

NIRF PREDICTION - 2024

The National Institutional Ranking Framework ([NIRF](#)) was approved by the MHRD and launched by the Honourable Minister of Human Resource Development on 29th September 2015. This framework outlines a methodology to rank institutions across the country. The methodology draws from the overall recommendations and broad understanding arrived at by a Core Committee set up by MHRD, to identify the broad parameters for ranking various universities and institutions. The parameters broadly cover “Teaching, Learning and Resources,” “Research and Professional Practices,” “Graduation Outcomes,” “Outreach and Inclusivity,” and “Perception”.

In this project, we are given the data of the last 5 years (2019-2023), and I have to predict the NIRF rank for the year 2024 based on that data.

To predict the rank, I have used the [scikit-learn](#) library. Scikit-learn is a free software machine-learning library for the Python Programming Language. It features various classification, regression, and clustering algorithms. As per the given data, the institutes are ranked using five different parameters. In addition, there are a total of 16 distinct sub-parameters for these parameters. I applied the **Multiple Linear regression** Model. Multiple linear regression refers to a statistical technique that uses two or more independent variables to predict the outcome of the dependent variable.

- 1) I calculated the Pearson coefficients of the 16 parameters with the rank to find out the highly significant attributes. I proceeded with the parameters with a Pearson coefficient of more than 0.5 or less than -0.5. There were such 6 parameters – FSR, FQE, FRU, FPPP, GUE, and WD. I applied the linear regression model on these parameters and got the rank 51.688.

	Year	Rank
Year	1.000000	0.306468
Rank	0.306468	1.000000
SS (20)	-0.638528	-0.189436
FSR (25)	-0.717267	-0.838188
FQE (20)	-0.052180	-0.963172
FRU (30)	-0.541968	-0.842073
PU (35)	0.907937	-0.087429
QP (35)	0.989312	0.173070
IPR (15)	0.707107	-0.301157
FPPP (15)	-0.682373	-0.586019
GUE (60)	-0.919251	-0.559229
GPHD (40)	0.912778	-0.029024
RD (30)	0.316064	0.062813
WD (30)	-0.130447	-0.893603
ESCS (20)	0.591738	-0.048126
PCS (20)	NaN	NaN
PREMP (100)	0.821128	0.362896

- 2) I also applied the linear regression model on 16 parameters and got the rank of 55.135. For this purpose, I have set the values of the OE for the years 2019 and 2020 as the mean of the years 2021-23 i.e. 3.5.

```
prediction.py > ...
1  import pandas as pd
2  import numpy as np
3  from sklearn.linear_model import LinearRegression
4  data = pd.read_csv("NIRF predictions.csv")
5  data = data
6  colnames = data.columns.values
7  OE_val_mean = np.mean(np.array(data[colnames[6]].values[2:5], dtype=float))
8  data.at[0,colnames[6]] = OE_val_mean
9  data.at[1,colnames[6]] = OE_val_mean
10 #print(data)
11 rank_matrix = np.array(data['Rank'].values[:-1], dtype=float)
12 #print(rank_matrix.shape)
13 data_matrix = np.array(data.values[:-1,2:], dtype=float)
14 #print(data_matrix.shape)
15 reg = LinearRegression()
16 reg.fit(data_matrix, rank_matrix)
17 new_data = np.array([data.values[-1, 2:]], dtype=float)
18 #print(new_data.shape)
19 prediction = reg.predict(new_data)
20 print("Predicted rank:", prediction)

PS C:\Users\91858\Desktop\Project> python -u "c:\Users\91858\Desktop\Project\prediction.py"
Predicted rank: [55.13466229]
```

Thus, I took the average of the two predictions and rounded off the resultant value.

The Final Predicted Rank is 53.

Accuracy: The model is 100% accurate when accuracy of the model is measured with respect to the training data itself because there are a greater number of parameters than number of datapoints. In addition, I have not split the data into the training and testing because the size of dataset is very small.