PUBLIC POLICY AND OPINION CELL
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

Assignment 2

PREDICITNG THE HDI USING LINEAR REGRESSION MODEL

REPORT

Mentored by Hitesh Anand

UNDERSTANDING VARIOUS DEVELOPMENT INDICATORS

BY PUBLIC POLICY AND OPINION CELL
IIT Kanpur

TEAM MEMBERS:
ARYAN KUMAR
ANUSHKA GUPTA
VISHAL HIMMATSINGHKA

PREDICITNG HDI OF INDIA

METHOD USED :LINEAR REGRESSION MODEL

MODEL PREPARED IN : GOOGLE COLLAB NOTEBOOK.

FOLLOWING PYTHON LIBRARIES ARE USED FOR IMPLEMENTING THE CODE

1- numpy array
2-matlplotlib for plotting graph
3-pandas for dataframe
4-sklearn for regression model

PREDICITNG HDI OF INDIA

Features Used in implementing the model

For accomplishing our task we have taken five features in our model which are as follows:-

1- year

2- Life expectancy at birth

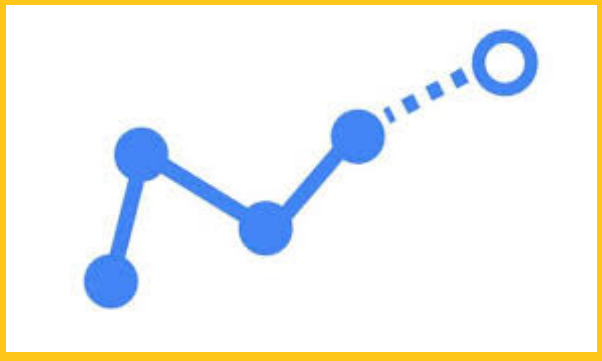
3- expected years of schooling

4- mean years of schooling

5- gross national income per capita(ppp)

NOTE:

The reason behind taking these features is because these are the indcators which is used to calculate HDI of any country by HDRO(human development report office) of UNDP(united nations development programme)



PREDICITNG HDI OF INDIA

Statistical analysis

Following are some statistics related to our regression model:

*For implementing our model we have used world bank data from 1990 to 2020.
*We used 21 data set to train our model and 9 data set to test the model, i.e splitting of data is 3:1.
*We also check our model with two more regression algorithm in addition to linear regression.
*For all the regression model we use R2 score as a metric ,since there is no proper accurate metric for regression because we are not expecting accurate predicted values. *
*Apart from R2 score we also use Mean squared error often known as L2 loss function to estimate the accuracy of our model.
*First one is decision tree regressor in which we got an accuracy of 97.42%
*Second is random forest regressor giving an accuracy of about 98.91%
*After implementing these regressor we found that the linear regression model is much more accurate and efficient than other by giving an accuracy of 99.95% and negligible mean squared error.