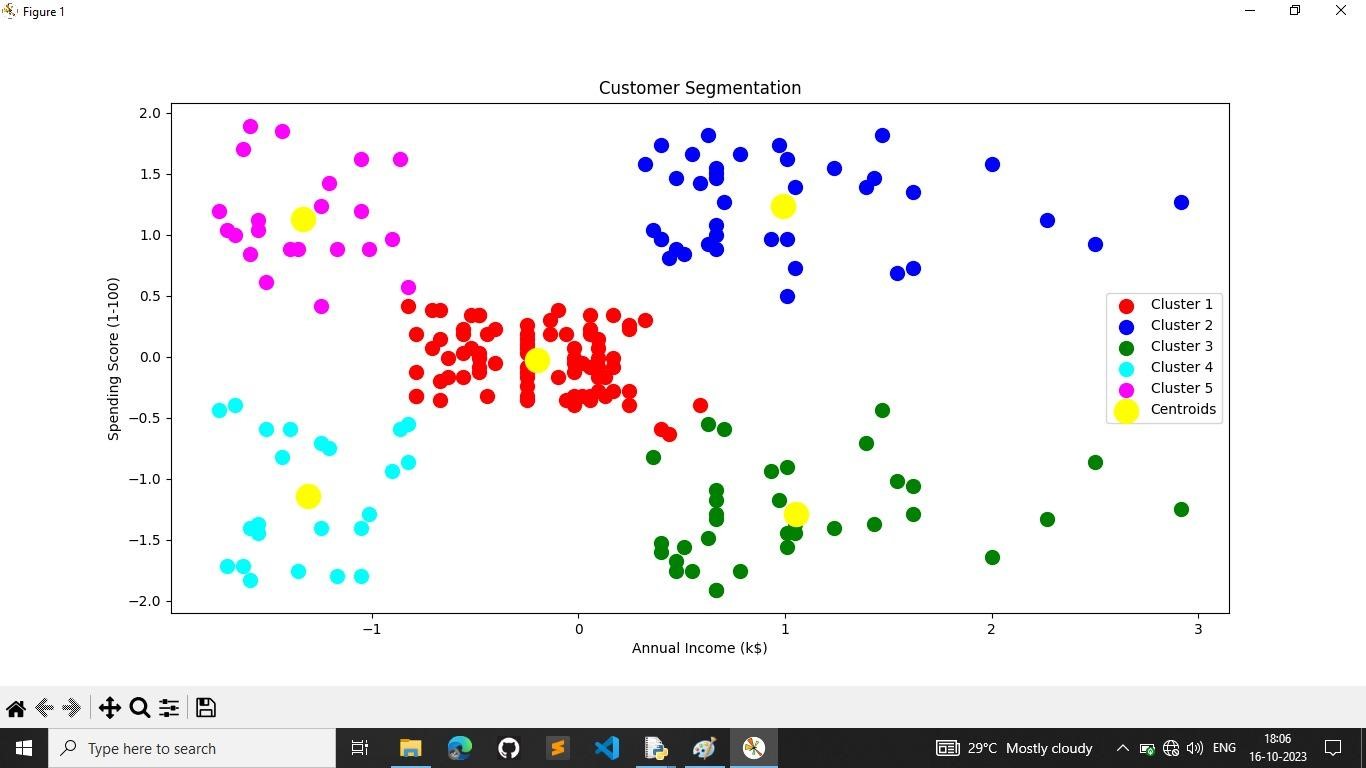
plt.xlabel('Date')

plt.ylabel('Electricity Price')

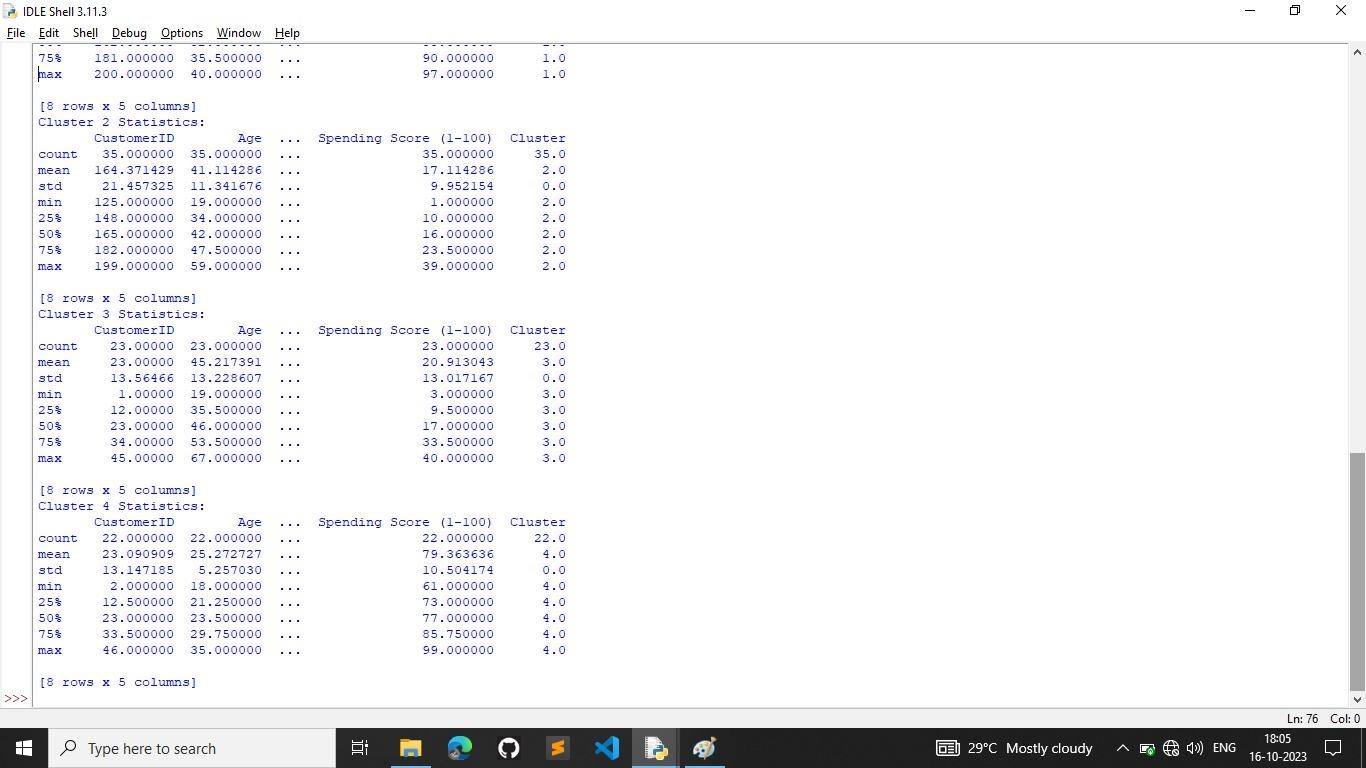
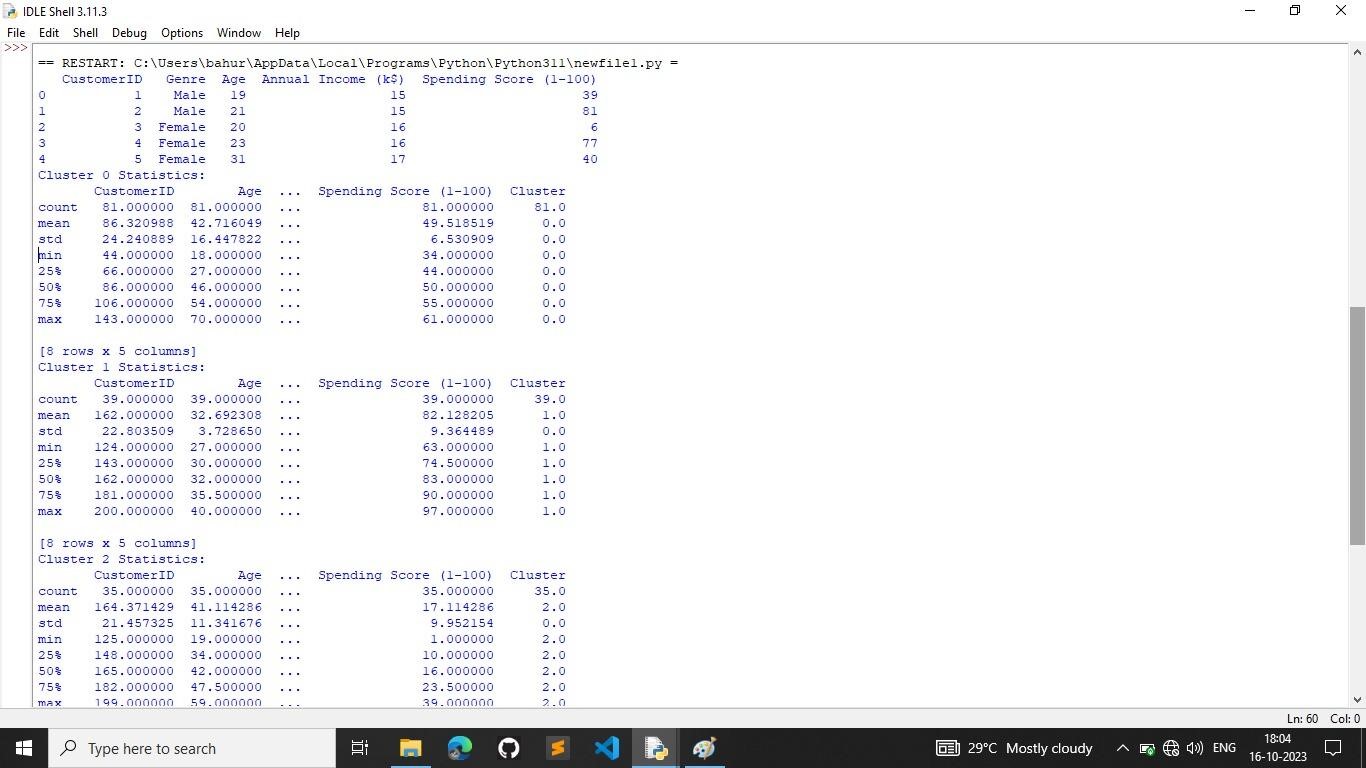
plt.legend()

plt.show()

**OUTPUT**:



:



**Preprocess the data:**

Convert the 'date' column to a datetime format using pd.to\_datetime.

Set the 'date' column as the index of the DataFrame using set\_index.

**Feature engineering:**

Two additional features are created:

'hour': Extracts the hour of the day from the 'date' index.

'dayofweek': Extracts the day of the week (0 for Monday, 6 for Sunday) from the 'date' index.

Split the data into training and testing sets:

X contains the features ('hour' and 'dayofweek').

y contains the target variable ('price').

The data is split into training and testing sets with an 80-20 split ratio using train\_test\_split.

**4.PERFORMING DIFFERENT ACTIVITIES LIKE FEATURE ENGINEERING, MODEL TRAINING, EVALUATION**

**Importing the necessary libraries**:-

Import numpy as np

Import pandas as pd

From sklearn.model\_selection import train\_test\_split

From sklearn.metrics import mean\_squared\_error

From math import sqrt

Import keras

From keras.models import Sequential

From keras.layers import Dense

From sklearn.preprocessing import StandardScaler

**Reading the dataset:-**

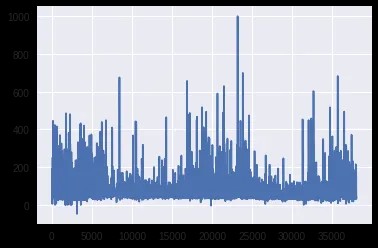
Df=pd.read\_csv(“/content/electricity\_prices.csv”, na\_values=[‘?’])

Df.head()

Df.shape

**Plotting the target feature**:-

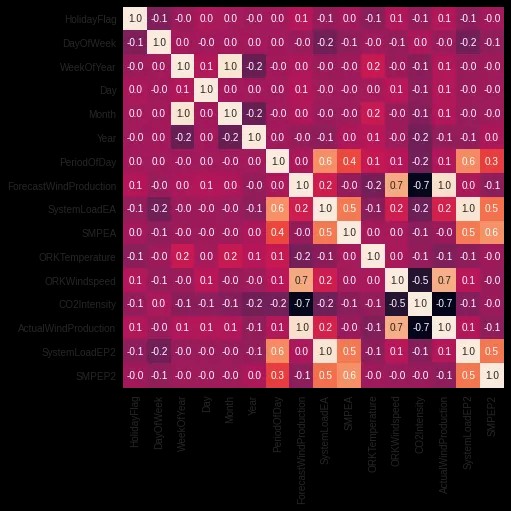
Plt.plot(“SMPEP2”, data=df)



**Correlation plot of Independent attributes:-**

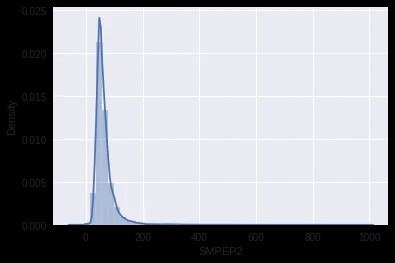
Plt.figure(figsize=(9,7))

Sns.heatmap(df.corr(), annot=True, square=True, fmt=’.1f’, cbar=False);



**Distribution plot of Target feature**:-

Sns.distplot(df[‘SMPEP2’])



**Splitting the independent features and target feature:-**

X = df[[‘ActualWindProduction’, ‘SystemLoadEP2’, ‘SMPEA’, ‘SystemLoadEA’, ‘ForecastWindProduction’,

‘DayOfWeek’, ‘Year’, ‘ORKWindspeed’, ‘CO2Intensity’, ‘PeriodOfDay’]]

Y = df[‘SMPEP2’]

**Compiling the model:-**

Model.compile(loss=’mse’, optimizer=’adam’, metrics=[‘mse’,’mae’])

**Evaluating the model on test set:-**

**fromsklearn.metricsimport** mean\_absolute\_error,r2\_score  
predictions = model.predict(X\_test)  
print(f"MAE: **{**mean\_absolute\_error(y\_test, predictions)**}**")  
  
print(f"R2\_score: **{**r2\_score(y\_test, predictions)**}**")

**Feature selection:**

**1. Domain Knowledge:**

Start by consulting domain experts who are familiar with the energy market. They can provide insights into which features are likely to be the most influential in price prediction. Consider factors such as supply and demand dynamics, weather conditions, market conditions, and regulatory policies.

**2. Correlation Analysis:**

Use statistical techniques to measure the correlation between each feature and the target variable (electricity prices). Features with strong positive or negative correlations are more likely to be relevant. You can use Pearson correlation for continuous features and point-biserial correlation for binary features.

**3. Feature Importance from Models:**

Train a preliminary model (e.g., a random forest, gradient boosting, or a linear regression model) and extract feature importances. Features that the model ranks as more important are likely to have a significant impact on price prediction.

**4. Recursive Feature Elimination:**

Implement RFE, a backward selection method, where you start with all features and iteratively remove the least significant ones based on a model's performance. Continue this process until you achieve the desired model accuracy.

**5. L1 Regularization:**

- Use L1 regularization techniques, such as Lasso regression, which automatically selects a subset of features while shrinking less important ones to zero. This method is particularly effective for high-dimensional datasets.

**Advantages:**

1. **Cost Optimization:**

By accurately predicting electricity prices, businesses and consumers can make informed decisions on when to buy or consume electricity. This helps them optimize their energy usage, reduce costs, and potentially save money on their electricity bills.

**2. Resource Planning:**

Utility companies and energy providers can use price predictions to plan their resource allocation effectively. They can adjust their generation and distribution strategies to meet expected demand, which minimizes operational costs and resource wastage.

**3.Market Participation:**

Traders and investors in energy markets can use price predictions to make profitable decisions. This includes buying electricity when prices are low and selling it when prices are high, maximizing revenue and profits.

**4.Grid Management:**

Electricity price prediction assists grid operators in managing and maintaining the stability of the electrical grid. By anticipating price fluctuations, they can take proactive measures to balance supply and demand, reducing the risk of blackouts or overloading.

**5. Risk Management:**

Businesses, especially those with high energy consumption, can use price predictions for risk management. They can hedge their exposure to volatile energy prices by locking in favorable rates or adjusting their energy usage during high-price periods.

**6. Renewable Energy Integration:**

Price predictions can help renewable energy producers, like wind and solar farms, determine the optimal times to generate and sell their electricity to maximize returns while minimizing grid congestion.

**7. Demand Response:**

Consumers can participate in demand response programs based on price predictions.

**8. Environmental Impact:**

Electricity price prediction can be used to promote environmentally friendly energy consumption. Consumers can choose to use electricity when it's generated from renewable sources or during periods of low emissions, reducing their carbon footprint.

**9. Energy Efficiency:**

Price predictions encourage energy-efficient practices. When consumers and businesses know when electricity is cheaper, they are motivated to use energy-efficient appliances and lighting, which is better for the environment and reduces energy bills.

**10. Regulatory Compliance:**

Electricity price prediction helps utility companies and grid operators comply with regulatory requirements related to pricing transparency, fair market practices, and adherence to environmental standards.

**Disadvantages:**

1. **Inherent Uncertainty:**

Electricity prices can be influenced by a multitude of factors, including unpredictable events like extreme weather conditions, natural disasters, and geopolitical events. As a result, there is inherent uncertainty in predicting electricity prices accurately.

1. **Complexity of Factors:**

Electricity price prediction models often need to consider a wide range of variables, including supply and demand, generation mix, fuel costs, regulatory policies, and weather patterns. Managing and incorporating all these factors into a model can be complex.

1. **Data Quality and Availability:**

Reliable historical data is crucial for training accurate prediction models. However, data may not always be readily available, and its quality can vary. Missing or inaccurate data can lead to less reliable predictions.

1. **Modeling Challenges:**

Developing and maintaining accurate prediction models can be challenging. Different modeling techniques may be required for different market conditions and regions. Model selection and hyperparameter tuning are complex tasks that require expertise.

1. **Market Dynamics:**

Electricity markets are subject to market dynamics that can change rapidly. Sudden shifts in demand or supply, policy changes, or unexpected events can disrupt the validity of existing models.

1. **Overfitting and Underfitting:**

Models can suffer from overfitting (capturing noise in the data) or underfitting (oversimplifying the problem) if not carefully tuned. Balancing model complexity with generalizability is a common challenge.

1. **Regulatory Changes:**

Changes in energy policies, regulations, or pricing structures can have a significant impact on electricity prices. Predictive models may struggle to adapt quickly to such changes.

**8. Behavioral Economics:**

Human behavior and market psychology can have a substantial influence on electricity prices. Predicting how people will react to price changes and market conditions is complex and uncertain.

**9.Lack of Transparency:**

In some regions, electricity markets may lack transparency, making it difficult to access and use relevant market data. This can hinder the development of accurate prediction models.

**10. Model Validation:**

Accurately assessing the performance of prediction models can be challenging. Traditional validation metrics may not always reflect the real-world performance of the models.

**Conclusion**

* Electricity price prediction is a crucial and multifaceted tool that serves various stakeholders in the energy sector, from utility companies and energy traders to consumers and policymakers.
* Its significance lies in its ability to provide valuable insights into future pricing dynamics, enabling more informed and strategic decisions.
* Accurate electricity price prediction allows for the optimization of energy consumption and resource allocation, cost reduction, and improved market efficiency.
* However, it also comes with inherent challenges, including the complexity of factors influencing prices, data quality and availability issues, and the dynamic nature of energy markets.
* Nonetheless, the advantages of electricity price prediction far outweigh its disadvantages.
* The incorporation of design thinking and human-centered approaches, as well as a commitment to addressing ethical and environmental considerations, can help unlock the full potential of electricity price prediction in driving sustainability, cost-efficiency, and reliability in the energy sector.
* Ultimately, while no prediction model can eliminate all uncertainties, electricity price prediction is a valuable tool that empowers individuals and organizations to make more informed, strategic, and responsible decisions in the dynamic world of energy consumption and pricing.

