

PROJECT REPORT

Submitted by: Aviral Upadhyay, Sanchit Hari & Kamal Raj Rawat

We have developed a machine learning model that predicts the Gender Development Index (GDI). The GDI is a measure of gender-based inequalities in health, education, and standard of living, and is an important indicator of a country's progress towards gender equality. Once trained, the model can be used to predict the GDI. The model is currently trained to predict the GDI of India but can be used to predict GDI of other countries by using different dataset from the United Nations website: [Human Development Data](#)

Process & Approach:

Out of some of the key development indicators approved by UN namely GDP per capita, HDI, GDI, Multidimensional Poverty Index, Happy Planet Index, etc., We decided to train the model to predict Gender Development Index (GDI). The features (factors) used are Life Expectancy at Birth (yrs), Mean Years of Schooling (yrs) and Gross National Income per capita (PPP\$).

Dataset:

Dataset for the same was taken from the official website of the United Nations. The link to which is : <https://hdr.undp.org/data-center/specific-country-data#/countries/IND>

The above dataset includes data for the various indicators approved by the United Nations from the year 1990-2021 for India.

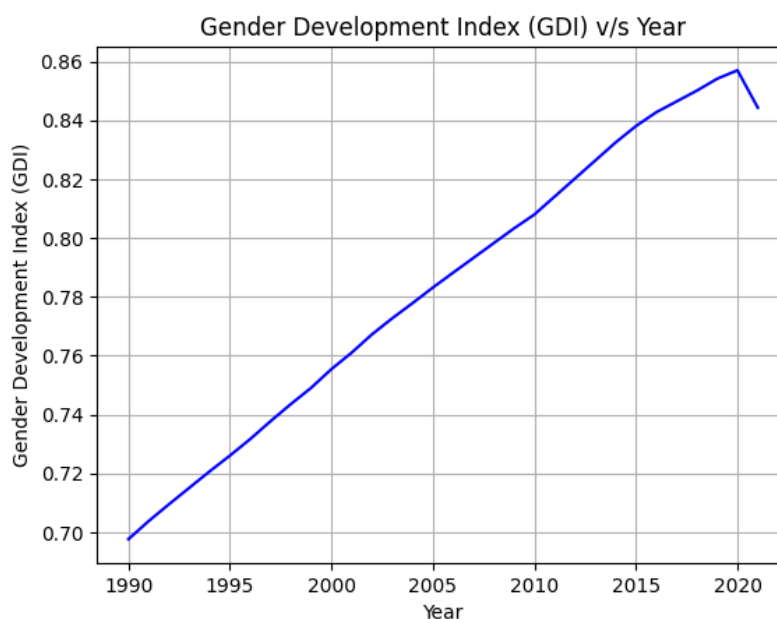
For designing the model, two separate arrays X and Y were created for input features and the dependent variable respectively. Array X was created by using 4 different arrays (X1, X2, X3, X4 for years, Life Expectancy at birth, Mean Years of Schooling and GNI per capita respectively). Array Y was directly taken from the dataset.

The dataset taken was split into two parts with 20% examples (7) for testing and 80% examples (25) for training the model.

Two different Machine learning models were trained. One using the Linear Regression Model for predicting the GDI while the other uses Random Forest Regression Model. The corresponding R2 score (A statistical measure of how well the regression line approximates the actual data) and Mean Squared Error is reported.

The Results through the two different Models are as follows:

Linear Regression Model



Plot of Gender Development Index v/s years

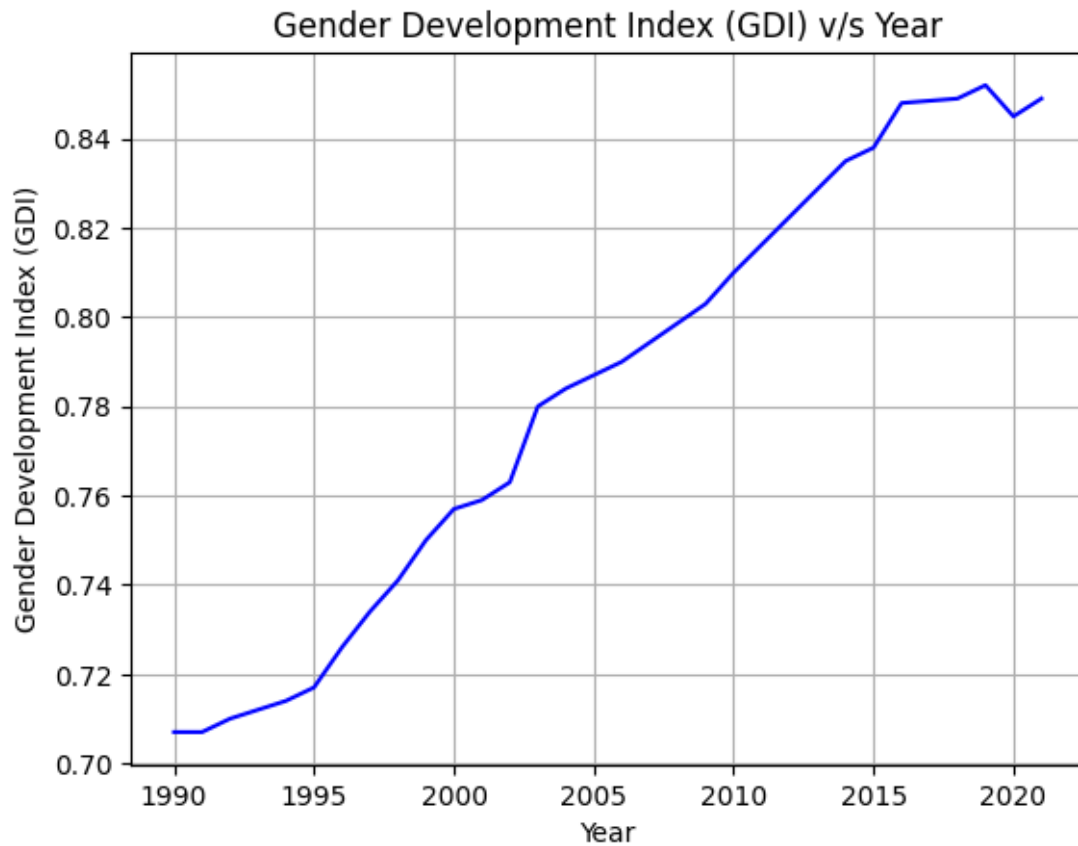
R2 Score: 0.9923081500796138

Mean Squared Error: 1.3534202376241883e-05

Link to the Notebook:

<https://colab.research.google.com/drive/1CVeNOqsFcQweSbVpuq88l-dareWCSNB4?usp=sharing>

Decision Tree Regression Model



Plot of Gender Development Index v/s years

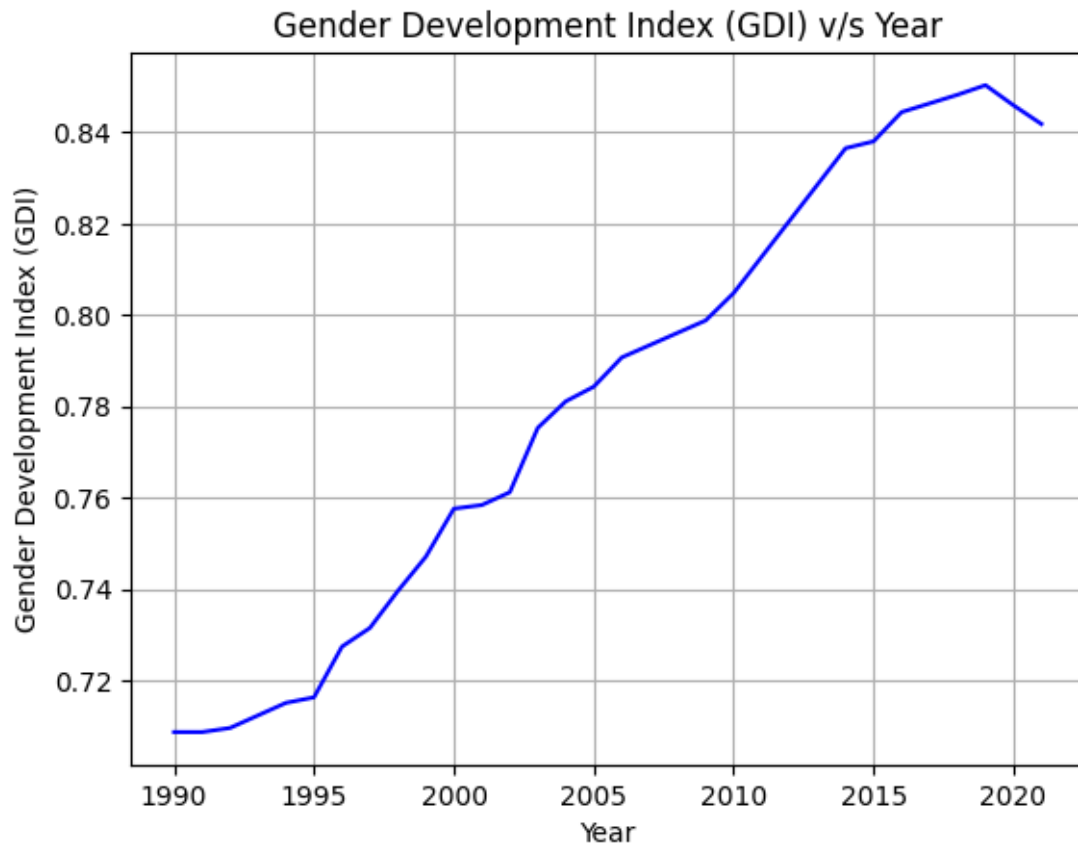
R2 Score: 0.9483634507875385

Mean Squared Error: 9.085714285714303e-05

Link to the Notebook:

https://colab.research.google.com/drive/1MdUYYLX_88huNqnKuEe6VrJlwEsmGx47?usp=sharing

Random Forest Regression Model



Plot of Gender Development Index v/s years

R2 Score: 0.992804170822798

Mean Squared Error: 1.2661428571428469e-05

Link to the Notebook:

<https://colab.research.google.com/drive/12rmloxzTP24BlqIWjc9pwMKpTa3oJBQQ?usp=sharing>

Steps to run the notebooks:

1. Open the links to the Google Colab Notebooks
2. Open the link: [Development Report India](#)
3. Download the Country Data (csv) and upload it to session storage
4. Run the cells (You can use 'Run all' from the 'Runtime' menu)