Paavana M Rao

paavanarao30@gmail.com

Abstract

To analyse the Energy consumption of smart homes with other key features and to predict the trends and variants of features to make better future decisions

Internship

On Data science using Python

**Contents**

1. Introduction
2. Python
3. Data science
4. Problem statement
5. Dataset
6. Libraries imported
7. Solutions/Summary
8. Conclusion

## **Introduction**

In the era of digital payments, scanning QR codes, applying for a job many such basic amnesties have been on our fingertips to utilize, learn and ensure the simplicity of understanding new technologies. Data has become an essential and crucial entity in the world and has indeed become important to Understand, analyse, store the processed information effectively.

Many industries have come to utilise and learn various new technologies to maintain equilibrium such that the data stored can be understood and analysed efficiently.

Such tools namely Excel, power BI ,Tableau ,Artificial Intelligence and many more have been improvised , developed which can enable us to understand the trends and the patterns ,hence to gain more insights about the information and help us to make better major decisions

*Information* which we collect through the insights and feedbacks of externa factors which in turn processed and formed into *Data*. There are various ways the data can be stored or formed (Tables, Csv files, etc.) These formats which have collection attributes and features in a specific dimension is called *Dataset.* After clearing all the noises (outliers, missing values, null values) With this specific set of data given we *Analyse* the pattern, trends and behaviour of the entities

**Python**

Python is a dynamic, interpreted (bytecode-compiled) language. There are no type declarations of variables, parameters, functions, or methods in source code. This makes the code short and flexible, and you lose the compile-time type checking of the source code. Python tracks the types of all values at runtime and flags code that does not make sense as it runs.

Python is a widely used general-purpose, high-level programming language. It was created by Guido van Rossum in 1991 and further developed by the Python Software Foundation. It was designed with an emphasis on code readability, and its syntax allows programmers to express their concepts in fewer lines of code. Python is a programming language that lets you work quickly and integrate systems more efficiently. There are two major Python versions: Python 2 and Python 3. Both are quite different.

**Data science**

Data science is the domain of study that deals with vast volumes of data using modern tools and techniques, including essential data science skills, to find unseen patterns, derive meaningful information, and make business decisions. Data science uses complex machine learning algorithms to build predictive models. The data used for analysis can come from many different sources and presented in various formats.

Data science is an essential part of many industries today, given the massive amounts of data that are produced, and is one of the most debated topics in IT circles. Its popularity has grown over the years, and companies have started implementing data science techniques to grow their business and increase customer satisfaction.

Fig:Data Analysis

**Objective/Problem statement**

To analyse patterns in energy consumption, understand the impact of environmental factors, and build predictive models.

Dataset Description

➢ Energy\_Consumption\_kWh: The total energy consumed(in

kilowatt-hours).

➢ Temperature\_C: The temperature in degrees Celsius.

➢ Humidity\_%: The relative humidity percentage.

➢ HVAC\_Usage\_kWh: The energy consumption of the HVAC(Heating,Ventilation, and Air Conditioning) system.

➢ Kitchen\_Usage\_kWh:The energy consumption in the kitchen.

➢ Electronics\_Usage\_kWh: The energy consumption of electronic devices.

➢ Occupancy: The number of occupants in the home.

➢ Weather Conditions: Categorical data describing the weather (e.g., Sunny,

Rainy).

➢ City: The name of the Indian city where the home is located

*Dataset:[..\Downloads\smart\_home\_energy\_consumption.csv](C:\\Users\\Intel\\Downloads\\smart_home_energy_consumption.csv)*

**Libraries Importe**d:

* Numpy
* Pandas
* Matplotlib
* Seaborn
* Sklearn
* Plotly

**Summary/Solutions**

***Missing /Null values***

Identifying the missing values and NaN values is very important step in the Data cleaning process

To identify missing values present in the data we used info()

Used a function called describe() which gives us the min, mean ,variance , standard deviation of the dataset .

By replacing mean() of the dataset to all the Nan values we have removed the noises

***Outlier Detection***

**Box plots** are used to detect outliers in columns like Energy\_Consumption\_kWh and Temperature\_C:

***Data Visualization***

**Histograms** and **line plots** are used for visualizing energy consumption, temperature, and time-based trends. **Pairplot** for visualizing relationships between multiple variables**: Jointplots** for scatter plots of energy consumption vs. temperature and humidity

***Time Series Analysis***

The data is visualized over time to understand patterns in energy consumption.Resampling techniques are applied to observe daily and weekly energy consumption trends.

***Feature Engineering***

**Resampling energy consumption** to daily and weekly sums*.* **Plotting temperature difference over time.** **Visualization techniques like joint plots and pair plots are used to understand the correlation between features (e.g., temperature and energy usage).**

***Modelling***

**Linear Regression** using sklearn and **Random Forest Regressor is applied** for continuous data prediction due to its robustness

***Model Evaluation***

**Mean Squared Error** and **R² Score** are calculated to evaluate the model, providing insights into how well the model fits the data.

**Conclusion**

* To handle a time series dataset and visualize key relationships between variables to gain insights into energy consumption patterns.
* The Random Forest Regressor is identified as a more accurate model compared to linear regression for predicting energy consumption, based on error metrics and model performance.
* By exploring the impact of external factors like temperature and humidity and shows the importance of feature engineering and visualization in improving predictive models.
* It provides a strong foundation for optimizing energy usage in smart homes by predicting future consumption based on historical data and environmental factors.