ELECTRICITY PRICES PREDICTION

PROBLEM DEFNITION:

Predicting electricity prices is of paramount importance to energy providers and consumers, as it informs crucial decisions related to consumption and investment. This abstract outlines a comprehensive approach to developing a predictive model for electricity price forecasting, leveraging data science techniques.

OBJECTIVE:

- The objective of this project is to develop a robust and accurate electricity price prediction model using some data science techniques. This model will leverage historical electricity pricing data and relevant features to forecast future electricity prices.
- 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.

DATA COLLECTION:

The first step is data collection. In this module we have to collect the data such as date, demand, supply, weather conditions, and economic indicators. Additionally, gather sources the external data sources like. Period of the day, forecast wind production.

DATA PREPROCESSING:

Data pre-processing is a crucial step in the process of electricity price prediction using data science techniques. Proper data pre-processing ensures that the dataset is clean, structured, and ready for analysis and modelling.

1. Handling Missing Values:

Identify and handle missing values in the dataset. Common strategies include imputation (replacing missing values with estimates like mean, median, or mode) or removal of rows or columns with missing data.

2. Dealing with Outliers:

Identify and handle outliers that can skew the model's predictions. Techniques like Winsorization or transforming data using the logarithm can be applied to mitigate the impact of outliers.

3. Data Normalization/Scaling:

Normalize or scale numerical features to ensure they have similar scales. Common methods include Min-Max scaling or Standardization (z-score scaling). This step is essential for models like neural networks and K-means clustering.

4. Handling Categorical Variables:

Encode categorical variables into numerical representations using techniques like one-hot encoding or label encoding. This allows machine learning algorithms to work with categorical data.

FEATURE ENGINEERING:

Feature engineering is a critical step in developing predictive models for electricity price prediction. It involves creating new features from the existing dataset or transforming existing features to better capture patterns and relationships that can improve the model's accuracy.

1. Moving Average:

- Compute various moving averages from the electricity price prediction such as Weekofyear, PeriodOfDay,etc

2. Lagged Values:

Create lagged features by including past stock prices or returns as predictors. Lagged features can capture autocorrelation patterns and dependencies over time.

MODEL SELECTION:

When selecting a model for stock price prediction, you should consider the characteristics of financial time series data, such as non-stationary, autocorrelation, and volatility clustering. Here are some model options commonly used in stock price prediction:

1. Time Series Models:

Autoregressive Integrated Moving Average (ARIMA): ARIMA models are well-suited for modeling the temporal dependencies in stock price data. They can capture trends, seasonality, and autoregressive behavior. - GARCH (Generalized Autoregressive Conditional Heteroskedasticity): GARCH models are used to model volatility clustering and changing variances in financial time series data. - Prophet: Developed by Facebook, Prophet is a robust time series forecasting model that can handle irregularly spaced data, holidays, and seasonal effects. It is relatively easy to use and offers good out-of-the-box performance.

2. Machine Learning Models:

Linear Regression: Simple linear regression models can capture linear trends and relationships in stock price data. Multiple linear regression can be used to account for multiple predictors. - Random Forest: Random Forest is an ensemble learning method that can capture non-linear relationships and feature importance. It's robust and works well with both numerical and categorical features.

Gradient Boosting (e.g., XGBoost, LightGBM): Gradient boosting algorithms are powerful for regression tasks and can capture complex patterns in the data. They require careful tuning but often deliver high accuracy.

Long Short-Term Memory (LSTM): LSTM is a type of recurrent neural network (RNN) that is well-suited for sequence modeling tasks like time series prediction. It can capture long-term dependencies in data. - Convolutional Neural Networks (CNN): CNNs can be applied to analyze stock price data, especially when considering patterns in technical indicators or visual data. - Ensemble Models: Combining predictions from multiple models (e.g., blending, stacking) can often lead to improved accuracy.

3. Hybrid Models:

ARIMA-GARCH Hybrid: Combine the strengths of ARIMA and GARCH models to capture both the trend and volatility of stock prices. - Machine Learning with Technical Indicators: Use machine learning models in conjunction with technical indicators as features to capture both fundamental and technical aspects of stock price movements.

4. Deep Learning Models:

Recurrent Neural Networks (RNNs): Similar to LSTM, RNNs can capture sequential dependencies and are suitable for time series forecasting

MODEL TRAINING:

Accurate prediction of electricity prices is of paramount importance for the efficient functioning of energy markets and informed decision-making by stakeholders. This abstract outlines the methodology and objectives of model training in the context of electricity price prediction, emphasizing the utilization of data science techniques.

This research project focuses on training predictive models using a meticulously pre-processed dataset containing historical electricity price data and relevant influencing factors. The objective is to harness the power of data science algorithms to develop accurate and robust models capable of forecasting future electricity prices.

1. Data Preparation:

Prepare your historical stock price and related financial data, ensuring it is cleaned, preprocessed, and split into training, validation, and test sets.

2. Select Features:

Choose the relevant features (predictors) that will be used to train your model. These features can include historical stock prices, trading volumes, technical indicators, economic indicators, and any engineered features.

3. Choose a Model:

Select the machine learning, time series, or hybrid model you intend to use for stock price prediction based on your project's objectives and the nature of your data. For example, you might choose an LSTM neural network for sequential data or an Boost ensemble model for tabular data.

4. Model Training:

Train the selected model on the training dataset. This involves feeding the historical data and associated target values (future stock prices) into the model and optimizing its parameters. Use appropriate loss functions and optimization algorithms for training.

5. Hyper parameter Tuning:

Tune the hyper parameters of your model using the validation dataset to improve its performance. This might involve adjusting learning rates, batch sizes, model architecture, or regularization techniques.

EVALUATION PHASE:

1. Testing Data Preparation:

Use the test dataset, which should be kept separate and not used during model training or validation, to evaluate the model's performance on unseen data.

2. Prediction Generation:

Use the trained model to generate predictions for the test dataset. These predictions represent the model's estimated future stock prices.

3. Performance Metrics:

Calculate various performance metrics to evaluate how well the model's predictions align with the actual stock prices. Common evaluation metrics for regression tasks in stock price prediction include:

- Mean Absolute Error (MAE): Measures the average absolute difference between predicted and actual prices.

• Mean Squared Error (MSE): Measures the average squared difference between predicted and actual prices.

• Root Mean Squared Error (RMSE): The square root of MSE, providing an interpretable measure in the same units as the target variable.

• R-squared (R^2) Score: Measures the proportion of variance explained by the model.

• Mean Absolute Percentage Error (MAPE): Measures the percentage difference between predicted and actual prices.

4. Visualization:

Visualize the model's predictions alongside actual stock prices using plots and charts to gain insights into how well the model is capturing trends and patterns.

CONCLUSION:

In conclusion, electricity price prediction using data science techniques represents a powerful and transformative approach to addressing the complexities of energy markets. This endeavour is underpinned by the recognition that accurate price forecasts are not just advantageous but imperative for various stakeholders in the energy sector, including utilities, consumers, and market operators. Electricity price prediction through data science is a cornerstone of informed decision-making in the energy sector, with the potential to revolutionize how we manage and consume electricity, ultimately benefiting both industry professionals and consumers alike.