Mahtab Nahayati

Report on Modeling Approach and Visualization for Trust Enhancement

1. Modeling Approach

In this analysis, the objective was to predict the Happiness Score using key predictors such as GDP per Capita, Social Support, CO2 Production, and HDI. The dataset was preprocessed to address missing values and scaled for consistent feature ranges. Two models were implemented and evaluated:

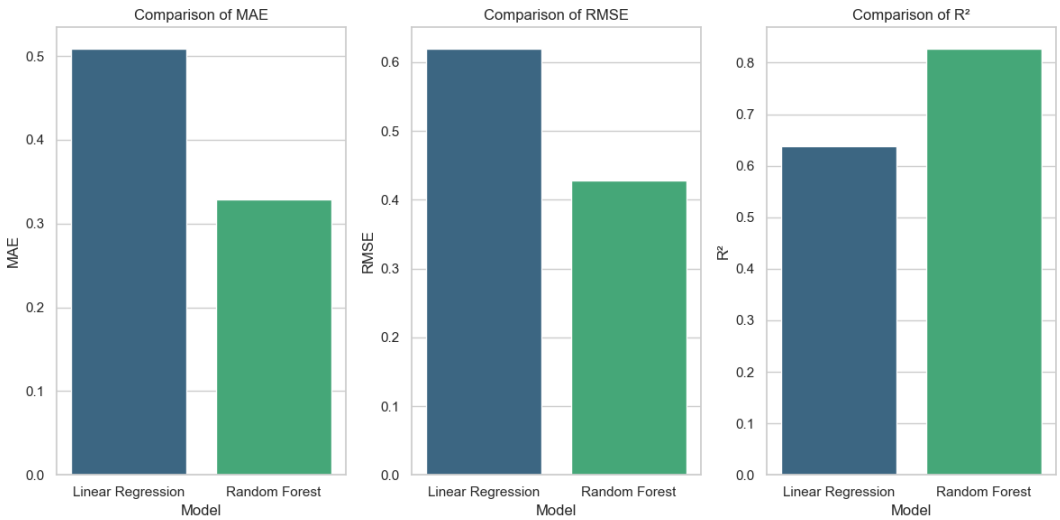
- **Linear Regression:** A baseline regression model to establish initial performance metrics. This model is interpretable, providing insight into feature importance and directionality.
- **Random Forest Regressor:** A non-linear, ensemble-based model selected for its ability to handle feature interactions and non-linear relationships.

Both models were trained and tested using an 80-20 train-test split. Evaluation metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2, were computed to assess model performance. Random Forest outperformed Linear Regression with lower MAE (0.33 vs. 0.51) and RMSE (0.43 vs. 0.62), and a higher R^2 (0.83 vs. 0.64), indicating superior predictive capability.

2. Enhancing Trust through Visualization

Data visualization was integral in fostering trust and understanding of the modeling process and results. The following strategies were employed:

- **Model Performance Comparison:** Bar charts compared MAE, RMSE, and R^2 across the two models. These plots demonstrated the clear performance advantage of the Random Forest model, reinforcing confidence in its selection.



- **Feature Importance:** A feature importance plot for the Random Forest model was created. This visualization clarified which predictors significantly influenced Happiness Score, ensuring transparency.