Electricity Prices Prediction:

2.1 INTRODUCTION:

Electricity Prices Prediction is to 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 the consumption and investment.

It involves data and predictive modeling techniques to forecast future electricity prices. This type of prediction is valuable for various stakeholders, including energy producers, consumers, and market regulators.

Goals of the Project:

1. To help businesses and consumers make better decisions about energy consumption and investment:

Business: to plan its production schedule so that it uses energy during times when prices are low.

Consumer: to decide when to use energy intensive appliances, such as the washing machine and dishwasher.

2. To support the development of renewable energy and energy storage technologies:

Renewable energy company: to identify the best times to generate and sell renewable energy.

3. To improve the efficiency and reliability of the electricity grid:

Grid operator: to identify times when there is a high risk of blackouts or brownouts, and to take steps to mitigate these risks, such as by dispatch power plants or reduce demands.

2.2 Dataset and its details:

Dataset:

https://www.kaggle.com/datasets/chakradharmattapalli/electricity-price-prediction.

A good model for predicting the electricity demand requires analysis of the following types of variables:

Calendar data: Season, hour, bank holidays, etc.

Weather data: Temperature, humidity, rainfall, etc.

Company data: Price of electricity, promotions, or marketing campaigns.

Social data: Economic and political factors that a country is experimenting with.

Demand data: Historical consumption of electricity.

2.3Columns used:

SystemLoadEA – **Target variable**

Features includes other columns like ‘Year’, ‘Month’, ‘Day’, ‘PeriodOfDay’, 'ForecastWindProduction', 'ORKTemperature', 'ORKWindspeed' etc…

2.4 Libraries used:

1. **Pandas**: For data manipulation and handling data in a tabular format

2. **Numpy**: For scientific and mathematical computations.

3. **Matplotlib**: For creating data visualizations to visualize the actual and predicted values.

4. **statsmodels.tsa.arima_model.ARIMA:**

It is part of the statsmodels package used for time series analysis. It provides the ARIMA model for time series forecasting.

5.**sklearn.metrics.mean_squared_error:**

It is from scikit-learn (sklearn) and is used for calculating the Mean Squared Error (MSE), which is a measure of the accuracy of the ARIMA model's predictions.

2.5 Training and Testing data:

The given data is split into training and testing sets, and an ARIMA model is trained on the training data and then used to make predictions on the testing data.

Step 1: Split the dataset:

In this code 80% of the data is used for training and 20% is used for testing.

Step 2: Train the ARIMA Model:

An ARIMA model is created using the training data (train). The model is configured with an order of (5, 1, 0), which means it's an ARIMA(5,1,0) model (p=5, d=1, q=0).

Step 3: Make Predictions:

Predictions are made on the testing data (‘test’) using the fitted ARIMA model. The ‘start’ and ‘end’ indices are set to specify the range of data for which predictions are made.

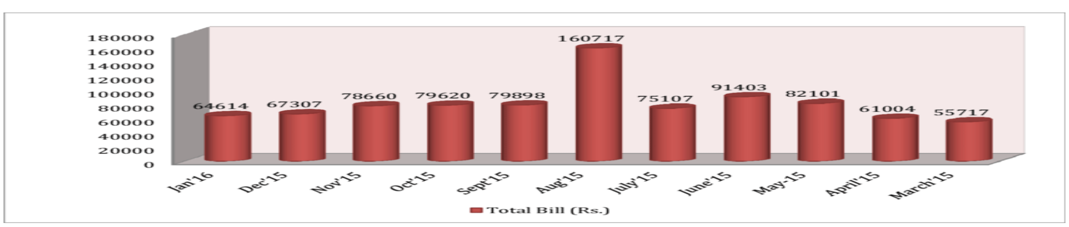
Step 4: Calculate Mean Squared Error (MSE):

```
mse = mean_squared_error(test[target_variable], predictions)
```

The Mean Squared Error (MSE) is calculated to assess the accuracy of the ARIMA model's predictions. The MSE quantifies the squared differences between the actual and predicted values.

Step 5: Visualize the Predictions:

- Using plt.legend() , plt.show() the code then visualizes the actual values and the predicted values using Matplotlib.
- It displays the plot, showing both the actual and predicted values on the same graph. The actual values are in one color, and the predicted values are in red.



2.6 Explanation:

Steps of how electricity price prediction works:

1. **Feature Engineering**: Feature engineering involves selecting and transforming the relevant data attributes (features) that will be used in the prediction model. For electricity price prediction, features might include historical price trends, weather conditions (temperature, wind speed, sunlight), demand and supply patterns, economic indicators, and more.

2. **Model Selection**: Machine learning and statistical models are commonly used for electricity price prediction. Models like regression analysis, time series analysis, neural networks, and more advanced techniques like Long Short-Term Memory (LSTM) for time series data can be employed. The choice of model depends on the complexity of the problem and the quality of the available data.

3. **Training the Model:** Historical data is used to train the selected prediction model. The model learns patterns and relationships within the data to make future price predictions. This training process involves adjusting model parameters to minimize prediction errors.

4. **Validation and Testing**: After training, the model is validated and tested to assess its accuracy and generalization capabilities. This involves using a portion of the data that the model hasn't seen during training to check how well it performs.

5. Prediction: Once the model is validated, it can be used to make future electricity price predictions. Users input relevant data, and the model generates forecasts. These predictions can range from short-term (e.g., hourly or daily) to long-term (e.g., monthly or yearly).

6. Continuous Monitoring and Updating: Electricity price prediction is an ongoing process, and the models need to be continuously monitored and updated. Changes in market conditions, regulations, or other external factors can impact the accuracy of predictions.

7. Decision Making: Stakeholders, such as energy providers, consumers, and traders, use the predictions to make decisions. For example, energy providers may adjust their generation strategies, consumers may modify their consumption patterns, and traders may make investment decisions based on predicted price movements.

It's important to note that predicting electricity prices is a complex task due to the many variables that can influence them, including weather conditions, energy supply, demand, and market dynamics. Therefore, models can only provide estimates, and actual prices may deviate from predictions. However, accurate forecasting can still offer valuable insights and help stakeholders make more informed decisions in the energy market.

Project link:
https://colab.research.google.com/drive/1DA8txYdHVLOMotL4SGruHnsNsOLTuVJj?usp=sharing

2.7 Metrics Used:

1. **Mean squared error (MSE):**
    - MSE is a measure of the average squared difference between the actual and predicted values. It is a common metric for evaluating the performance of regression models and other times series forecasting tasks as well.
    - It quantifies the average squared differences between the predicted prices and the actual prices in your test dataset.
    - **Lower MSE**: Indicates that the model's predictions are close to the actual electricity prices, suggesting that the model is making accurate predictions.

      Higher MSE: The model's predictions are further from the actual prices, indicating less accurate predictions.

      Units: Typically expressed in the square of the units of the target variable. For electricity price prediction, this would usually be the square of the price units (e.g., cents, dollars, etc.).

    ```
    # Calculate Mean Squared Error
    mse = mean_squared_error(test[target_variable], predictions)
    ```

2. **AIC**:

    AIC (Akaike information criterion) is a measure of the relative quality of statistical models. It is a penalty function that penalizes models with more parameters. The model with the lowest AIC is considered to be the best model.

The MSE is used to evaluate the performance of the ARIMA model on the testing set. The AIC is used to select the order of the ARIMA model.