ENERGY CONSUMPTION FORECASTING USING MACHINE LEARNING FOR SMART GRID TECHNOLOGY

Submitted to

SASKATCHEWAN POLYTECHNIC

ARTIFICIAL INTELLIGENCE & DATA ANALYTICS

PROJ-613-R05 - Capstone Project

Under the guidance of

Rahim Samei

Submitted by,

- Anish Appu Kuttan
- Febin Lawrence Pravada Rence

Table of Contents

INTRODUCTION:

This project aims to develop a machine learning-based solution for predicting household energy consumption in smart grids, with a focus on optimizing electricity usage and enhancing grid efficiency. The sequential structure of energy data will also be better handled by time-series models like LSTM (Long Short-Term Memory) and ARIMA (Auto Regressive Integrated Moving Average), which will provide more accurate predictions over time. To identify underlying patterns in energy usage, conventional regression models using methods like Linear Regression, Random Forest Regressor, and Gradient Boosting (XG Boost) will be used. The accuracy and robustness of the model will be evaluated by applying feature engineering approaches and employing evaluation measures such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

OBJECTIVE:

The main objective of this project is to develop a machine learning-based predictive model that can forecast household energy consumption in real-time. The project will explore how smart grids can use these predictions to implement demand response programs, encouraging homeowners and power distributing companies for better incentives. This solution will ultimately reduce energy costs, promote efficiency, and contribute to a more resilient energy grid.

FIELD OF THE PROJECT:

The multidisciplinary project **"Energy Consumption Forecasting using Machine Learning for Smart Grid Technology"** crosses many academic and professional domains. By utilizing varied knowledge and approaches, this junction promotes all-encompassing solutions. The primary fields involved in this project include:

Electrical Engineering

This discipline adds the information required to build and improve electrical grids for better energy distribution, with a focus on **smart grid technology** and **power systems management**. **Integration of renewable energy** sources is also supported, as this is necessary for sustainable energy management.

Data Science and Machine Learning

The development of **algorithms** and **predictive analytics** are made possible by data science and machine learning, which are essential to this project. The prediction of energy usage will be made more accurate by utilizing methods like time-series forecasting and neural networks.

Computer Science

System integration, **data management**, and **software development** are all heavily dependent on computer science. This field ensures that a scalable system is built that can process large energy datasets, put machine learning models into practice, and visualize results for users.

Environmental Science and Sustainability

The project's objective of lowering **carbon footprint** through **energy efficiency** is driven by this field. Precise energy forecasting facilitates more sustainable energy consumption management and helps integrate renewable energy sources more effectively.

Statistics and Applied Mathematics

The machine learning models are based on statistical modeling, time-series analysis, and **optimization techniques**. The mathematical precision required for accurate and trustworthy energy consumption projections is ensured by this field.

IMPACT ON AREA OF STUDY:

Integrating machine learning into energy consumption forecasting within smart grids will have several significant impacts:

- **Enhanced Grid Efficiency**: Better demand management and decreased energy waste are the results of improved forecasting, which also increases grid reliability.
- **Renewable Energy Integration**: Precise forecasts make it easier to integrate intermittent renewable energy sources like wind and solar power.
- **Economic Benefits**: Optimizing energy distribution can lead to cost savings for both energy providers and consumers.

- **Academic Contribution**: By integrating machine learning into actual energy systems, the initiative will add to the body of knowledge in this area and maybe open new avenues for study and advancement.

AIM OF THE PROJECT:

1. **Electricity Demand Forecasting:** Accurately predict future electricity demand using historical data and machine learning models. This can help energy providers optimize power generation and supply management. (1)
2. **Renewable Energy Generation Prediction:** Forecast the generation of renewable energy (wind and solar) to assist in balancing the energy grid and maximize renewable energy utilization. (2) (5)
3. **Holiday Impact on Energy Demand:** Analyze and predict the changes in energy demand during holidays to improve grid reliability and planning during high-demand periods. (6)
4. **Energy Storage Optimization**: Forecast when to store surplus energy from renewable sources to optimize the use of energy storage systems and improve efficiency in energy distribution. (7)
5. **Time-Series Anomaly Detection for Energy Data:** Identify unusual patterns or anomalies in energy demand or generation, which can indicate outages, abnormal conditions, or potential inefficiencies. (3)

We intend to improve the project in the future by incorporating IoT-enabled smart meters and time-series data to enable dynamic load forecasting and energy optimization. Grid operators will be better equipped to balance supply and demand in real time as a result. Predictive analytics and real-time data streams will also be incorporated, helping to better manage peak demand and reduce energy waste, ultimately enhancing the smart grid's sustainability. In the long run, this technology will assist power firms in making better grid management decisions and managing the distribution of their energy more effectively.

MILESTONES AND TIMELINE:

Project Duration: September 23rd – November 30th
Sprint Duration: 1 Week (10 Sprints Total)

Sprint 1 (Sept 23 – Sept 29): Project Setup and Initial Research

- **Objectives:**
  - Define project goals, objectives, and deliverables.
  - Research energy forecasting techniques and machine learning models.

- Set up project management tools – GitHub, Google Collaboratory or VS Code, MS Excel and Power Bi.

- **Deliverables:**

  - Project plan and timeline.
  - Research documentation on energy forecasting and relevant machine learning methods.
  - Tools and workspace setup.

Sprint 2 (Sept 30 – Oct 6): Data Preprocessing

- **Objectives:**

  - Clean and preprocess the collected data (remove duplicates, handle missing values, normalize data).
  - Split the data into training and test sets.
  - Conduct exploratory data analysis (EDA) to identify trends and patterns.

- **Deliverables:**

  - Preprocessed dataset.
  - EDA report (visualizations, trends, summary statistics).
  - Training and test data ready for model development.

Sprint 3 (Oct 7 – Oct 13): Feature Engineering (5)

- **Objectives:**

  - Identify relevant features for energy consumption forecasting.
  - Engineer new features based on domain knowledge and data patterns.
  - Create feature sets for different models.

- **Deliverables:**
  - Feature sets prepared for model training.

Sprint 4 (Oct 14 – Oct 20): Model Selection and Baseline Model Implementation (3)

- **Objectives:**
  - Research different machine learning models suitable for time-series forecasting - Linear Regression, Random Forest Regressor, and Gradient Boosting (XG Boost)
  - Implement baseline models for comparison.
  - Set up evaluation metrics - MAE, RMSE.
- **Deliverables:**
  - Baseline models implemented
  - Evaluation metrics document.
  - Initial model performance results.

Sprint 5 (Oct 21 – Oct 27): Advanced Model Training and Optimization

- **Objectives:**
  - Train advanced machine/deep learning models – ARIMA, LSTM. (3)
  - Tune hyperparameters using grid search or other optimization techniques.
  - Evaluate model performance on test data.
- **Deliverables:**
  - Trained advanced models.
  - Hyperparameter tuning results.
  - Performance metrics for advanced models.

Sprint 6 (Oct 28 – Nov 3): Model Validation and Testing

- **Objectives:**
  - Perform cross-validation on models to ensure accuracy and robustness.
  - Test models on out-of-sample data or unseen data to check generalization.
  - Analyze errors and refine models as needed.
- **Deliverables:**
  - Cross-validation results.
  - Final test results and error analysis.
  - Refined models.

Sprint 7 (Nov 4 – Nov 10): Model Comparison and Selection (4)

- **Objectives:**
  - Compare all models based on performance metrics, complexity, and efficiency.
  - Select the best-performing model(s) for the forecasting task.
  - Prepare documentation on the model selection process.
- **Deliverables:**
  - Model comparison report.
  - Selected model(s) finalized for deployment.

Sprint 8 (Nov 11 – Nov 17): Model Performance Enhancement (7)

- **Objectives:**
  - Fine-tune the selected model(s) to enhance performance (reduce error, improve speed).
  - Optimize the model for production-scale forecasting (if applicable).

- Conduct additional testing on specific scenarios or edge cases.

- **Deliverables:**

  - Final optimized model.
  - Model performance enhancement report.

Sprint 9 (Nov 18 – Nov 24): Final Documentation and Report Preparation

- **Objectives:**

  - Document the entire project process, including data preprocessing, model development, validation, and results.
  - Prepare visualizations for model performance and energy forecasting outcomes.
  - Draft the final project report for submission.

- **Deliverables:**

  - Complete project documentation.
  - Visualizations (graphs, charts).
  - Draft of final project report.

Sprint 10 (Nov 25 – Nov 30): Final Presentation and Project Wrap-Up

- **Objectives:**

  - Prepare and finalize the project presentation.
  - Review the entire project for any final adjustments or improvements.
  - Conduct a project retrospective and gather feedback.

- **Deliverables:**

  - Final project presentation slides.
  - Final project report.
  - Project retrospective document.

PROJECT METHODOLOGY:

DATASET:

The dataset utilized for this project is sourced from Kaggle and is titled "Electricity Consumption UK 2009-2024.

Link-https://www.kaggle.com/datasets/albertovidalrod/electricity-consumption-uk-20092022?resource=download

DATASET ARCHITECTURE:

The dataset consists of the following key columns:

- **settlement date**: Indicates the specific date of the recorded energy data, useful for time-series analysis.
- **settlement period**: Represents time intervals (typically 30 minutes) within each day for detailed consumption analysis.
- **Nd (National Demand)**: Total electricity demand across the country at any given time, crucial for balancing supply and demand.
- **tsd (Transmission System Demand)**: Demand connected to the transmission grid, excluding local generation, offering insights into grid load.
- **England Wales demand**: Electricity demand specifically for England and Wales, enabling regional consumption analysis.
- **Embedded wind generation**: Amount of electricity generated by localized wind farms, highlighting contributions from small-scale renewables.

- **Embedded wind capacity**: Total capacity of embedded wind systems, important for evaluating utilization rates.
- **Embedded solar generation**: Electricity generated by localized solar systems, reflecting solar energy contributions.
- **Embedded solar capacity**: Maximum theoretical output from solar systems, aiding in capacity planning.
- **Non bm stored**: Amount of electricity stored in facilities outside the national grid's balancing mechanism.
- **Pump storage pumping**: Energy consumption of pumped storage systems, critical for managing excess energy.
- **interconnector flows (e.g., ifa flow)**: Represents electricity flow between the UK and neighboring countries, assessing energy imports and exports.
- **Is holiday**: Indicates if the date falls on a public holiday, important for analyzing variations in consumption patterns.

TECHNOLOGIES OR SERVICES:

- **Programming**: Python
- **Data Storage**: Azure Cloud Storage
- **Platforms**: Google Collab, Excel
- **Machine Learning**: Scikit-learn (Random Forest, Linear Regression)
- **Time-Series Analysis**: Pandas, NumPy
- **Data Visualization**: Matplotlib, Seaborn
- **Evaluation Metrics**: Scikit-learn (MAE, RMSE)

BUDGET

Category	Details	Cost Range (CAD)
Software	Python, Pandas, Scikit-learn, ARIMA, Power Bi	$0 - $94
Data Storage (Optional)	AWS S3, Google Cloud Storage	$0 - $13
Research Materials	Books, Journals, Papers	$20 - $100
Presentation/Printing	Project Presentation Materials, Printing	$20 - $60
Miscellaneous	Unforeseen Costs	$70
Total Estimated Budget		**$20 - $200**

REFERENCE:

1. https://www.geeksforgeeks.org/ml-linear-regression/
2. https://www.mdpi.com/2071-1050/15/9/7087
3. How to Create an ARIMA Model for Time Series Forecasting in Python - MachineLearningMastery.com
4. Short-Term Firm-Level Energy-Consumption Forecasting for Energy-Intensive Manufacturing: A Comparison of Machine Learning and Deep Learning Models (mdpi.com)
5. Forecasting Renewable Energy Generation with Machine Learning and Deep Learning: Current Advances and Future Prospects (mdpi.com)
6. Thanksgiving holiday causes unique electricity usage patterns across the country - U.S. Energy Information Administration (EIA)
7. Multi-Time-Scale Energy Storage Optimization Configuration for Power Balance in Distribution Systems (mdpi.com)

CONCLUSION:

The goal of the "Energy Consumption Forecasting with Machine Learning for Smart Grid" project is to create a robust forecasting model that predicts energy demand from historical data using the ARIMA algorithm. This model will aid in optimizing energy distribution, lowering operating costs, and advancing sustainable practices within smart grid systems. Insights into energy consumption patterns that may be used will be made possible by the project's successful completion, allowing for more effective resource allocation, demand-response plans, and general energy management.

Additionally, this study lays the framework for future improvements, like using more sophisticated machine learning methods to increase forecast accuracy and broaden its application to other energy systems. By taking these steps, we aim to play a key role in the advancement of smart grid technologies and contribute to the future of sustainable energy management.