Project Report

Energy Consumption Prediction using Linear Regression

1. Introduction

Energy consumption is a critical factor in building management, urban planning, and sustainable development. Accurately predicting energy usage enables organizations to improve efficiency, reduce costs, and implement effective energy conservation strategies.

This project focuses on building a Linear Regression model to predict energy consumption using various building and environmental features. The approach emphasizes data preprocessing, feature selection, model training, and performance evaluation to ensure robust and accurate predictions.

2. Objectives

- To preprocess the dataset and prepare it for modeling.
- To develop a predictive model using Linear Regression.
- To evaluate the model's performance using statistical metrics.
- To analyze the predictive power of features influencing energy consumption.

```
Building Type           0
Square Footage          0
Number of Occupants     0
Appliances Used         0
Average Temperature     0
Day of Week             0
Energy Consumption      0
```

```
Building Type           0
Square Footage          0
Number of Occupants     0
Appliances Used         0
Average Temperature     0
Day of Week             0
Energy Consumption      0
```

3. Dataset and Preprocessing

The dataset consisted of features related to building specifications, occupancy, and environmental conditions. To ensure high-quality inputs, the following preprocessing steps were performed:

```
   Square Footage  Number of Occupants  Appliances Used  Average Temperature  \
0       -1.287805             0.951134        -1.106957             1.012925
1        1.323516             0.606869         1.375645            -0.825544
2       -0.434466            -0.391498        -0.610436            -1.164652
3       -0.853716            -1.183306         1.091919             1.430504
4       -0.846017            -0.770188        -0.539505            -1.498155
```

	Building Type_Industrial	Building Type_Residential	Day of Week_Weekend
0	False	True	False
1	False	False	False
2	True	False	True
3	False	True	False
4	False	False	False

0	2713.95
1	5744.99
2	4101.24
3	3009.14
4	3279.17

	Square Footage	Number of Occupants	Appliances Used	Average Temperature	\
0	-0.062950	-1.148879	-1.532546	0.827957	
1	0.148425	0.262605	-0.184847	0.064264	
2	1.389378	-1.527570	1.304714	1.534198	
3	1.131249	1.226545	-0.610436	0.669613	
4	0.787940	0.331458	1.517508	-0.775098	

	Building Type_Industrial	Building Type_Residential	Day of Week_Weekend
0	False	True	False
1	False	False	True
2	False	False	False
3	False	True	True
4	False	True	False

0	2865.57
1	4283.80
2	5067.83
3	4624.30
4	4820.59

1. Handling Missing Values – Missing data points were appropriately treated to maintain data consistency.
2. One-Hot Encoding – Categorical variables such as Building Type and Day of Week were transformed into numerical representations.
3. Feature Scaling – Numerical features were scaled for uniformity: Square Footage, Number of Occupants, Appliances Used, Average Temperature.

4. Methodology

The project employed the Linear Regression algorithm, a fundamental statistical method for modeling the relationship between dependent and independent variables.

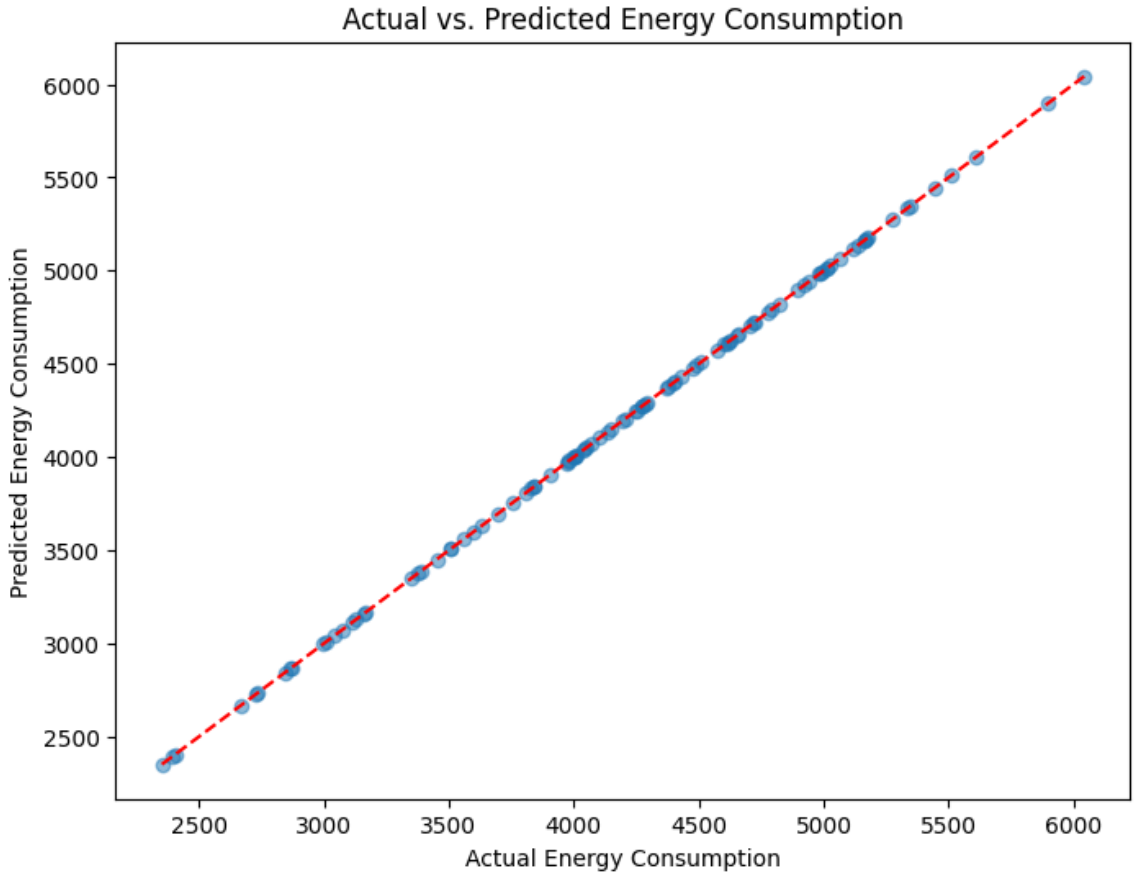
Steps followed:
1. Splitting the dataset into training and testing sets.
2. Fitting the Linear Regression model on the training data.
3. Predicting energy consumption on the testing dataset.
4. Evaluating predictions using performance metrics.

A scatter plot of Actual vs. Predicted Energy Consumption was also created to visually confirm the model's accuracy.

5. Results and Analysis

The trained Linear Regression model demonstrated exceptional accuracy:

- Mean Squared Error (MSE): 0.0002015
- R-squared Score (R^2): 0.9999999997



These results indicate that the model explains nearly 100% of the variance in energy consumption. The scatter plot confirmed this finding, as predicted values closely aligned with actual energy consumption points.

6. Discussion

The outstanding performance can be attributed to the strong correlation between the selected features and energy consumption. For example, Square Footage and Number of Occupants directly influence energy demand, while Average Temperature affects heating and cooling needs.

Although Linear Regression worked well in this case, future work could explore:
- Feature Importance Analysis – Examining coefficients to interpret influence of individual

variables.
- Regularization Techniques – Using Ridge or Lasso regression to prevent overfitting.
- Comparative Modeling – Testing non-linear models such as Random Forest or Gradient Boosting.

7. Conclusion

This project successfully developed a Linear Regression model that accurately predicts energy consumption. The preprocessing pipeline ensured data quality, and the model achieved extremely low error rates and an almost perfect R^2 score.

The findings highlight that building characteristics and environmental features play a crucial role in energy prediction. This model can serve as a reliable tool for decision-making in building energy management and lays the foundation for more advanced predictive modeling in the future.

8. References

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning.
- Scikit-learn documentation: https://scikit-learn.org
- Pandas documentation: https://pandas.pydata.org