Time Series Forecasting for Energy Consumption: Optimizing Resource Allocation in Smart Grid Systems

Name: SREELAKSHMI M

Organization: Entri Software Pvt Limited

Date: 13/05/2024

1. INTRODUCTION

In a world increasingly reliant on energy, understanding and predicting energy consumption patterns are crucial for effective resource management and sustainable development. The aim of this project is to develop predictive models for energy consumption based on various factors such as temperature, time of day, and economic indicators. By leveraging machine learning techniques and advanced data analysis, we seek to uncover underlying trends and relationships within the data to enhance our understanding of energy usage dynamics.

The project dataset comprises comprehensive information, including temperature, humidity, time of day, and previous energy consumption, spanning a significant period. Despite extensive preprocessing efforts to ensure data quality, challenges remain in extracting meaningful insights and building accurate prediction models due to the complex interplay of factors influencing energy consumption.

Through this project, I endeavor to address these challenges by employing a multifaceted approach encompassing feature engineering, advanced modeling techniques, and cross-domain analyses. By delving into the intricacies of energy consumption patterns, I aim to not only improve prediction accuracy but also inform evidence-based decision-making in energy policy and management domains.

2. PROBLEM STATEMENT

This project aims to develop a time series forecasting model for predicting energy consumption patterns in smart grid systems. By leveraging historical data on energy usage, weather conditions, holidays, and other relevant factors, the model seeks to forecast future energy consumption accurately. The goal is to optimize energy distribution and resource allocation, enabling efficient management of electricity grids. By forecasting energy consumption patterns, this project contributes to reducing costs, improving grid reliability, and promoting sustainability in energy management.

3. MODELS

Choosing the appropriate machine learning model is crucial for the success of the predicting energy consumption patterns in smart grid systems. Considerations include the complexity of the data, the interpretability of the model, and computational efficiency. So, models include linear regression, Random Forest regressor, Support Vector Regressor and multi-layer perceptron(MLP) regressor are used.

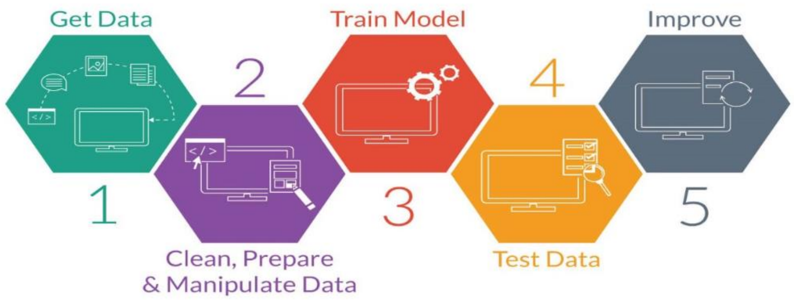
Support Vector Regression (SVR) is a supervised learning algorithm effective for capturing complex, non-linear relationships in high-dimensional spaces.

Multi-layer Perceptron Regressor (MLPRegressor) is an artificial neural network model capable of learning complex patterns, though requiring careful hyperparameter tuning.

Random Forest Regressor is an ensemble learning method based on decision trees, known for its robustness to outliers and noise in high-dimensional datasets.

Linear Regression is a simple and interpretable regression algorithm, providing insights into linear relationships between features and the target variable.

4. MODEL ARCHITECTURE



5. EXPERIMENTAL RESULTS AND DISCUSSION

5.1. Dataset

The dataset was given by the ENTRI Elevate team. The given dataset has 43825 rows and 15 columns.

The feature are 'Equipment_ID', 'Sensor_1', 'Sensor_2', 'Sensor_3', 'Environmental_Temperature','Environmental_Humidity','Production_Volume', 'Operating_Hours', 'Error_Code', 'Equipment_Age', 'Power_Consumption', 'Voltage_Fluctuations', 'Current_Fluctuations', 'Vibration_Analysis', 'Temperature_Gradients', 'Pressure_Levels' and 'Failure_Maintenance_Indicator'.

'Failure_Maintenance_Indicator' is the target variable.

5.2 Data Preprocessing

- Preprocessing steps involve handling missing values, normalizing attributes and splitting the dataset into training, validation, and test sets.
- Performing EDA to get insights of the data like identifying distribution, outliers etc.
- Check any null values present in the dataset. If present, then impute or remove those null values.
- Checking for duplicate values.
- Encode the categorical features/columns.
- Removing unnecessary columns.
- Perform several visualization tools like Boxplot, Scatter plot, Count plot and Histogram.
- Checking the correlation between features and target variables using correlation matrix and heatmap.
- Perform Standard Scalar to scale down values.

5.3. Model Selection

Choosing the appropriate machine learning model is crucial for the success of the predicting energy consumption patterns in smart grid systems. Considerations include the complexity of the data, the interpretability of the model, and computational efficiency. So models include linear regression, Random Forest regressor, Support Vector Regressor and multi-layer perceptron(MLP) regressor are used.

5.4. Feature Selection

Using feature selection methods like Selectkbest, SelectFromModel with Lasso (L1 Regularization), Recursive Feature Elimination (RFE) with Random Forest Regressor and Variance Threshold.

5.5. Model Training

Split 80% of the dataset into training data (used for model training) and 20% into testing data (used for model evaluation).

During training, the model learns patterns and relationships in the training data to make predictions.

5.6. Model Evaluation

Use the trained model to make predictions on the test data (X_test) and calculating performance metrics (Eg: mean absolute error(MAE), mean squared error(MSE) and r2 score) to assess how well the model generalizes to unseen data.

6. RESULT AND ANALYSIS

Without using any feature selection methods or hyperparameter tuning linear regression gives the best r2 score (0.00081291). The other models give negative r2 score.

Among all these feature selection techniques, Recursive Feature Elimination (RFE) with Random Forest Regressor method performed well and gave Linear regression as the best model with an r2 score of 0.00093776. The performance of the remaining models from the feature selection method Recursive Feature Elimination (RFE) with Random Forest Regressor is given in **Fig.1.**

The r2 score of models is comparatively low and it is because of the problems in the dataset. The features given in the dataset have a very low correlation with the target variable.

```
SVR Mean Squared Error: 1348656.0436222448
SVR Mean Absolute Error: 1009.6460480032588
SVR R^2 Score: -3.4081276090836e-05
MLPRegressor Mean Squared Error: 1366772.9730380636
MLPRegressor Mean Absolute Error: 1014.846711595511
MLPRegressor R^2 Score: -0.01346785999941269
Random Forest Mean Squared Error: 1368263.8915046924
Random Forest Mean Absolute Error: 1013.3790755299904
Random Forest R^2 Score: -0.014573382260691714
Linear Regression Mean Squared Error: 1347345.4114553772
Linear Regression Mean Absolute Error: 1009.0905722435082
Linear Regression R^2 Score: 0.0009377579419830306
```

Fig.1

7. LIMITATIONS

Despite the absence of missing values and outliers, the low correlation between features and the target variable suggests that the selected features may not adequately capture the underlying patterns or predictive signals related to equipment failures or maintenance needs.

The low correlation between features indicates that the chosen features may not fully capture the complex relationships and interactions within the dataset. This limitation could hinder the models' ability to generalize well to unseen data or accurately predict future maintenance events.

Limited discriminatory power despite using multiple feature selection methods. Model accuracy may be influenced by complexity, hyperparameter sensitivity, or overfitting.

The dataset's limited scope or representation of industrial machinery and operational conditions may also contribute to the low accuracy of the models.

External factors and context not accounted for in the dataset may introduce uncertainty in model predictions.

8. CONCLUSION

This project highlights the effectiveness of machine learning techniques for a time series forecasting model for predicting energy consumption patterns in smart grid systems . The examination of four machine learning models that are SVR, MLP Regressor, Random Forest Regressor and Linear Regression, for predicting machine failure shows that Linear Regression is the best performing model with an r2 score of 0.00093776.The feature selection method, Recursive Feature Elimination (RFE) with Random Forest Regressor, showcased promising results by identifying a subset of features conducive to linear regression modeling. These findings underscore the challenges posed by the dataset's limited feature representation and sparse correlations with the target variables.

9. FUTURE WORKS

Enhancing feature engineering methodologies could prove fruitful, involving the exploration of additional features or transformations to better encapsulate the intricacies of energy consumption patterns. Advanced modeling techniques, such as ensemble methods or deep learning architectures, could be investigated to potentially uncover nonlinear relationships within the data. Moreover, delving into cross-domain analyses may yield valuable insights by examining correlations between energy consumption and variables from other domains, such as weather patterns or economic indicators. Spatial analysis offers another promising avenue, enabling the assessment of regional variations in energy usage patterns.