Model Comparison and Evaluation

Linear Regression, Random Forest and Support Vector Regressor

Chaney Chetty
September 2024

Introduction

Climate change is a major global challenge, impacting ecosystems, economies, and communities. The agri-food sector significantly contributes to CO2 emissions.

This project, led by environmental consultants and data scientists, analyzes and predicts the sector’ s CO2 impact on climate change. Using data from FAO and IPCC, we performed regression analysis to understand and predict how CO2 emissions from the agri-food sector affect climate change. By analyzing the data, we identified patterns and relationships between emissions and temperature variations. This helped us provide actionable insights for reducing the sector’ s carbon footprint and promoting sustainable practices.

Objective

To compare the performance of three regression models: Linear Regression, Random Forest, and Support Vector Regressor based on a dataset that contains CO2 emissions data from various agricultural activities, demographic information, and average temperature increases for different regions spanning 30 years, from 1990 to 2020.

Model Overview

Linear Regression Model: Simple, interpretable model

Random Forest: Ensemble method, good for capturing non-linear relationships

Support Vector Regressor: Effective in high-dimensional spaces, but can be sensitive to outliers

Evaluation Metrics

- **Mean Squared Error (MSE):** Measures the average squared difference between actual and predicted values.
- **R-squared (R2):** Indicates the proportion of variance in the dependent variable that is predictable from the independent variables.

Model Performance

Linear Regression Model:

- MSE: 0.00
- R2: 1.00
- Performance: Excellent, with predictions almost perfectly aligned with actual values.

Random Forest:

- MSE: 0.01
- R2: 0.99
- Performance: Very good, with predictions closely following the actual values, though slightly less perfect than Linear Regression.

Support Vector Regressor:

- MSE: 1.30
- R2: 0.02
- Performance: Poor, with significant deviations from actual values, especially for higher values.

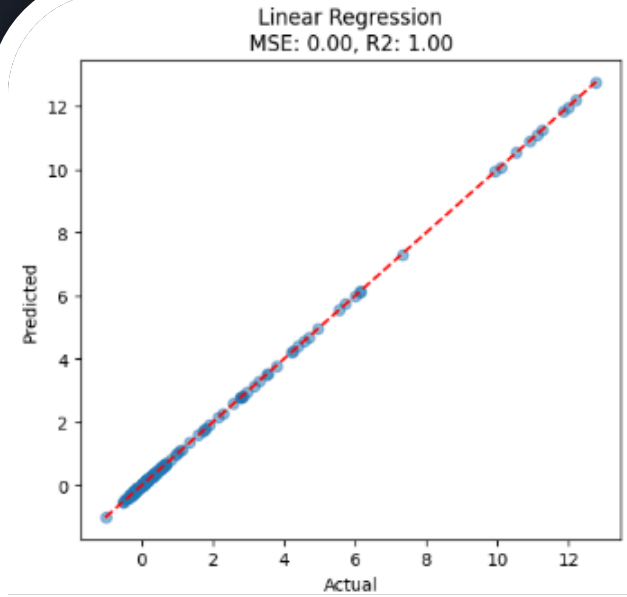
Model Comparison

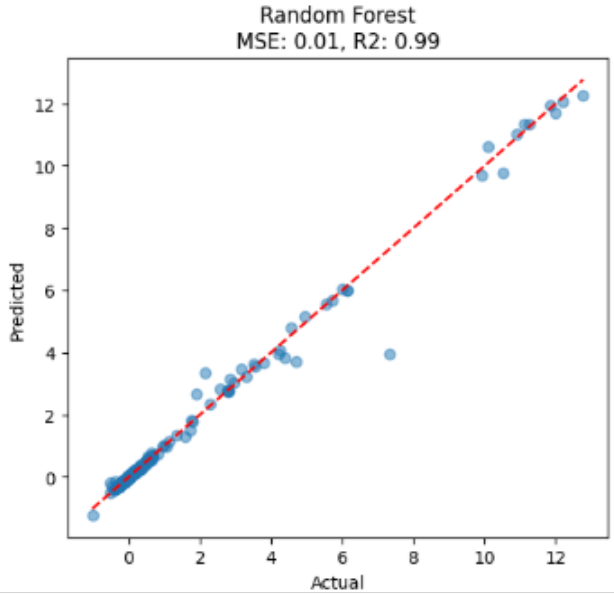
Best Model:
Linear Regression, with the lowest MSE and highest R2, indicating near-perfect predictions.

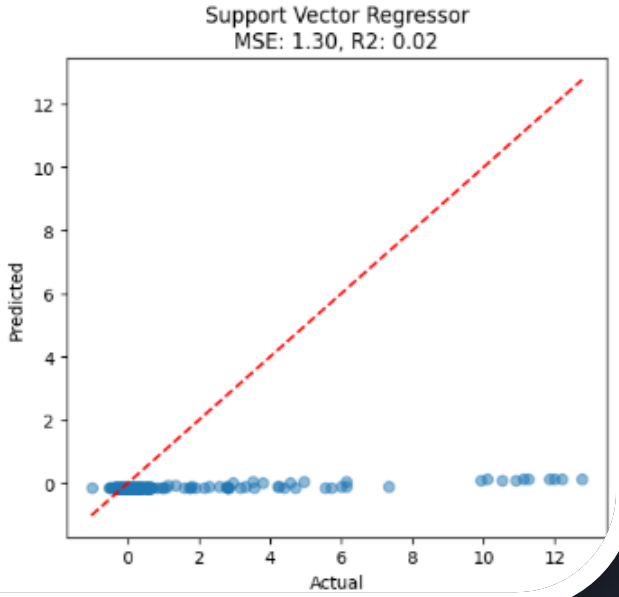
Second Best:
Random Forest, with slightly higher MSE and lower R2 than Linear Regression but still very accurate.

Least Effective:
Support Vector Regressor, with much higher MSE and lower R2, indicating poor predictive performance.

Visual Comparisons







Conclusion

Conclusion

- The analysis revealed that the Linear Regression model performed exceptionally well, achieving a Mean Squared Error (MSE) of 0.00 and an R-squared (R2) score of 1.00, indicating near-perfect predictions.
- The Random Forest model also demonstrated strong performance with an MSE of 0.01 and an R2 of 0.99.
- In contrast, the Support Vector Regressor showed significantly poorer performance, with an MSE of 1.30 and an R2 of 0.02.
- These findings suggest that Linear Regression and Random Forest are highly effective for predicting the target variable in this dataset.