Energy Consumption of Houshold Appliances

Harshilkumar Soni
Master of Engineering in Systems and Technology
Hamilton, Canada
sonih5@mcmaster.ca

Karan Patel
Master of Engineering in Systems and Technology
Hamilton, Canada
patelk96@mcmaster.ca

Dhruv Upadhyay
Master of Engineering in Systems and Technology
Hamilton, Canada
upadhd3@mcmaster.ca

Dhruv Padia
Master of Engineering in Systems and Technology
Hamilton, Canada
padiad@mcmaster.ca

***Abstract*— This paper presents a study that analyzes household appliance energy consumption patterns by integrating smart home technology and weather data. The goal is to better understand and optimize energy use in homes. Using a dataset spanning 350 days with minute-by-minute records of temperature, humidity, and appliance energy consumption, the study explores how environmental conditions impact energy usage. The main objective is to create accurate regression models that predict appliance-specific energy patterns, using data from various sensors. This work contributes to energy conservation and smart home technology, providing insights for more efficient energy practices. The findings highlight the potential of data integration and predictive modeling for a sustainable future.**

I. Introduction

This document serves as a comprehensive exposition of a formal project aimed at discerning and analyzing household appliance energy consumption patterns through the integration of advanced smart home technology and meticulously collected weather data. The central objective of this endeavor is to conduct an exhaustive analysis of energy utilization trends across a spectrum of household appliances, facilitated by a meticulously curated dataset enriched with real-time meteorological information. This dataset encompasses a time frame of 350 days, providing minute-by-minute granularity essential for comprehending the intricate dynamics of energy consumption. Noteworthy parameters, including indoor temperature, humidity levels, and precise energy consumption data for a diverse array of appliances, are captured.

The principal aspiration of this project is the development of robust regression models calibrated to predict energy usage patterns specific to individual appliances. These models are constructed using data gathered from an array of sensors, promising an innovative perspective on the complex interplay between domestic environments, technological ecosystems, and the variable influences of weather phenomena.

II. Problem Statement

At the crux of this project lies a fundamental problem statement: How can the intricate relationship between household appliance energy consumption patterns and the confluence of weather conditions and indoor environmental variables be effectively characterized and modelled? The project addresses the intricate challenge of unravelling the multifaceted dependencies that govern energy utilization trends, offering a proactive avenue for managing energy resources within a domestic context. By devising accurate regression models that encapsulate the intertwined effects of temperature, humidity, and appliance-specific attributes, the project endeavors to furnish homeowners with informed insights, thereby empowering judicious practices in energy consumption. This formal endeavor holds the potential to contribute substantively to the domains of energy conservation and smart home technology implementation.

III. Project Objectives

A. Thoroughly Load and Analyse the Smart Home Dataset in Conjunction with Weather Information to Extract Profound Insights

Explore the integrated dataset meticulously, aiming to extract comprehensive insights that illuminate the intricate interplay between household appliance energy consumption and varying weather conditions.

B. Utilize Effective Data Visualization Techniques to Scrutinize Dataset Characteristics and Decipher Complex Appliance Energy Consumption Trends

Employ advanced visualization methods to dissect the dataset's inherent patterns, comprehending the distribution of data and unravelling the nuanced energy consumption trends among diverse appliances.

C. Execute Robust Pre-Processing Protocols to Manage Missing Data and Normalize Dataset Attributes, Paving the Way for Effective Model Construction

Apply meticulous pre-processing techniques, adeptly managing any missing data points, while concurrently normalizing dataset attributes to establish an optimal foundation for subsequent modelling endeavors.

D. Construct and Fine-Tune a Suite of Diverse Machine Learning Models, including Linear Regression, Decision Trees, and Random Forests, to Accurately Predict Appliance Energy Consumption

Ingeniously formulate, calibrate, and fine-tune a range of machine learning models, encompassing Linear Regression, Decision Trees, and Random Forests, with the primary objective of attaining precise predictions of appliance energy consumption.

E. Architect and Execute an Advanced Feedforward Neural Network (FNN) Model to Provide an Intriguing Comparative Analysis with Traditional Models

Engineer a sophisticated Feedforward Neural Network (FNN) model, implementing it meticulously to facilitate a captivating comparative exploration against conventional machine learning models, enriching the understanding of predictive capabilities.

F. Conduct Rigorous Performance Assessment and Metric-Based Comparison of Diverse Models to Ascertain Optimal Predictive Accuracy

Engage in a comprehensive evaluation process, meticulously assessing the performance of the diverse models, guided by appropriate metrics, to discern the model or models that best encapsulate appliance energy consumption patterns.

IV. Theory and Datasets

A. Therory

In the realm of cyber-physical systems, the fusion of intelligent technologies and real-time data has brought about revolutionary capabilities to comprehend and enhance various processes, including the energy consumption dynamics within households. The integration of smart home technology, complete with sensors and interconnected devices, has paved a path for intricate data gathering and analysis within domestic settings. This harmonious convergence of technology and data serves as the foundational theory underpinning our project, enabling us to untangle the intricate interrelationships between energy usage of household appliances and external factors, such as prevailing weather conditions.

Central to our analysis are regression models, pivotal tools that enable the anticipation of energy consumption patterns through a fusion of multiple variables. Alongside traditional methods like Linear Regression, Decision Trees, and Random Forests, we extend our exploration to advanced machine learning models like the Feedforward Neural Network (FNN), empowering us to delve into the intricate intricacies of data dependencies and interactions.

B. Datasets

Our project's foundation rests upon a meticulously curated dataset sourced from Kaggle, a prominent platform for data science enthusiasts and researchers. This dataset encompasses a span of 350 days and offers a detailed account of energy consumption patterns in households, captured at a 1-minute interval. Comprising readings from smart meters and enriched with weather information for a specific region, this dataset presents a holistic view of energy consumption dynamics within real-world settings. The dataset is provided in CSV format, allowing for easy access and manipulation using data processing tools and programming languages. It includes a multitude of attributes, with each row representing a minute of recorded data. Key attributes encompass:

- Timestamp: We have ranged the data from 01/01/2016, 05:00 AM for better understanding.
- Energy Consumption: The energy consumed by various household appliances during the specified minute in kW.
- Indoor Temperature: The indoor temperature of the household during that minute.
- Humidity: The humidity level within the household at that particular time.
- Weather Data: Weather-related information such as outdoor temperature, humidity, and other meteorological conditions for the specific region.

The integration of indoor environmental conditions and external weather data provides a unique opportunity to explore the intricate interplay between these factors and appliance energy consumption. By analyzing this dataset, we aim to uncover patterns, correlations, and insights that contribute to a more profound understanding of household energy utilization.

V. Project Implementation

- Data Analysis and Visualization: Load the smart home dataset and weather information using Python and Pandas. Perform exploratory data analysis (EDA) to identify trends and correlations. Create visualizations (line plots, scatter plots, etc.) to visualize appliance energy usage patterns. Analyze the correlation between weather conditions and energy consumption of appliances.
- Data Pre-processing: Handle missing or erroneous data points through imputation or removal. Normalize the data to ensure consistent scales for different features. Split the dataset into training and testing sets for model evaluation.
- Regression Models: Implement Linear Regression, Decision Trees, and Random Forests models using scikit-learn. Perform hyperparameter tuning for each model to optimize performance. Evaluate the regression models using metrics such as Mean Squared Error (MSE) and R-squared.
- Feedforward Neural Network (FNN) Model: Design a Feedforward Neural Network architecture using TensorFlow/Karas. Experiment with different activation functions and layer configurations. Train the FNN model using the pre-processed dataset and validate its performance.
- Feature Importance Analysis: Identify the most relevant features for accurate energy consumption predictions. Use techniques like Recursive Feature Elimination (RFE) or feature importance scores from models. Provide insights into which features or data combinations are crucial for specific appliance predictions.

VI. Explanation of Sourcecode

A. Step 1

Importing the necessary libraries and importing the csv file

```
In [1]: import pandas as pd
        import matplotlib.pyplot as plt
        %matplotlib inline
        import seaborn as sns
        from sklearn.model_selection import train_test_split
        from sklearn.linear_model import LinearRegression
        from sklearn.metrics import mean_squared_error
        from sklearn.preprocessing import StandardScaler
        import numpy as np
```

```
In [2]: data = pd.read_csv('HomeC.csv', low_memory=False)
        data.head()
```

Figure 1

B. Step 2

Knowing the datatypes and removing the data row with garbage values

```
In [ ]: data.info()
```

```
In [ ]: data.tail()
```

```
In [ ]: data=data[:-1]
        data.tail()
```

Figure 2

C. Step 3

Converting the time values and ranging it from 01-01-2016, 5:00 AM

```
In [ ]: #given that the time is in UNIX format, let's check
        time_index = pd.date_range('2016-01-01 05:00', periods=len(data),  freq='min')
        time_index = pd.DatetimeIndex(time_index)
        data = data.set_index(time_index)
```

Figure 3

D. Step 4

Combining the similar equipment readings and checking again if the values are in required format.

```
In [ ]: data.columns
```

```
In [ ]: data.columns = [col.replace(' [kW]', '') for col in data.columns]
        data.columns
```

```
In [ ]: data['Furnace'] = data[['Furnace 1','Furnace 2']].sum(axis=1)
        data['Kitchen'] = data[['Kitchen 12','Kitchen 14','Kitchen 38']].sum(axis=1)
        data.drop(['Furnace 1','Furnace 2','Kitchen 12','Kitchen 14','Kitchen 38','ico
        data.head(5)
```

```
In [ ]: data.info()
```

```
In [ ]: data['cloudCover'].unique()
```

```
In [ ]: data['cloudCover'].replace(['cloudCover'], method='bfill', inplace=True)
        data['cloudCover'] = data['cloudCover'].astype('float')
```

Figure 4

E. Step 5

Plotting the data to the know the relation each data with time and neglecting the non-impacting parameter.

```
In [ ]: energy_per_day = energy_data.resample('D').sum()
        plt.figure(figsize=(20,10))
        plt.title("Overall energy consumption per day")
        sns.lineplot(data = energy_per_day.filter(items=[ 'Dishwasher', 'Home office',
        #'Dishwasher', 'Furnace', 'Home office', 'Fridge','Wine cellar', 'Garage door'
```

```
In [ ]: energy_per_day = energy_data.resample('D').sum()
        plt.figure(figsize=(20,10))
        plt.title("Overall energy consumption per day")
        sns.lineplot(data = energy_per_day.filter(items=[ 'Barn', 'Well', 'Garage door
        #'Dishwasher', 'Furnace', 'Home office', 'Fridge','Wine cellar', 'Garage door'
```

```
In [ ]: energy_per_month = energy_data.resample('M').sum() # for energy we use sum to
        plt.figure(figsize=(20,10))
        sns.lineplot(data= energy_per_month.filter(items=[  'Dishwasher', 'Furnace', '
                                              'Wine cellar', 'Garage door', 'Kitchen',
                                              'Microwave', 'Living room']) , dashes=Fal
```

Figure 5

F. Step 6

Knowing the relationship between sensor data and consumption by plotting the graph.

```
In [ ]: sns.regplot(x=energy_per_day['Kitchen'], y= sensor_per_day['temperature'])
```

```
In [ ]: sns.regplot(x=energy_per_day['Fridge'], y= sensor_per_day['temperature'])
```

```
In [ ]: sns.regplot(x=energy_per_day['Furnace'], y= sensor_per_day['temperature'])
```

Figure 6

G. Step 7

Feature Importance, scaling down the values from actual to values between 0 & 1 and plotting the value.

```
In [ ]: # Standardize the features
        scaler = StandardScaler()
        X_train_scaled = scaler.fit_transform(X_train)
        X_test_scaled = scaler.transform(X_test)
```

```
In [ ]: # Linear Regression Model
        model_lr = LinearRegression()
        model_lr.fit(X_train_scaled, y_train)
```

```
In [ ]: y_pred_lr = model_lr.predict(X_test_scaled)
```

```
In [ ]: mse_lr = mean_squared_error(y_test, y_pred_lr)
```

```
In [ ]: feature_importance_lr = np.abs(model_lr.coef_)
        feature_names = X.columns

        # Visualize feature importance
        plt.figure(figsize=(12, 6))
        plt.barh(feature_names, feature_importance_lr)
        plt.xlabel('Feature Importance')
        plt.ylabel('Features')
        plt.title('Linear Regression - Feature Importance')
        plt.show()
```

Figure 7

H. Step 8

After feature importance, training the 80% of the total data on different algorithms and mapping the accuracy of each model and then testing on 20% of the data.

```
In [ ]: # Random Forest Regressor
        model_rf = RandomForestRegressor(n_estimators=100, random_state=42)
        model_rf.fit(X_train_scaled, y_train)
        y_pred_rf = model_rf.predict(X_test_scaled)
        r2_rf = r2_score(y_test, y_pred_rf)
        print(f"Random Forest R-squared: {r2_rf}")
```

```
In [ ]: # Support Vector Regressor
        model_svr = SVR()
        model_svr.fit(X_train_scaled, y_train)
        y_pred_svr = model_svr.predict(X_test_scaled)
        r2_svr = r2_score(y_test, y_pred_svr)
        print(f"SVR R-squared: {r2_svr}")
```

```
In [ ]: # K-Nearest Neighbors Regressor
        model_knn = KNeighborsRegressor()
        model_knn.fit(X_train_scaled, y_train)
        y_pred_knn = model_knn.predict(X_test_scaled)
        r2_knn = r2_score(y_test, y_pred_knn)
        print(f"K-Nearest Neighbors R-squared: {r2_knn}")
```

Figure 8

VII. Results and Discussions

Learning Algorithm	KNN Regressor	Radom Forest Regressor	Support Vector Regressor	FNN (Multi-Layer Perception)
Accuracy	91.57%	97.22%	--	94.08%

Figure 9

VIII. Recommendations for Future Work

Subsequent research endeavors may prioritize enhancing the precision of predictions and gaining deeper insights into model behavior. By incorporating advanced time-series models to capture temporal patterns, exploring ensemble techniques for enhanced forecasting, and leveraging interpretable AI methodologies for meaningful insights, we can further enrich our understanding. Furthermore, for applications focused on energy efficiency, delving into model generalization and seamless integration into real-time systems holds promising avenues for exploration.

IX. Challenges

1. Data Quality and Consistency:
   - Ensuring the accuracy and reliability of the collected dataset, particularly weather data, required rigorous cleaning and validation processes to maintain data integrity.
2. Model Complexity and Performance:
   - Developing and fine-tuning complex machine learning models like the Feedforward Neural Network (FNN) necessitated careful parameter tuning and computational resources to achieve optimal performance.
3. Interpretable Models:
   - Balancing model accuracy and interpretability remains a challenge, especially for advanced models like the FNN, where understanding the intricate relationships can be complex.

X. Conclusion

In conclusion, this study signifies a significant step forward in understanding and optimizing household energy consumption patterns. By harnessing advanced smart home technology and real-time meteorological data, the project unveiled valuable insights into the complex relationships that govern energy utilization. The developed regression models, coupled with comprehensive data analysis, provide homeowners with predictive tools for informed energy consumption decisions. This work holds substantial implications for energy conservation and smart home technology, offering a pathway toward more sustainable and efficient energy practices.

References

[1] https://www.kaggle.com/datasets/taranvee/smart-home-dataset-with-weather-information