Gender Inequality

Chaitanya Shekar, Sangeetha Ramesh, Weiye Zhu, Yaxin Ye

Introduction

Research Questions:

How do we understand various aspects of gender inequality worldwide, and what factors influence these disparities and their changes over time?

A Deeper Dive:

Examining these factors helps identify causes, track changes, and inform policies for gender equality.

Who are affected?

- ***Women:*** Inequality limits women's access to opportunities, resources, and decision-making power.
- ***Men:*** Social expectations and norms limit men's choices.

History:

Persistent gender inequality impacts women in education, employment, and politics. Although 20th-century progress improved women's rights, ongoing efforts are still needed to address remaining disparities.

Why it matters:

- ***Economic Growth:*** Reducing gender inequality boosts productivity, innovation, and economic growth by workforce inclusivity.
- ***Social Equality:*** Promoting gender equality supports social justice, human rights, and individual potential realization.

Analysis & Methods

Analysis:

This study applied diverse methods to analyze global gender inequality, informing policies and actions for promoting equity. It compares gender disparities worldwide, considering social, economic, and political factors, and predicts Gender Inequality Index (GII) values.

Various regression models, assessed by *Root Mean Square Error (RMSE)*, help examine relationships between disparities and their characteristics.

Methodologies:

Linear Regression Analysis: *Linear regression* predicts one variable's value based on another variable. In this case, the Gender Inequality Index value serves as the predictor variable.

Ridge Regression Analysis: *Ridge regression*, a tuning method for analyzing data with multicollinearity, was employed to address high correlation in our data.

LASSO Regression: To obtain more accurate results and to shrink the model, *lasso regression* was used.

Random Forest: To improve accuracy and control overfitting, the *Random Forest* method was used.

To compare the performance of these methods, we used *RMSE* as a comparison metric.

Results

Gender inequality remains a persistent global issue, with the highest levels in southern Africa, South Asia, and South America (*Figure 1*). These regions face challenges include limited education, healthcare access, and workforce opportunities for women.

The correlation plot (*Figure 2*) shows that reducing maternal mortality and adolescent birth rates can lower gender inequality, while secondary education and labor participation have negative correlations with the GII.

All four models were trained to predict GII values, with the *Random Forest* model performing best based on its lowest *RMSE* (*Table 1*). The *Random Forest* model's *mean squared residuals* (MSR) is 0.00543, indicating the model's predictions were close to the actual values. Additionally, the model has a high variance explained of 88.91%, which indicates that the selected variables explained most of the variance in GII. *Figure 3* shows that increasing the number of trees in the model improves its performance.

The variable importance plot (*Figure 4*) highlights maternal mortality ratio as the most crucial factor for predicting GII, followed by female secondary education and adolescent birth rates.

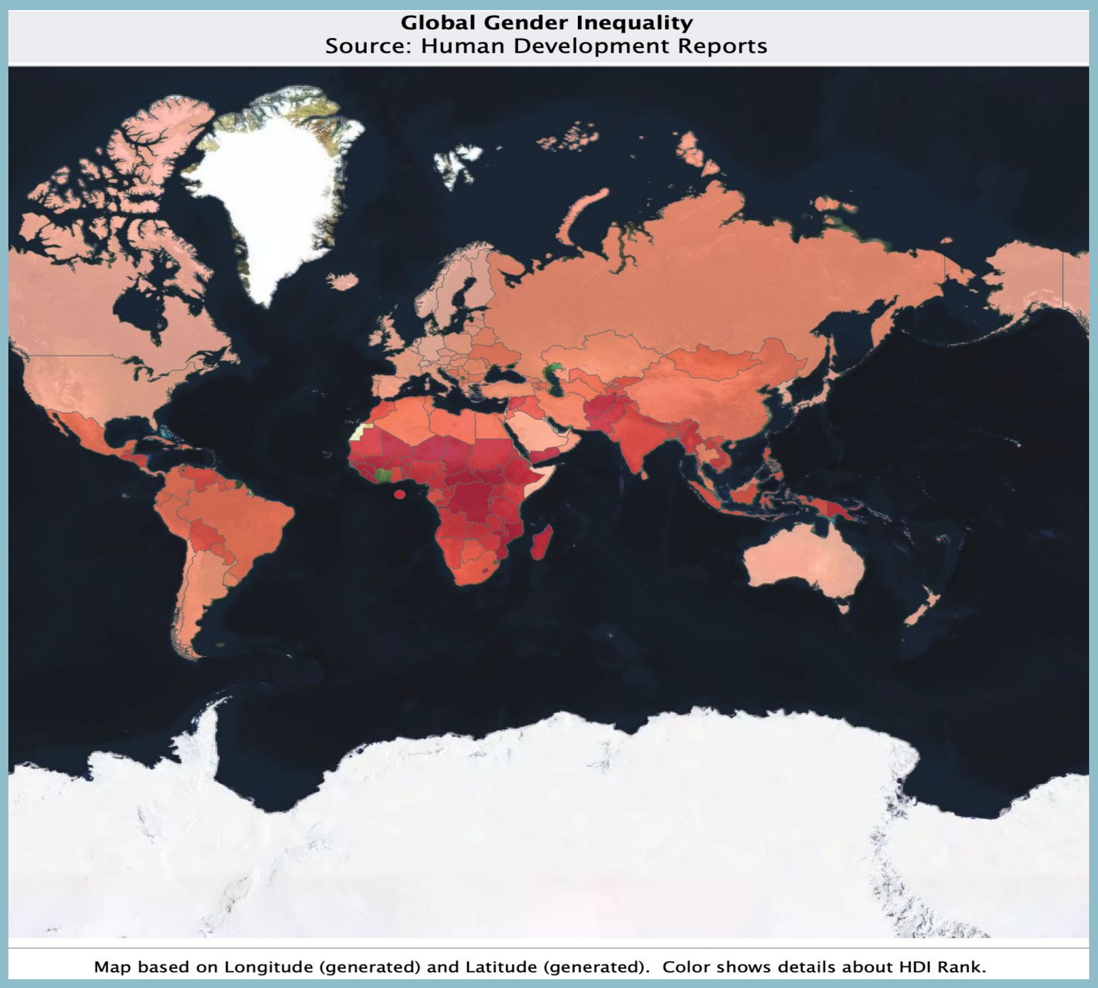

Figure 1: This figure shows global rankings of gender inequality, with some regions in Africa and Asia displaying particularly high levels of inequality according to the Gender Inequality Index.

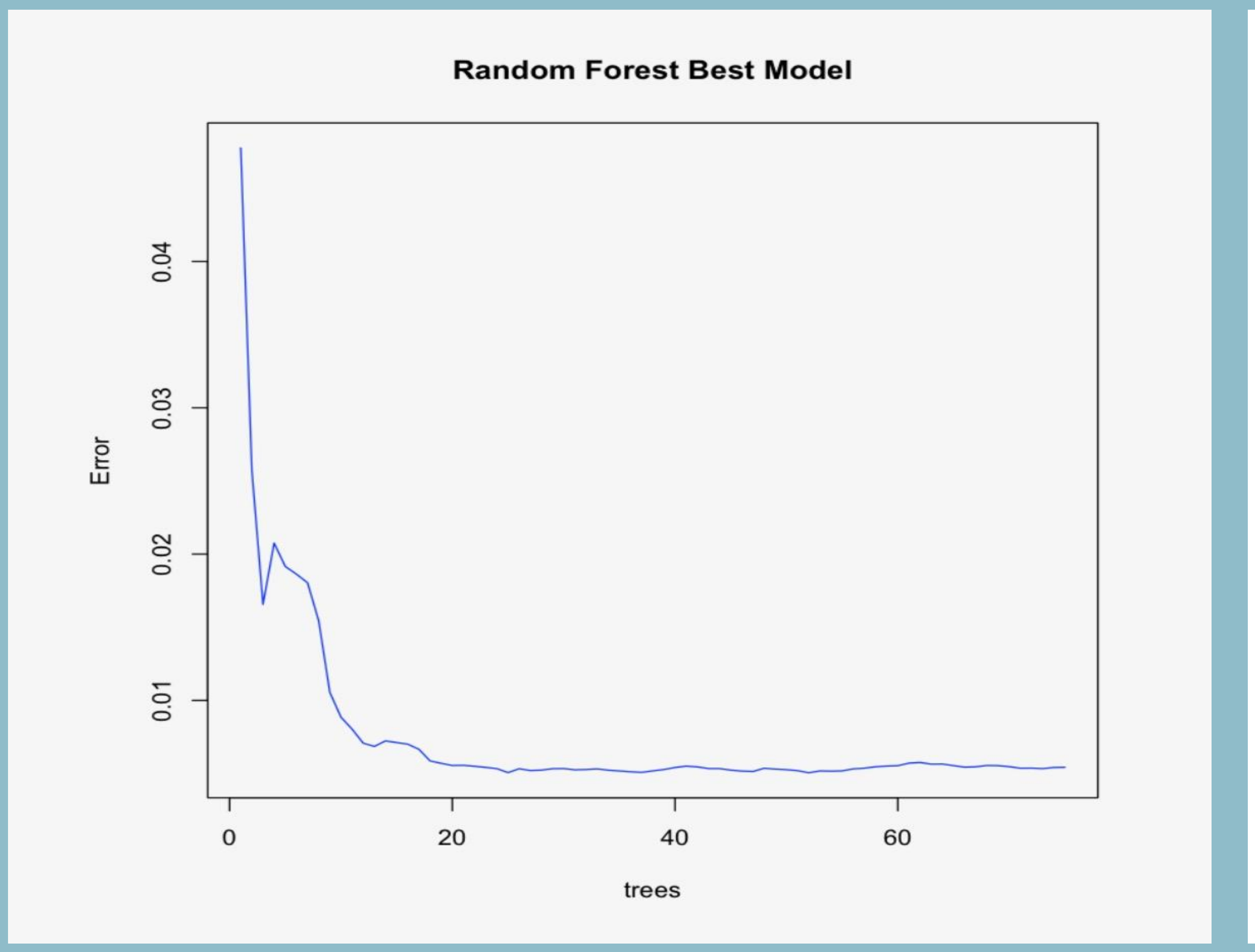

Figure 3: Performance of the best model, random forest with respect to the number of trees in the model.

Machine Learning Models	*RMSE*
Linear Regression	0.7298689
Ridge Regression	0.6027686
Lasso Regression	0.484732
Random Forest	0.08925188

Table 1: RMSE values for different models, providing an overview of their predictive performance.

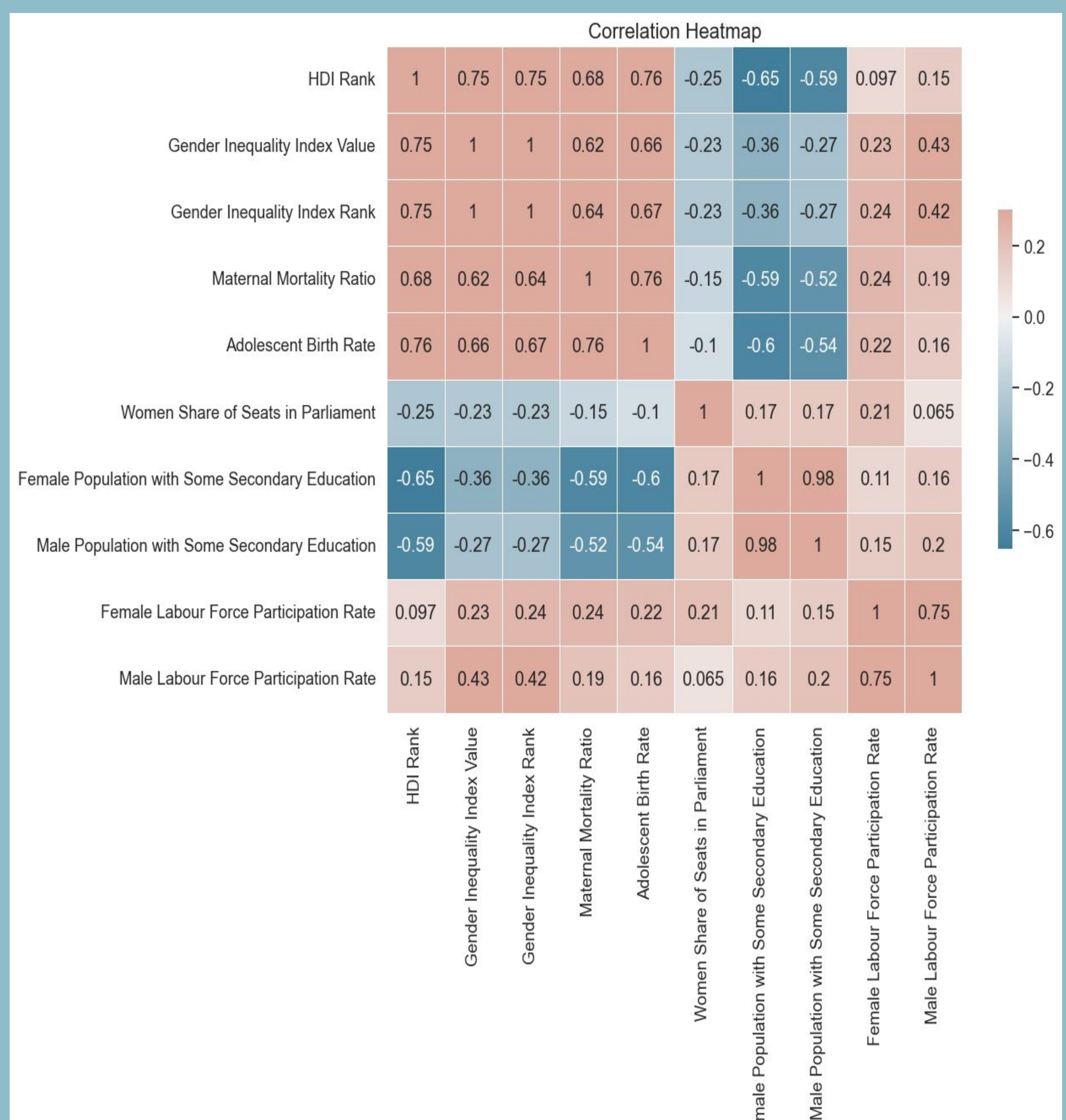

Figure 2: Correlation plot showing the positive relationships between predictor variables and the response variable.

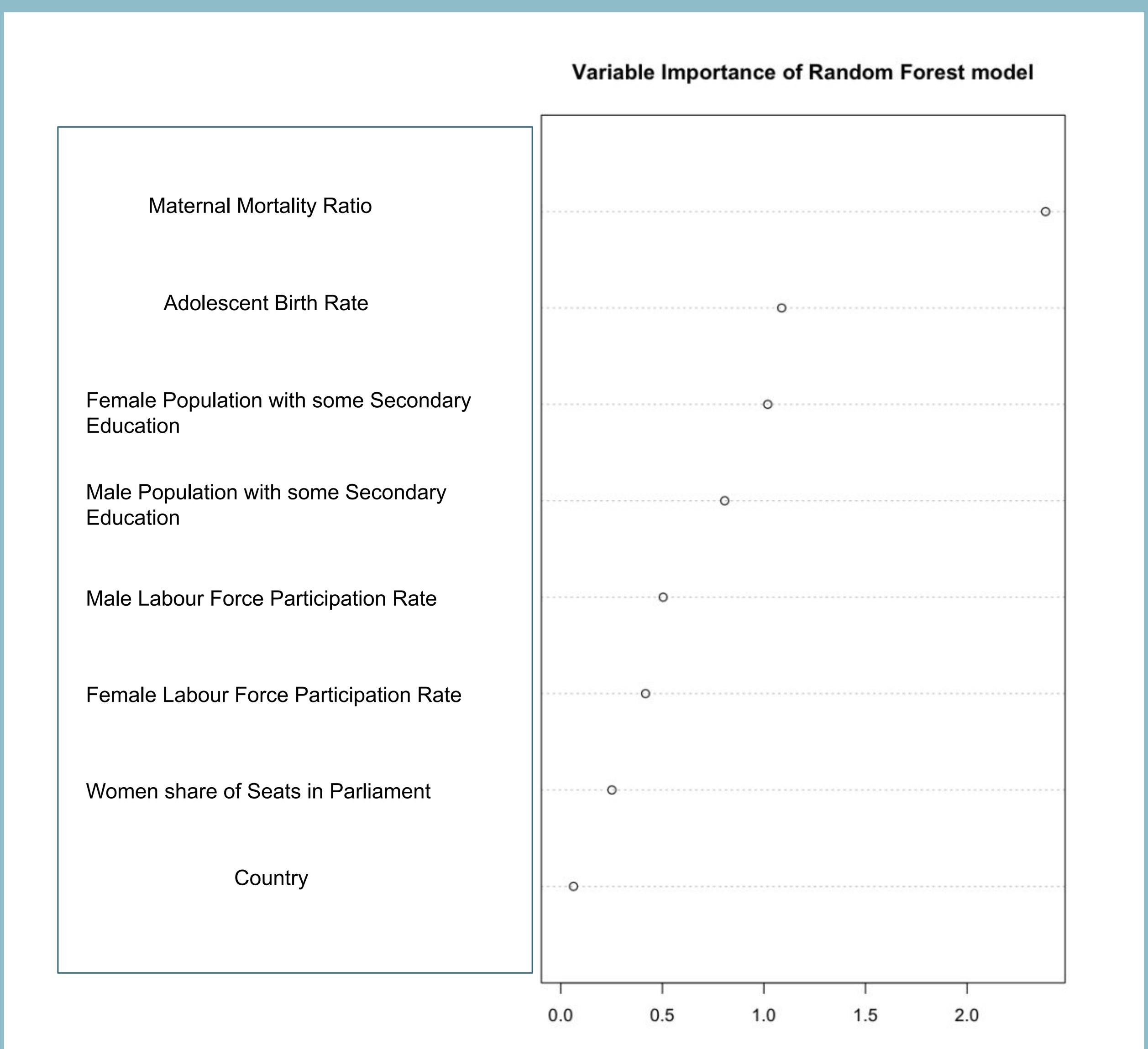

Figure 4: Variance Importance plot for the best model, which is a random forest model. It ranks the variables based on their contribution to the variance in the model, with the most important variables at the top.

Conclusion

Gender inequality significantly impacts regions such as southern Africa, South Asia, and South America, due to discriminatory norms, limited education and healthcare access, and fewer workforce opportunities for women.

Our study found the *Random Forest* model, with the lowest *RMSE* value of 0.08925188, to be the most accurate in predicting gender inequality based on factors like disparities in work and pay. Though *Linear*, *Ridge*, and *Lasso Regression* models may be relevant with less data or variables, the *Random Forest* model provides better predictions for complex data, such as the GII. This insight can guide policymakers in promoting gender equity.