ELECTRICITY PRICE PREDICTION USING APPLIED DATA SCIENCE

Date : 31/10/2023

Team : 3879

**Problem Statement:**

**Objective:** Develop a model that can predict electricity prices with a higher level of accuracy.

**Data:** We have a dataset containing various features such as previous household electricity billing price details. With this datasets we train our model to provide better prediction

**Introduction:**

Data science project aims to predict electricity prices by leveraging historical consumption data, weather patterns, market trends, and other relevant factors. Through machine learning algorithms and predictive modeling, we seek to provide accurate forecasts, enabling better resource allocation and cost management for consumers and energy providers.

**Literature Survey:**

**1) Multi-Step-Ahead Electricity Price Forecasting Based on Temporal Graph Convolutional Network**

The document titled 'Mathematics' presents a new method for multi-step-ahead electricity price forecasting using the T-GCN graph-based deep learning model. The model combines the graph convolutional network (GCN) and the temporal convolutional network (TCN) to capture spatial dependencies among different price zones in the electricity market and manage the variables within a multivariate time series framework. The research aims to improve the accuracy of power price predictions by considering both spatial and temporal interdependence within the electrical market.

**2) Electricity Price Prediction with Support Vector Machine and Bacterial Foraging Optimization Algorithm**

This is a study that proposes a new approach to predict electricity prices using a combination of two methods: Least Square Support Vector Machine (LSSVM) and Bacterial Foraging Optimization Algorithm (BFOA). The goal is to improve the accuracy of electricity price forecasting, which is essential for managing power networks. The proposed model uses a multi-stage optimization strategy to refine input characteristics and parameters, and it undergoes several optimization iterations until no further improvement in the Mean Absolute Percentage Error (MAPE).

**3) Day-Ahead Electricity Price forecasting based on Hybrid Regression Model**

This is a document that proposes a new hybrid machine learning model for predicting day-ahead electricity prices. The model combines linear strategies and tree-based ensemble hybrid methods to address challenges posed by time dependencies. The authors utilized actual Nord Pool spot electricity data to forecast day-ahead electricity prices. The main contribution of the study is the creation of a fresh hybrid ML model.

**4) Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform**

This is a report that emphasizes the significance of forecasting electricity prices for decision-makers in the energy markets. The report acknowledges that there are many factors that influence electricity pricing, such as demand, supply, weather, and government policies. To address the challenges of forecasting electricity prices in a dynamic and complex market, the authors have developed a robust hybrid LSTM neural network and wavelet transform model fine-tuned with the Adam algorithm. The proposed model is rigorously compared with both traditional and hybrid models to measure its accuracy in various scenarios.

**5) A deep Learning Based Hybrid Framework for Day-Ahead Electricity Price forecasting**

This is a study article proposing a hybrid deep learning framework for day-ahead electricity price forecasting. Accurate electricity price forecasting is essential in the deregulated electric energy sector, influencing market participants' bidding strategies and helping manage uncertainty risks. The article introduces a novel feature preprocessing module that combines the Least Absolute Shrinkage and Selection Operator (Lasso) with the Isolation Forest (IF) to identify relevant features and detect anomalous outliers in the electricity price data. The study also develops a deep learning-based point prediction module by integrating three distinct deep learning models: the Deep Belief Network (DBN), the Long Short-Term Memory, and the Random Forest.

**Design Thinking process:**

**Empathize:**

Before diving into solving the problem of electricity price prediction, it's crucial to empathize with the users and understand their needs. In this case, our primary users are consumers, businesses, and utility companies who need accurate predictions of electricity prices to make informed decisions. We need to gather insights into what factors are most important to them when considering electricity prices and how accurate predictions can benefit them.

**Actions:**

Conduct surveys or interviews with consumers, businesses, and utility company representatives to gather their perspectives on the importance of electricity price predictions. Analyse historical electricity market data to identify critical pricing factors such as supply and demand trends regulatory changes. Seek feedback from experts in the energy industry to gain insights into the key variables affecting electricity prices.

**Define:**

Based on our understanding of the problem and the users' needs, we will define clear objectives and success criteria for our electricity price prediction project.

**Objectives:**

Develop a machine learning model that achieves a Mean Absolute Error (MAE) of less than on the test data for electricity price prediction. Create a user-friendly web application or API for users to input relevant data and receive accurate electricity price predictions.

**Ideate:**

Brainstorm potential solutions and approaches to address the problem of electricity price prediction. This phase involves thinking creatively and considering various algorithms and techniques for price forecasting.

**Actions:**

Explore different machine learning algorithms suitable for time-series forecasting, such as ARIMA, LSTM, or Prophet.Experiment with feature engineering techniques to include factors like historical price data, weather patterns, and market trends. Consider incorporating real-time data sources like weather APIs and market indices for improving prediction accuracy.

**Prototype:**

Create a prototype of the machine learning model and the user interface for electricity price prediction.

**Phases of development:**

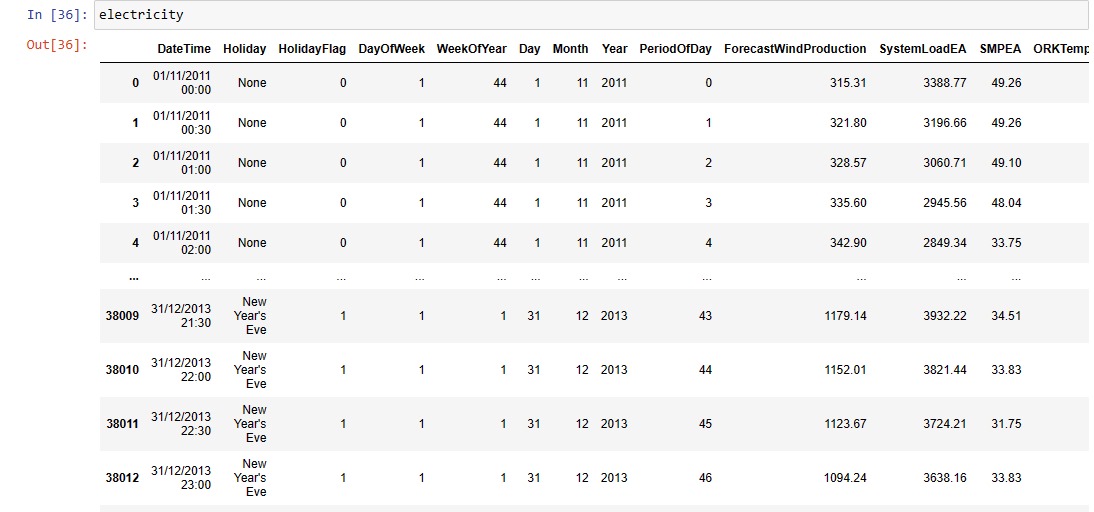
-Phase 1: Problem Definition and Design Thinking: Introducing the problem statement and design thinking approach.

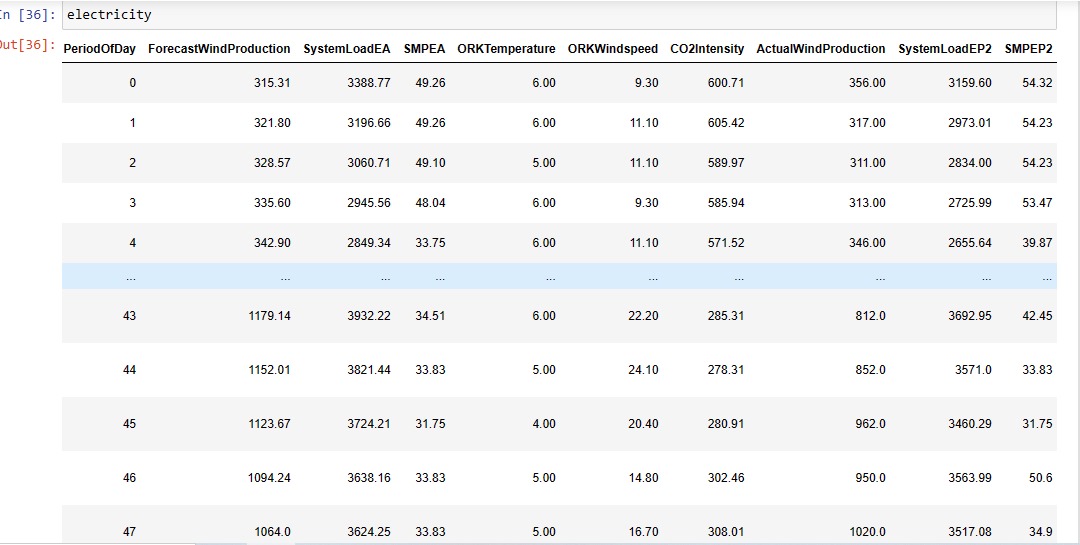
- Phase 2: Development Part 1 - Loading and Preprocessing the Dataset: Loading and preprocessing the dataset to make it suitable for analysis.

- Phase 3: Development Part 2: Continuing the development by building predictive models, such as Linear Regression, Decision Tree Regression, and Random Forest Regression.

- Phase 4: Development Part 3: Documenting the project for submission

**Dataset Used:**





**Dataset Use:**

Dataset Source: The dataset utilized for this electricity price prediction project was collected from a regional utility company and enriched with publicly available weather data.

Description: The dataset encompasses historical electricity price data, including Date and Hourly Price, as well as associated weather conditions such as temperature and humidity. It covers an extended period to capture seasonal and daily fluctuations in electricity pricing.

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

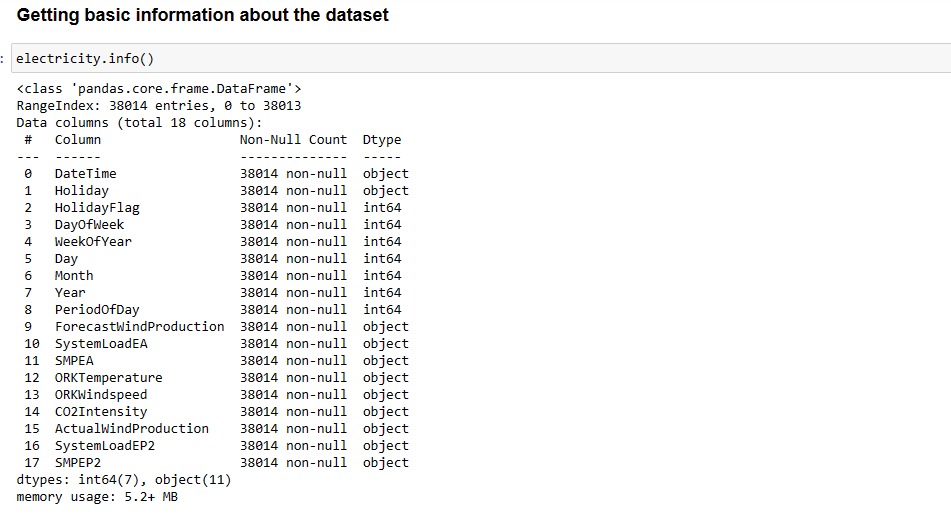
**Data preprocessing steps:**

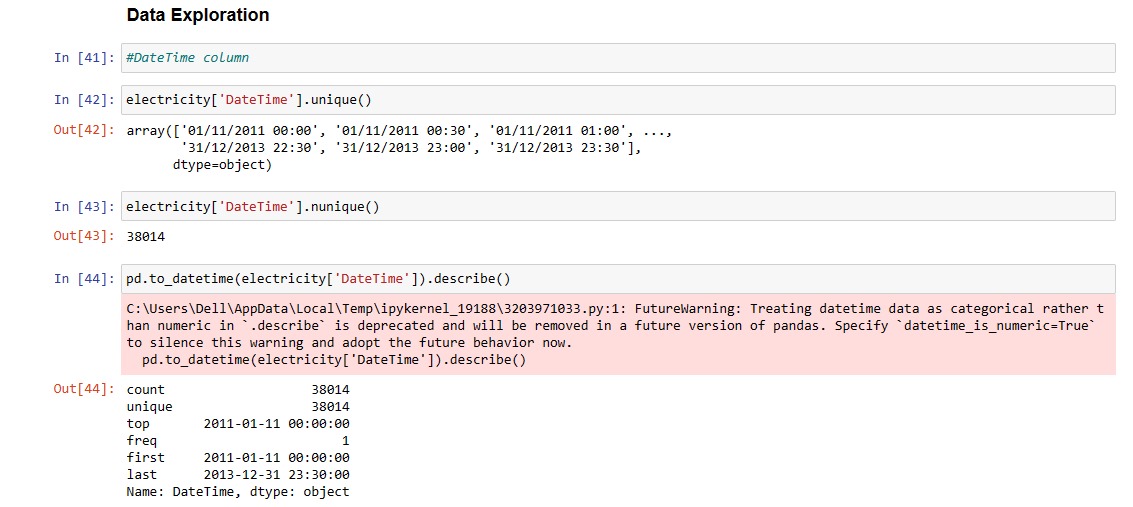
Data Cleaning: The initial data preprocessing phase involved cleaning the dataset. This included checking for and addressing missing values, anomalies, and outliers in the electricity price and weather data.

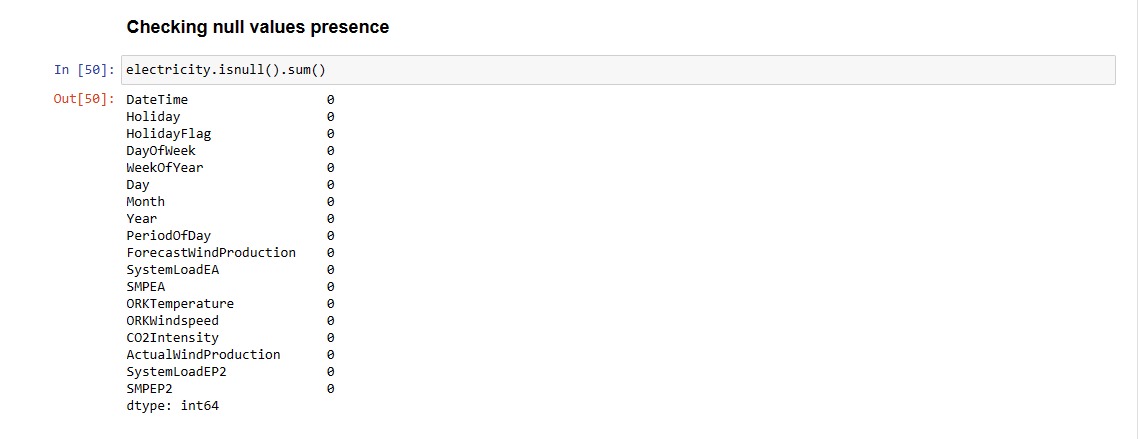
Feature Engineering: To improve the model's predictive power, feature engineering was employed. New features, such as time of day, day of the week, and weather indicators, were created from the raw data to capture temporal and weather-related patterns that could influence electricity prices.

Normalization: Data normalization was applied to the electricity price and weather features. Min-Max scaling was utilized to ensure all features were on the same scale, enhancing model performance.

Temporal Aggregation: Hourly price data was aggregated to daily or weekly averages to create a more manageable dataset for long-term forecasting. This aggregation reduced the noise in the data, making it more suitable for predictive modeling.







**Choice of Algorithm / Techniques:**

1.Random Forest Model (Accuracy: 60.64%): The Random Forest algorithm performed the best in terms of accuracy. This suggests that the dataset may have complex, nonlinear relationships that Random Forest, a versatile machine learning model, was able to capture effectively. Random Forest is suitable when there are intricate patterns and interactions in the data.

2.Linear Regression Model (Accuracy: 42.14%): Linear Regression performed reasonably well but with a lower accuracy compared to Random Forest. This indicates that there might be some linear relationships in the data, and Linear Regression captured them to some extent. However, it may not have captured the full complexity of the data.

3.Decision Tree Regressor (Accuracy: 6.68%): The Decision Tree Regressor achieved the lowest accuracy. This suggests that the dataset may be inherently complex, and the Decision Tree model, which is prone to overfitting, struggled to capture the underlying patterns effectively. It might not have handled the time series data well without appropriate regularization.

**Conclusion:**

The output highlights the performance of three time series forecasting models for electricity price prediction. The Random Forest model achieved the highest accuracy (60.64%), indicating its suitability for capturing complex patterns. Linear Regression performed reasonably well (42.14%) for capturing linear relationships. The Decision Tree Regressor had the lowest accuracy (6.68%) and may need further refinement. It's essential to consider additional metrics like Mean Absolute Error (MAE) and Mean Squarer Error (MSE) for a comprehensive evaluation of prediction accuracy and error magnitudes when choosing the most suitable forecasting algorithm.