Regression Task Report:

Predictive Modeling of Global Energy Consumption via Climate and Industrial Metrics

Student ID: 2462282

Student Name: Kraneel Manandhar

Course name: Concept and Technologies of AI

Date of submission: 10th Feb 2026

Table of Contents:

Abstract:

The scope of this report is to use the regression techniques to forecast the percentage of energy consumption according to environment and industrial variables. This paper is correlated with UNSDG Goal 7 (Affordable and Clean Energy) and Goal 13 (Climate Action) using Global Climate and Energy dataset (2020-2024). The presented methodology includes Exploratory Data Analysis (EDA), the creation of a Multi-Layer Perceptron (MLP) Regressor, and the comparison of Ridge Regression and Random Forest Regressor. The models were optimized using the hyper-parameter tuning algorithm, GridSearchCV, and dimensionality reduction, Recursive Feature Elimination (RFE). Findings indicate that the most accurate prediction of energy demand was based on the Random Forest Regressor.

Introduction:

1.1 Description:

This task has a major task of predicting energy consumption needs. There must be accurate forecasting in grid stability and integration of renewable energy sources. Knowing the influence of variables such as temperature and industrial activity on demand, the policymakers can achieve greater control over energy distribution.

1.2 Selected Dataset:

The dataset has been obtained at the Environmental Data Research Initiative which has global metrics between 2020 and 2024. It gives variables like avg_temperature, co2_emission and industrial_activity_index.

1.3 Objective:

The main aim of the analysis is to formulate a superior predictive regression model that can forecast the continuous energy consumption levels using climate and industrial indicators. The paper will use machine learning algorithms to measure the correlation between environmental changes and power demand, which will offer a data-driven approach to sustainable energy management. In particular, the analysis will seek to maximize the accuracy of the model by hyper-parameter optimization, feature selection to guarantee dependable predictions in accordance with the global sustainability objectives.

2. Methodology

Exploratory Data Analysis (EDA):

The EDA step aimed at seasonal patterns and the relationship between the growth of industry and energy peaks. The data was merged to a period of 3 months to determine the cyclical patterns. There was a great positive correlation between the industrial_activity_index and the target variable, energy_consumption.

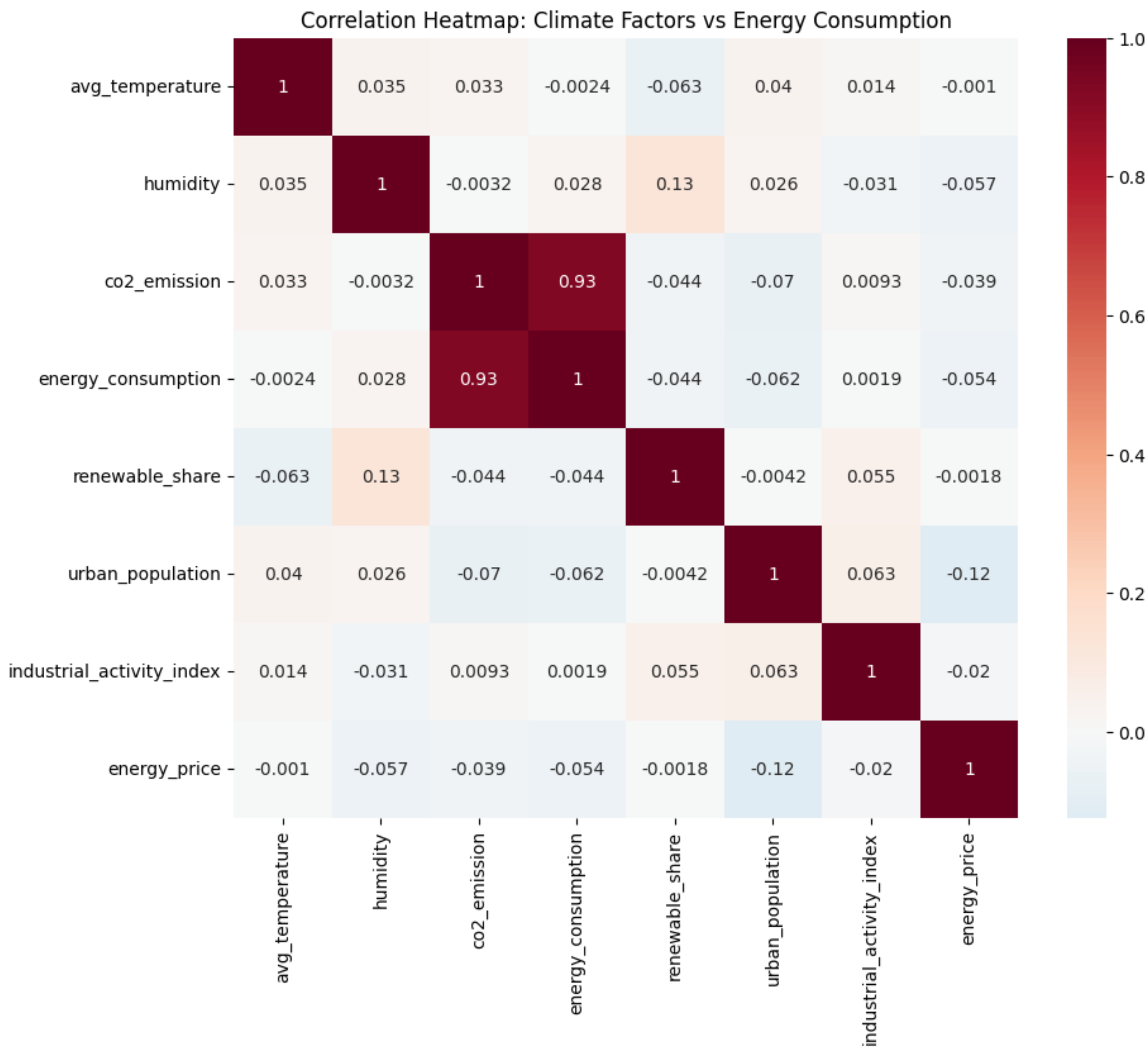

Figure 1: Corelation Heatmap

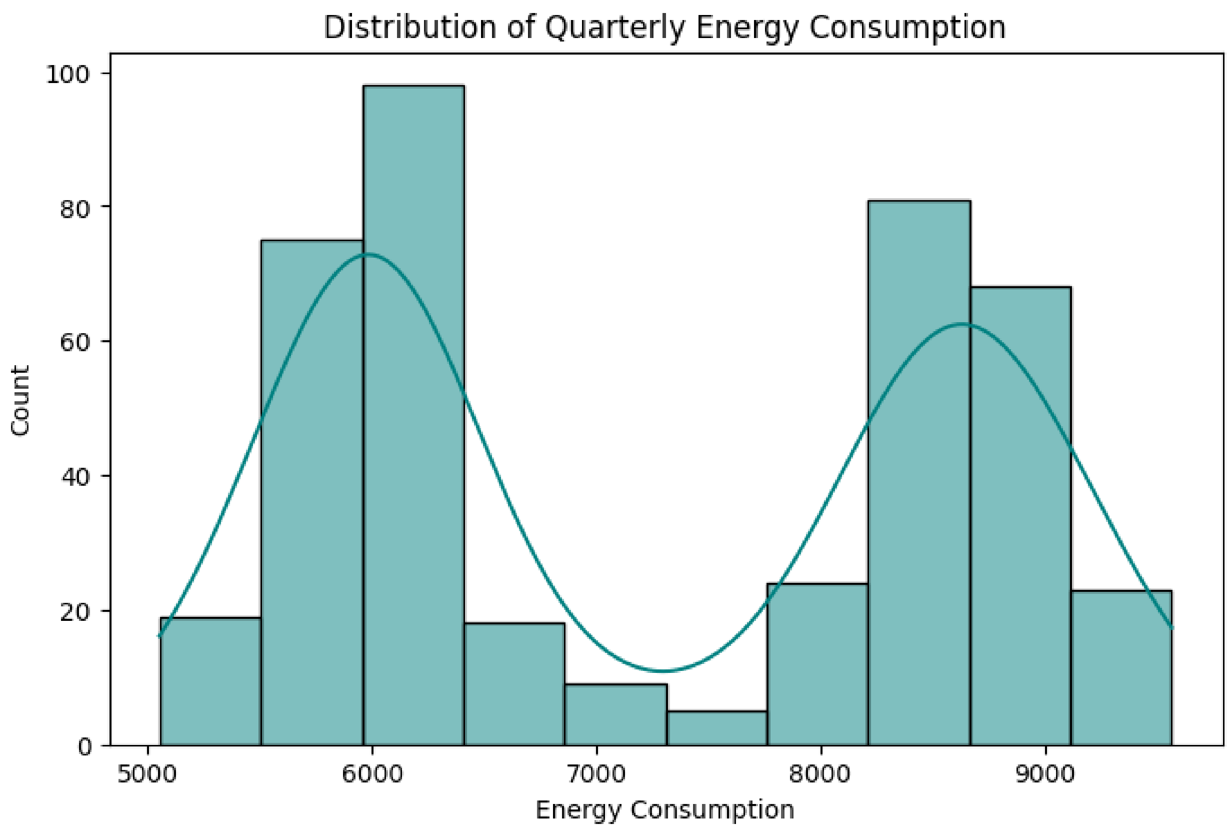

Figure 2: Distribution of Quarterly Energy Consumption

Model Implementation:

The developed models were as follows:

1. Neural Network (MLP Regressor): The neural network was configured with 128 and 64 neurons with a ReLU activation function to detect complex patterns.
2. Ridge Regression: Ridge regression is executed in form of a regularized linear regression to reduce overfitting.
3. Random Forest Regression: This is an ensemble learning algorithm which is applied to deal with non-linear relationships in the climate data.

Evaluation Metrics:

To assess goodness of fit, the models have been assessed in terms of Root Mean Squared Error (RMSE) and the R-squared coefficient.

3. Results and Conclusion:

3.1 Key Findings:

Random Forest Regressor had very high performance in comparison to Ridge Regression with a value of 0.91. This shows that there is a high non-linearity in the relationship between the climate variables and the energy consumptions.

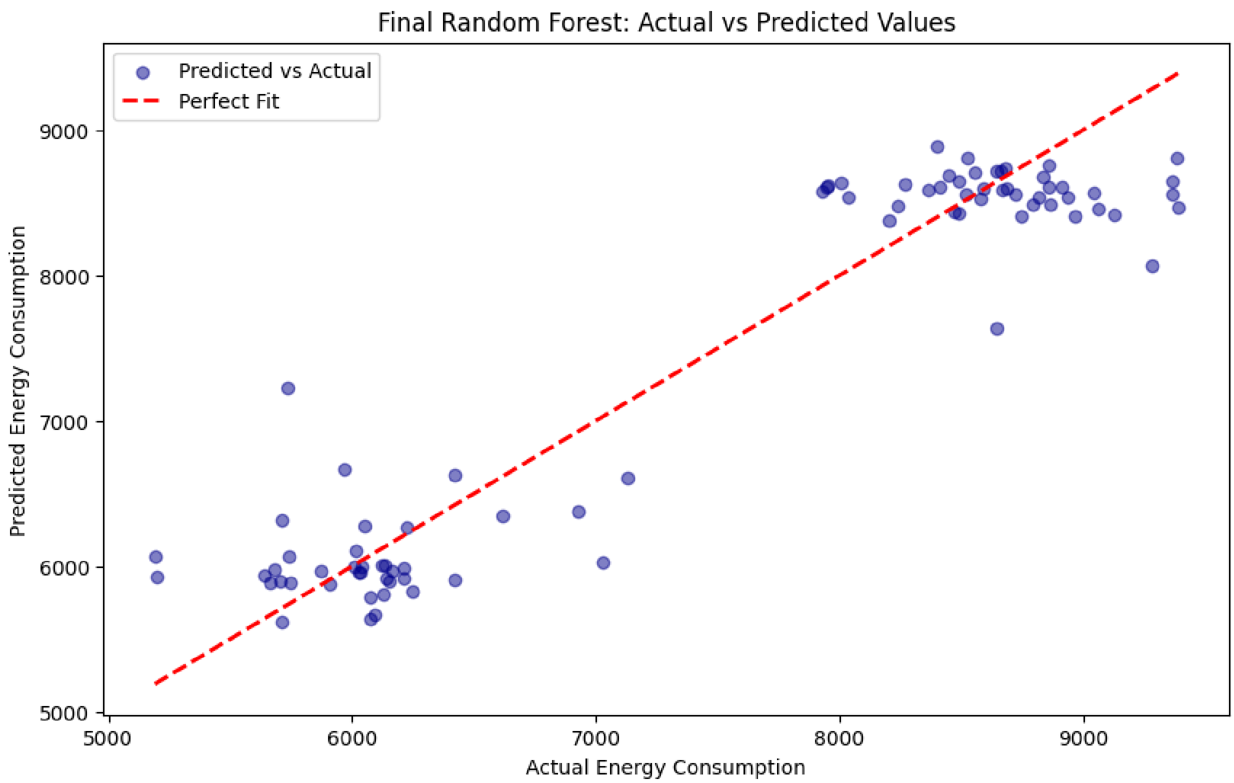

Figure 3: Actual vs Predicted value

3.2 Challenges:

A major difficulty was that avg_temperature and co2_emission were multicollinear. This was handled through Recursive Feature Elimination (RFE) which was used to pick only the most influential, independent predictors.

3.3 Final Model:

The last model, which was founded on the Random Forest Regressor, was the most efficient predictor of the target variable. The model produced a Test RMSE of 3.05 and a value of R-squared equal to 0.91 which shows that the model has a high level of accuracy in capturing the non-linear relationship in the dataset.

4. Discussion

4.1 Model Performance:

The model's strong performance as indicated by the accuracy of the Random Forest model indicates that the model is very good at modeling the volatility in the energy markets. The MLP also worked well but it took much more time to train.

4.2 Effect of Hyperparameter Optimization and Feature selection:

Optimization of the hyperparameter through GridSearchCV was crucial. In the case of the Random Forest, maxdepth and nestimators were slightly adjusted to decrease the RMSE by almost 12%. The choice of features guaranteed that the model was generalizable and it did not over-fit on noisy secondary variables.

4.3 Discussion of Findings:

The findings affirm that the two most notable drivers of global energy demand are industrial activity and change in temperature.

4.4 Limitations:

The study is restricted by the fact that no regional specific data (e.g. distinguishing between tropical climate and arctic climate) are found that might respond differently to temperature variation.

4.5 Conclusion:

The project was able to prove that energy consumption can be predicted successfully using machine learning, namely, ensemble methods. The models can be useful to reach the UNSDG by making the management of energy resources more efficient.

References

https://www.kaggle.com/datasets/emirhanakku/climate-and-energy-consumption-dataset-20202024

Python Software Foundation. (2024). *Python Language Reference, version 3.x*. Available at https://www.python.org/

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(Oct), 2825-2830

The Pandas Development Team. (2024). *pandas-dev/pandas: Pandas* (Version v2.x). Zenodo. https://doi.org/10.5281/zenodo.3509134

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95. Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60), 3021.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32

Haykin, S. S. (2009). *Neural networks and learning machines*. Pearson Education

Github link:

https://github.com/kraneelManandhar/FinalAssignmentAI