**7PAM2002-0901-2024 - Data Science Project**

**TITLE: Predicting Peak Electricity Usage to Reduce Energy Costs**

**Student name & ID: Aishwarya Sukumaran (22058088)**

**Email: as23adl@herts.ac.uk**

**1. Project Overview**

**Project Summary:**

This project focuses on predicting **peak electricity usage** in households to facilitate cost savings through optimized energy consumption. By leveraging the **Individual Household Electric Power Consumption dataset**, which comprises four years of minute-level electricity usage data, the project aims to identify patterns and trends in electricity consumption. Using machine learning techniques, specifically **Linear Regression and** **Random Forest Regression**, the model will forecast peak usage times across weekly, monthly, and yearly intervals. This information will empower households to shift energy-intensive activities to off-peak times, ultimately reducing energy costs and promoting sustainable energy usage.

**Research Question:**  
How can machine learning algorithms, be effectively utilized to predict **peak electricity usage** in households based on historical data, and what strategies can be developed to optimize energy consumption during high usage periods?

**Project Objectives:**

1. **Data Preprocessing:** Clean and preprocess the dataset by handling missing values, resampling the minute-level data to daily, weekly, and monthly intervals, and generating relevant time-based features (e.g., hour of the day, day of the week, season).
2. **Exploratory Data Analysis (EDA):** Conduct EDA to explore patterns and seasonal trends in electricity usage, identifying peak periods throughout different timeframes (daily, weekly, monthly, yearly).
3. **Model Development:** Build predictive models using Linear Regression and Random Forest Regression to forecast electricity usage, focusing on capturing peak usage times.
4. **Model Evaluation:** Assess the model’s accuracy with metrics like Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to ensure reliable predictions of peak periods.
5. **Insights and Recommendations:** Provide actionable insights for households, such as shifting high-energy activities to off-peak hours to maximize savings on electricity costs.

**Reference List:**

1. Gupta, R., & Mehta, A. (2019). Optimizing electricity consumption: A machine learning approach to peak load forecasting. *Energy Reports*, 5, 1234-1245. [Link](https://www.sciencedirect.com/science/article/pii/S2352484719301998)
2. Saha, S., & Ghosh, S. (2020). Predicting peak electricity demand using machine learning techniques. *IEEE Transactions on Power Systems*, 35(2), 1072-1081. [Link](https://ieeexplore.ieee.org/document/8940957)
3. Brown, T., & Green, M. (2021). A data-driven approach to forecasting electricity consumption and identifying peak usage periods. *Renewable Energy*, 167, 267-279. [Link](https://www.sciencedirect.com/science/article/pii/S0960148121001959)

**2. Project Plan: Task List and Timeline**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Week** | **Task** | **Objective** | **Activities** | **Milestone/**  **Assessment** |
| **Week 1 (Sep 25 - Oct 1)** | **Project Planning & Initial Research** | Define project scope and methodology | - Finalize project objectives and research question  - Identify relevant datasets  - Conduct initial literature review to frame the project | Project proposal completed |
| **Week 2 (Oct 2 - Oct 15)** | **Literature Review** | Analyse existing research on sales forecasting | - Search for at least 5 peer-reviewed articles  - Summarize key findings and methodologies  - Prepare a literature review draft | Literature review draft completed |
| **Week 3 (Oct 16 - Oct 31)** | **Data Collection & Preprocessing** | Prepare and clean the dataset for analysis | - Load the dataset and inspect its structure  - Handle missing values and outliers  - Convert data types (e.g., dates)  - Conduct exploratory data analysis (EDA) | Cleaned dataset ready for analysis |
| **Week 4 (Nov 1 - Nov 15)** | **Model Selection & Training** | Build and train machine learning models | - - Research different regression models.  - Split data into training and testing sets  - Train models on training data | Trained models |
| **Week 5 (Nov 16 - Nov 22)** | **Model Evaluation & Optimization** | Evaluate model performance and optimize | - Assess models using MAE and RMSE metrics  - Compare performance of different models  - Optimize the best-performing model by tuning parameters | Final model selected |
| **Week 6 (Nov 23 - Nov 29)** | **Final Report Preparation & Documentation** | Compile findings and prepare the final report | - Document methodology, results, and analysis  - Create visualizations of key findings  - Review and revise the report for clarity and coherence | Final report completed |
| **Week 7 (Nov 24 - Dec 1)** | **Project Presentation Preparation** | Prepare to present the project findings | - Create presentation slides summarizing key results and methodologies  - Rehearse presentation delivery  - Seek feedback from peers or advisors | Final presentation completed |

**3. Data Management Plan**

**Overview of the Dataset:**The dataset "Individual Household Electric Power Consumption" is collected from a smart meter deployed in a household located in France. The dataset includes measurements of electrical power consumption, recorded in kilowatt-hours (kWh) at 1-minute intervals over a four-year period, providing a comprehensive view of electricity usage patterns.

**Data Collection:  
The data can be accessed from the UCI Machine Learning Repository at the following link:** [**UCI Electric Power Consumption Dataset**](https://archive.ics.uci.edu/dataset/235/individual+household+electric+power+consumption)**.**

**Metadata:**

* Format: CSV
* Number of Records: Approximately 2 million rows
* Expected Size: ~200 MB

**Document Control:**

* GitHub Address: [GitHub Repository](https://github.com/AishwaryaSukumaran/Final-Project_22058088)
* Commit Frequency: Weekly commits to track code changes and updates.
* File Naming: Descriptive names (e.g., energy\_data.py, model\_results.py) and version control (e.g., v1.0, v1.1).

**ReadMe File:**The ReadMe file will include a brief project overview, instructions for running the code, dependencies required, and a summary of key findings. It will help others understand the project and replicate the results.

**Security and Storage:**

* Backups will occur weekly, both on GitHub and on a cloud service (e.g., OneDrive) to ensure data safety.
* Data will be shared securely with staff and markers through the GitHub repository link, ensuring appropriate access control.

**Ethical Requirements:**

1. GDPR Compliance: The dataset is anonymized, and no personal data is included, ensuring compliance with GDPR regulations.
2. Conformity to Ethical Policies: The project adheres to ethical guidelines set by the University of Hertfordshire, ensuring responsible data usage.
3. Permission to Use Data: The dataset is publicly available for research purposes, providing implied consent for its use.
4. Ethical Data Collection Assurance: The data was collected by the original organization (UCI), ensuring ethical standards were maintained during its collection.