Technical Report

Introduction

This project set out to predict the **Index of Economic Freedom "Overall Score"** using the underlying indicators that make up the index. A central challenge was that *Monetary Freedom* is included both as an input indicator and as part of the overall score. Using it as a predictor for a target that already contains it would introduce circularity. To address this, a new outcome variable was defined: the **"Overall Score excluding Monetary Freedom."**

Data and Preparation

The analysis relied on the Heritage Foundation's Index of Economic Freedom, covering the period **1995–2024**.

Data preparation involved several key steps:

- Filling missing values with each country's median score, then dropping any remaining gaps.
- Standardizing column and text formats.
- Creating the adjusted dependent variable.
- Splitting the dataset temporally: all data before 2020 was used for training, while data from 2020 onward was reserved for testing.

Exploratory analysis showed that many indicators are strongly correlated with one another (multicollinearity), and that **Monetary Freedom has a moderate positive correlation (0.55)** with the adjusted overall score. This finding directly supports the hypothesis that countries with weaker Monetary Freedom tend to also have lower scores in other areas of economic freedom.

Modeling Approach

Several regression approaches were tested:

- Linear Regression

- Ridge, Lasso, RidgeCV
- Decision Tree, Random Forest, Gradient Boosting, XGBoost
- **Polynomial Regression (degree 2)**

The modeling pipeline was simplified after encountering compatibility issues with more complex preprocessing steps during cross-validation. The final approach involved applying imputation before splitting, scaling the features, and fitting the models.

Results

Model performance was assessed using **R^2, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE)** on the independent test set.

The **Polynomial Regression model** performed best:

- **$R^2 = 0.9370$**
- **MSE = 8.8914**
- **RMSE = 2.9818**
- **MAE = 2.0175**

This demonstrates that accounting for non-linear relationships significantly improves predictive accuracy compared to linear and regularized models.

Residual analysis showed that the model's errors were generally scattered randomly but tended to increase at the extremes. Learning curves pointed to **possible overfitting**, a known trade-off with polynomial models. Performance was stronger for countries with higher actual freedom scores, suggesting some bias across the score distribution.

Key Decisions

- Excluding *Monetary Freedom* from the dependent variable to avoid circularity.
- Using country-wise median imputation to handle missing values.
- Applying a temporal split to better reflect real-world prediction.
- Removing preprocessing steps (e.g., PowerTransformer, SelectKBest) that caused issues with time-based cross-validation.
- Selecting **Polynomial Regression** as the final model due to its superior test performance.

Ethical Considerations

The analysis identified potential bias: predictions are more reliable for higher-scoring countries. This should be acknowledged in any application of the model, as over-reliance on predictions for lower-scoring countries may misrepresent their economic conditions.

Conclusion

The study shows that **Polynomial Regression is the most effective model** for predicting the “Overall Score excluding Monetary Freedom.” The analysis confirms that *Monetary Freedom* is positively related to other aspects of economic freedom, even when excluded from the target variable. At the same time, the findings highlight the need for caution: the model is complex, prone to overfitting, and less accurate at the lower end of the scale.

Future research could explore stronger regularization techniques, different model families, or direct regressions between Monetary Freedom and the other dimensions of economic freedom to deepen understanding.