Webology (ISSN: 1735-188X)
Volume 18, Number 6, 2021

ML Use For Forecasting The NIRF Ranking Of Engineering Colleges In India And PCA To Find The Correct Weightage For The Best Result

Dr. Bishnu Prasad Mishra[1] , Dr. Ratikanta Dash[2] , Dr. Banashri Rath[3]

[1]Director (R&D), GITA Autonomous College, Bhubaneswar, Odisha, India.

[2]Professor --HR, GITA Autonomous College, Bhubaneswar, Odisha, India.

[3]Divisional Head, IDCO, Bhubaneswar, Odisha, India.

Corresponding Author:_
Dr. Ratikanta Dash, professor --HR, GITA Autonomous College, Bhubaneswar, Odisha, India

Abstract

Evaluation provides factual position of any person or organization in the long queue. Annual evaluation for all participating institutions is carried out with respect to particular scale for global as well as local ranking. National Institutional Ranking Framework (NIRF) is adopted in India to rank institute of eminence in various sectors such as University, Engineering, Management, Pharmacy, College, Medical, Law, Architecture and Dental. TRL, RP, GO, OI and Peer Perception are the criteria used to evaluate ranking with 30%, 30%, 20%, 10% and 10% weightage respectively. All above parameters are further sub-divided into 17 parameters with different sub-weightage. Some parameters have strong points for rejecting the application. In this paper, Engineering ranking factual data are considered for developing regression model using ML under individual heading. Multiple regression model has also been developed to cross-check the accuracy of the model. It is felt that the weightage given to each major 5 components needs verification. Machine Learning model using Python software has been developed to train computer to forecast the rank of any participating institution. Training has been imparted to this ML software with 80% random data from NIRF rank list. Subsequently, testing has been done with 20% test data from same NIRF rank list. A few numbers of test data has been fed to the system and accordingly, accurate prediction has been made. Many findings from this ML plots need further detailed interpretation and discussion for refinement of the weightage. Further, the overall combined evaluation has been studied using PCA. The synergy components under different Principal Components have been computed along with their contribution towards the overall evaluation and final ranking. The result encourages the scope for changing the weightage of five main components and modify the weightage for increasing accuracy of the evaluation process. The weightage can be altered and evaluation process can be made more

accurate. ML can be used for forecasting the ranking of any interested institution with the correct input. On-line evaluation of NIRF ranking will be feasible once the parameters are finalized. ML plots show that the scattered plot is not evenly spread. This implies that the funds availability and spending capacity of the institutions are not at par with each other. Many institutions scoring zero in peer perception have obtained ranks within top 200. This needs to be re-examined. The PCA highlights that peer perception weightage criteria is correct and the teaching learning resources weightage is to be reduced for a better judgement. Graduation outcome, Outreach and Industries criteria are not provided with proper weightage. The lower correlation at PC3 and PC2 reveals above facts.

The combination of PCA and ML has given valuable insights for finalizing the NIRF ranking weightage.

Keywords: NIRF, PCA, ML, Weightage, TRL, RP, GO, OI, Peer Perception, Python, Training, Testing

INTRODUCTION:

Evaluation is a continuous process. Annual evaluation by NIRF considering 17 parameter covering academic and non-academic parameters is not judicious. Any Institute has many more activities, which the students like and enjoy along with the study. Presently the industrial ready technical graduates has many more skills to learn along with study. Different parameters like teaching-learning and institute resources, research-professional practice, graduation-outcomes, outreach and inclusivity and peer-perception are the thrust parameters. (Srimathi H. Krishnamoorthy, 2020). While reviewing the activities in an engineering institute/ university campus, generally more than 100 activities are reviewed in every year. These activities also train the students to make them market ready (MHRD, 2020).

The ranking survey conducted by NIRF and publication is highly regarded by all educational institutes. This ranking shows the standard of the institution they maintain with respect to teaching, infrastructure, research, publications, students placement, institute interaction with industries, appreciation by the different elite mass and educationist. The five parameters related to five major parameters are further subdivided in to 17 parameters. All these 17 parameters are measured using different equations and conditions. The data for last three years are considered for evaluation purpose. The weightages provided to all 17 sub parameters and 5 major parameters are empirical ones. Like weightages provided during ANN of the ML program are changed to develop the best predictive model, attempt has been made to change the weightages of the NIRF ranking computation.

Goals:

Considering the above facts, one attempt has been made to find out the justification of weights assigned to each five major parameters and suggest the changes

Find out the point where any institution can qualify for the ranking within 200 position using Machine Learning teaching.

REVIEW OF LITERATURE:

The data available with respect to the ranking are published every year in the NIRF website since2016 based on previous data.

Table 1 OVERALL WEIGHTAGE USED FOR RANKING IN NIRF, INDIA

HEADING	SUB HEADING	WEIGHTAGEFOR RANKING
RP	QP	0.12
RP	PU	0.105
PR	PREMP	0.1
TLR	FSR	0.09
TLR	FRU	0.09
GO	GPHE	0.08
TLR	SS	0.06
TLR	FQE	0.06
GO	MS	0.05
RP	IPR	0.045
GO	GPHD	0.04
GO	GUE	0.03
OI	RD	0.03
OI	WD	0.03
RP	FPPP	0.03
OI	ESCS	0.02
OI	PCS	0.02

The details of the MHRD in ranking reviewed. MHRDnamed as NIRF rank and know their rank over net. Measurement and the institutes fill up procedure followed byengineering college wasconducts ranking survey many institution apply to Awell-definedparameters are circulatedthe format and upload in the NIRF website. All these parameter are listed with individual weight and the cumulative weightage to be computed (Table 1.). Some applications are also rejected under specificsituation. Any institution must work in this direction so that the total score is above the cutoff score to be with in the rank. It has ranking and RP (QP) has highest priority and it has a mean of
.0588. The parameters has value more than .0588 may be considered with priority. They are RP(PU), PR (Premp), TLR (all four parameters) and Go (GPHE).

The contribution of OI is not at higher percentage of mark. These points are to have some minimum

values for qualifying for ranking. It will be a fair process by including some cutoff values. AICTE is advocating for rural link and moral building of the students. Some parametersof such must be included in this evaluation. The Alumni contribution must be considered at higher level. This is a quite fair process of evaluation. Still modifications of the weightage and inclusion of some parameters to bring synergy among the other institutions to be given priority.

ML program can be developed using python (colab, 2020). It is comfortable with any PC or Laptop to work on ML use in day to day problem solving. The GPU, TPU and Python loaded to cloud and available free of cost is prompting many researchers to carry out their research usingColab. Some review papers shows only the algorithms and the categorization of the technique (Dey, 2016) and (Rekha Nagar, 2019) for use on case to case basic. In this research regression model and unsupervised technique has been adopted.

Google Colab has been used in this data analysis. Simple regression has steps like Importing directory, importing dataset, splitting dataset into the training and testing set, predicting the testset result, visualizing the results. The data from the MHRD website was obtained and the data sheet was prepared. The five parameters with weightage was recorded for analysis. Pandas, Numpy and Matplotlib libraries were used in this program. The library like Sklearn was of muchuse for development of program. Python program has been developed and used to find the results required for analysis and interpretation. The data sanitization, normalization, has been performed before conducting the analysis and plotting (Vasiley, 2019), (Zeliko Iyezic, 2014).

Principal Component Analysis (PCA)

A dimension reduction tool has been selected for computing the synergy and contribution of all5 major parameters in the ranking process. Finding the contribution of all five components towards all principal components. The Eigen vector computation and the plots like scree plot,

score plot, loading plot will be used to explore the contribution of individual parameters to principal components (Jolliffe, 2010). The statistical tool has other plots like outlier plot and Bi-plot; but these plots are not been considered in this analysis as the objective was to find thecorrectness of the weight used in the NIRF ranking process. The clustering of data are used to infer conclusions on change of weightage value (Vidal, 2016), (Jolliffe I. T., 2016). The uncertainty and fluctuation in data are examined and components contributing and non-contributing are separated for best result.

MATERIALS AND METHOD

Statistical software (Minita-18), Python programing with Colab has been used to analysis the data. The data (Secondary) are collected from the website of the MHRD, Government of India. The last year ranking data has been entered and .csv file has been created for processing usingegression models. Single and multiple regression has been developed for training and testing data. The same was plotted to see the accuracy of the prediction. The one to one regression model has been developed and scatter and egression plot prepared. Viewing the position of each data point on the graph and we can infer many facts about

the behavior of each parameters of the data set. The agglomeration of the dataset prompts the researcher to give the physical meaning to the plots. Any institution must be around that mark to be placed within 200 rank. The ML is very accurately predicting the parameters. More programming to address different dimensions of ranking can be performed with ML and Deep Learning application. The mean, standard deviation, median and mode value at 25%, 50%, 75% data was computer to see the behavior of data at different level. The training was done with 80% data and testing was done with rest 20% data selected randomly. The model predict the ranking for any set of data for any institution precisely.

PCA use in the research:

Tabulated data in excel worksheet was used to perform PCA analysis. The five parameters are analyzed for the correlation analysis. The Eigen matrix was computed. Important plots was be prepared for interpretation. Score plot was used to see how many institution are above the average value. If is less, then we can infer that the evaluation processed is biased one. The scree plot and loading plot will help the researcher to find the correctness of evaluation. If the slope of the line and length of the line is positive and high the parameters are contributing parameters. Other parameters will be evaluated accordingly and findings may be implemented to improve the evaluation and ranking effectiveness.

RESULTS AND DISCUSSIONS:

Results obtained through ML programming and PCA were presented and the conclusive remarks are highlighted. ML used Google Colab notebooks which executes python programs inGoogle's could with GPU and TUP results are analyzed. The data containing individual score oftop 200 engineering colleges on TRL, RP, GO, OI, Peer Perception were tabulated and converted to CSV file for use in ML programming in Google Colab cloud server. The ML model steps are described as written below.

The Simple regression used steps like Importing the libraries, Importing the dataset, Splitting the dataset into the Training set and Test set, Training the Simple Linear Regression model on the Training set, Predicting the Test set results, Visualizing the Training set results and Visualizing the Test set results. Fiver regression analysis were carried out.

A scatter plot (Figure 1., FigureS 2.) for TRL VS overall score of all 200 aspirant institution was plotted. The scatter plot shows that countable institution has excellent infrastructure. The fundflow may be high and comparing those institution with privately managed institution which arefunded by students needs to be debated and corrective measure to be adopted for equity. Dataare with good correlation. The evaluation parameters are to be altered a little for best contribution from RTL to overall score. It clearly shows that about 11 institutions are having best infrastructure. With privatization of education, it is very difficult for the private colleges to compete with IITs and NITs funded by Central Government and state government. Such biasness will not create healthy evaluation, as private institute never get Government funding.

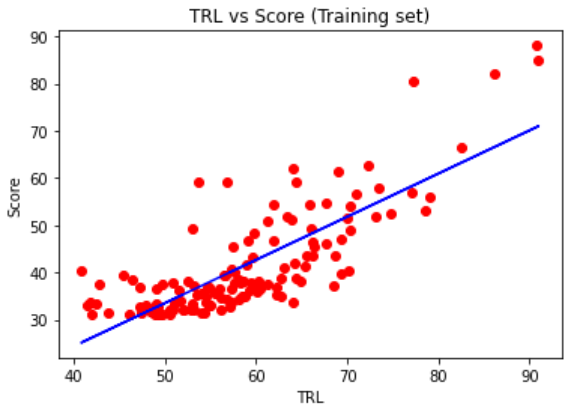

Figure 1 TRL training plot

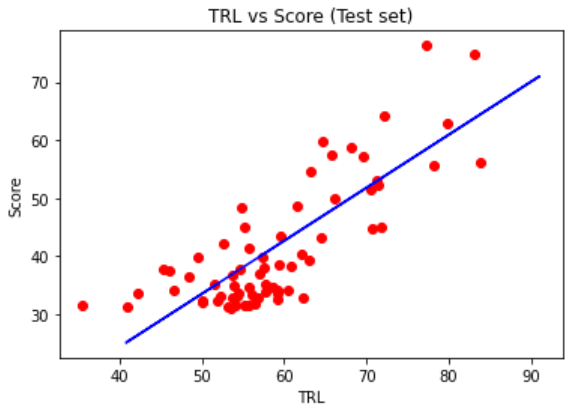

Figure 2 TRL Test Plot

R= 0.914

RP analysis and plot clearly shows, there is high concentration of institution at score 0 to 10. This shows that many institution has poor progress in RP. Scoring 0 and getting a rank within 200 institution in India is a point to be debated. How we can take care of institution so that theyare otherwise evaluated or facility is created so that they come up to that level.

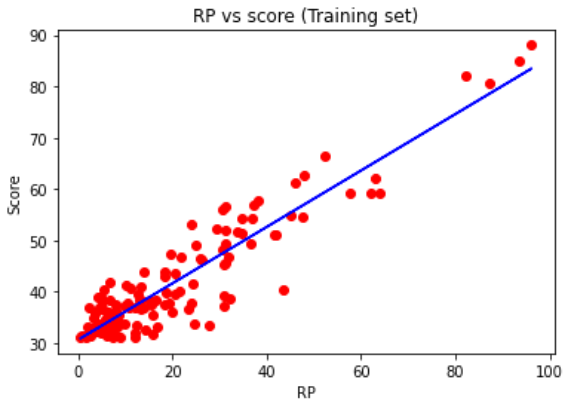

Figure 3 RP Training set

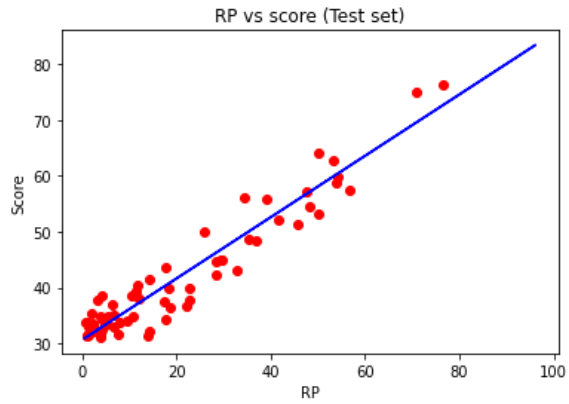

Figure 4 RP Testing set

R = 0.55

Poor regression coefficient shows that some sizable work to be performed in this area. Five institution get very good mark in research front. Can they adopt ten more so that the distributionis best and the research development in India improves? Much work to be done with well spelled out achievable SMART objectives so that all institute do well in research front.

Graduation outcome has been plotted. The regression coefficient is fair (0.77). Only 6 institutions have outstanding mark. But none of the institution has scored below 30 marks. Thedistribution provided

utmost satisfaction as the achievement of most of the institution is more than 50 marks. Such a distribution is expected in all other parameters used for ranking institutions. As all the graduates are placed in some way or other, this is a win-win story.

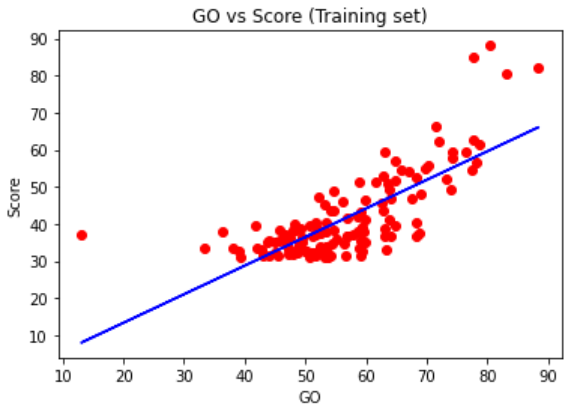

Figure 5 GO Training set

Figure 6 GO Test set

R = 0.77

The regression coefficient is less (0.6). The plot shows wide variation. A few institutes of eminence has better access to the industrial world. Students are debarred from enjoining such facilities. However, most of the institute has score of 50% in this parameter. Industries must bemapped to some institution so that the student get chance to read in practice mode. They morepractical than theoretical. They are more comfortable with the latest machines and gadgets. The induction time of industries will be less and productivity can start flowing quickly once these graduates are employed after graduation.

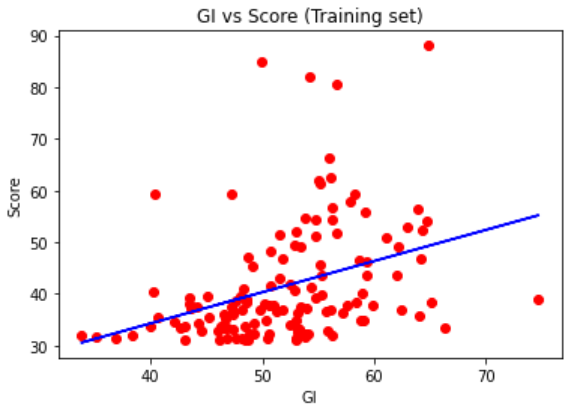

Figure 7 GI Training set

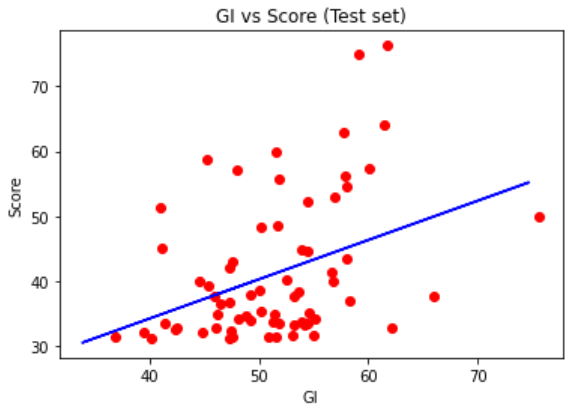

Figure 8 GI Test set

R = 0.60

Considering peer perception, this parameter seems to be higher biasness in the weightage. Regression coefficient is 0.53. It has a higher contribution to ranking in NIRF. Eight institutes arewell known. It is

not possible for people to know engineering colleges. It seems there is a great change required for such a parameter. This parameter needs to be friendly in ranking the institutes. Here many institutes have obtained 0 score and still they are within 200 rank. We have to look into this parameter with greater care.

The evaluation is perfect. The weights in ML (ANN) changes so that the result has least error. Similarly MHRD can rework on the weightage parameters.

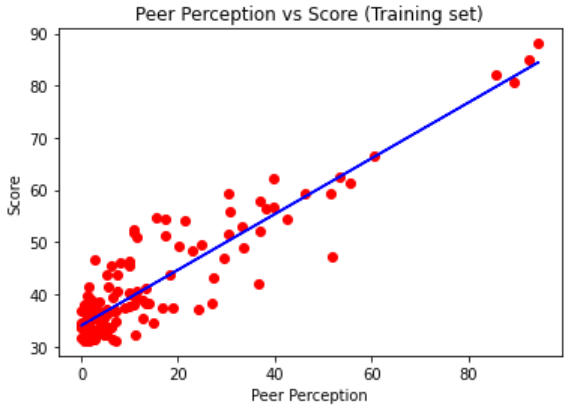

Figure 9 Peer Perception Training set

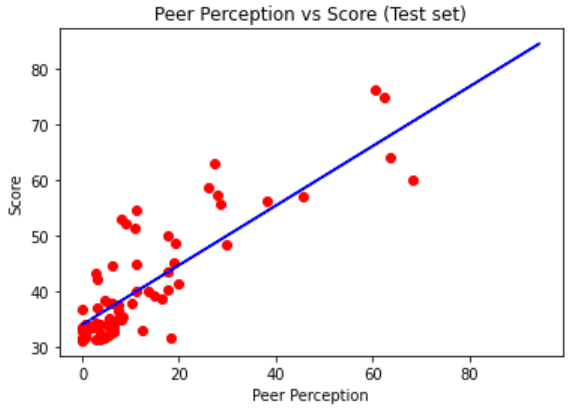

Figure 10 Peer Perception Test set

R = 0.535

The table below measuring the central tendency shows same facts. The standard deviation is highest in research and outreach and inclusivity. The third parameter with third highestst andard deviation is Peer Review. All these three parameters are to be evaluated with appropriate weightage. (Table 2)

Table 2 central tendency evaluation of all 5 parameters

count	199.000000	199.000000	199.000000	199.000000	199.000000	199.000000
mean	58.867387	19.933819	56.566935	51.826281	13.783568	41.516281
std	9.843033	19.272948	10.960868	6.837751	18.318533	11.242451
min	35.510000	0.460000	13.060000	33.800000	0.000000	31.090000
25%	52.520000	5.385000	50.055000	47.350000	2.425000	33.515000
50%	57.490000	13.290000	54.990000	51.850000	6.650000	37.520000
75%	64.380000	30.550000	63.070000	55.890000	17.605000	46.610000
max	91.000000	96.150000	89.650000	75.700000	94.460000	88.080000

```
[0.29978102 0.30002756 0.19995163 0.1001455  0.10003369]
```

The regression coefficient obtained from the Python programming matches with the prevailing weightage. This shows the model developed is quite accurate and the prediction will be best once this ANN model is used for forecasting the NIRF ranking. Putting the score of the institution its ranking can

be obtained on line correctly. It can be a method to check the rankingdata before publication of such result.

PCA

Principal component analysis has been performed to find the synergy within the parameters under evaluation. The parameters like C1 (TRL), C2 (RP), C3 (Go), C4 (OI) and C5 (PP) have beenused in PCA. The Eigen vectors shows that the main contribution for better ranking is from C5, C2, and C1 (with 0.45 as cut off value). PP, RP and TRL has been given more priority as correlation coefficient is higher. In PC1, contribution from C4 and C3 is least. In present contestgraduate outcome and outreach and inclusivity has higher role with respect to make student employable. The PC1 contributes about 62% of the total value. The three major parameters are Peer Perception, Research and Professional Practice and Teaching, Learning and Resources. Only resources and Peer perception cannot be the best parameter to evaluate the rank.

Analyzing the PC2 of the above PCA, it is observed that the outreach and inclusivity is very lessfor most of the top 200 institutions. It has correlation coefficient -0.892 (OI). The contribution from PC2 is 18% but the parameters are less correlated or one is negatively correlated with all other parameters. This give one indication that with less correlation the contribution to rankingis high. Ranking the correlation parameters C2, C5 and C3 has certain correlation. Presently research and professional practice has a greater role in making the institute visible.

PC3 contributes to 10 to the ranking. Main contribution is from teaching learning and resources.This prompts to think putting the TLR with in this weightage than present weightage. Graduation Outcome is negatively correlated at very high value (-0.762). More detailed analysisis required on this parameter with experts to fix weightage of this parameter.

PC4 contributes 5.4% to the ranking. The main contributing factor which has highest correlation is research and Profession practices. Many institute has less resources, hence it has to be incorporated in different dimension. The PC5 contributes 3.5% towards the ranking. The peer perception is highly correlated to the 3.5% contribution whereas it never goes well without research. Many institution has poor performance / reporting with respect to research. Other parameter dominates and they are placed within 200 rank published by NIRF, (MHRD).

Analysis of C1 shows that three parameters are negatively correlated. C2 has one negatively correlated parameter, C3 has three negatively correlated parameters, C4 has two negatively correlated parameter and only C5 has all positively correlated parameters. The change in weightage will improve this component synergy and the ranking process can be fairer.

Details are given in Table below.

Table 4 Eigen Correlation Matrix

Table 3Principal Component Analysis solution

Eigen analysis of the Correlation Matrix

Eigenvalue 3.1129 0.9012 0.5428 0.2687 0.1745

Proportion 0.623 0.180 0.109 0.054 0.035

Cumulative 0.623 0.803 0.911 0.965 1.000

Eigenvectors

Variable	PC1	PC2	PC3	PC4	PC5
C1	0.467	-0.202	0.571	-0.614	-0.195
C2	0.498	0.264	0.175	0.619	-0.519
C3	0.439	0.208	-0.762	-0.384	-0.188
C4	0.274	-0.892	-0.234	0.269	0.050
C5	0.517	0.224	0.087	0.143	0.809

Scree plot (Figure 11.) shows that the contribution from PC1 is 62.3 percent. The contributing sub-parameters are PR, RP, GO and TLR. The contribution from these sub components can be computed and used for computation of weightage for final NIRF ranking. Considering the PC2, RP, GO and PR contributes at very low correlation value. The negative correlation of OI to be considered while finding weightage of this sub parameter.

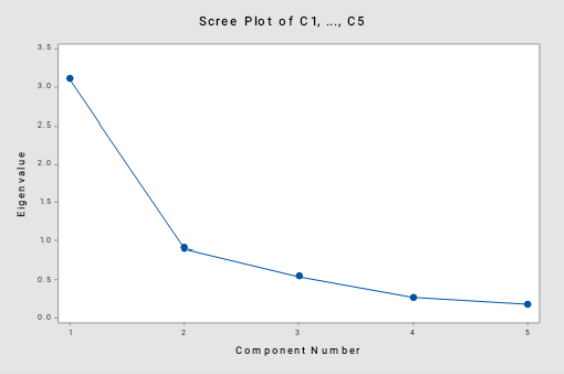

Figure 11 Scree Plot

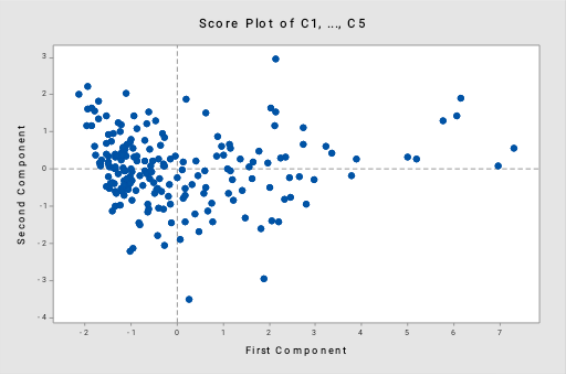

Figure 12 Score Plot

Figure 12. shows that the scatter plot is towards negative value. The positive quadrant is havingmore variation. This shows the weightage given to all parameters are not proper. Properrescaling will help the evaluation process to have higher efficiency. Figure 13. Shows all the five
sub- parameters have higher length. They are contributing parameter towards the evaluation process. The

parameter TLR and OI havenegative slop. The weightage are to be modifiedso that these parameters also contribute positive to the ranking process.

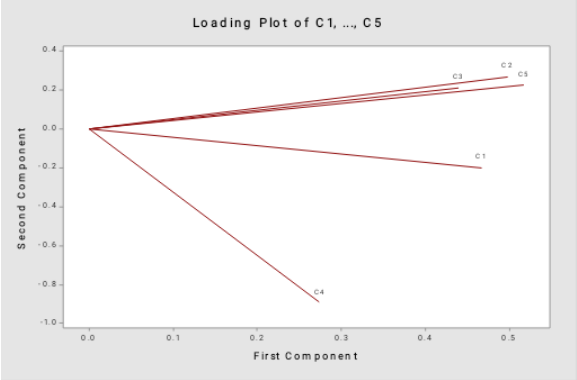

Figure 13 Loading Plot

First and second components are perpendiculario each other. They are not correlated but they are formed independently with synergy for best output. The weightage parameters computed are nor reflected in this paper because of legal issues, however it is suggested to incorporate similar changes in the weightage so that the evaluation process is best one with higher efficiency.

CONCLUSIONS:

There is always scope for refinement and development of better training models using neural network platform. This research is based upon use of machine learning and PCA on presently published NIRF ranking for 2020 by MHRD GOI. The data analysis shows some remarkable recommendations to be considered for better computation for ranking. Many institution have arank with scoring zero score in PR and RP parameters. This may be reviewed further and sub-division of criteria may be developed accordingly. The parameter TRL has been improperlygiven weight. Similarly in the parameter like RP and PR many institution has cored zero. In spiteof scoring zero in RP and PR the institutions have been ranked with in top 200. By changing thesub-criteria for RP and PR along with the weightages, the evaluation process will show proper distribution score. It has been further noted that the parameters like TLR and OI has improper contribution to tanking process. The PC1 and PC2 has less correlation. Data are spread apart inpositive quadrant and has been placed in a small gap in negative quadrant. The minimum valuescored in GO criteria is more than 30. The evaluation parameter with respect to OI seems to be well implemented by all the institutions. Similar easy of adoption to be visible in the parameterslike TLR and GO. Institutes with huge Government funding has scope to score more in these parameter, whereas privately funded institution lag behind. The evaluation processes must be streamlined so that proper evaluation is possible in national level ranking.

References:

1. colab, g. (2020). colab . https://colab.research.google.com/notebooks/intro.ipynb#recent=true.

2. Dey, A. (2016). Machine Learning Algorithms: a review. International Journal ofComputer Science and Information Technology, 1174-1179.

3. Jolliffe, I. T. (2016). Principal component analysis: a review and recent developments. Philosopjical Transnctions A, 374-381.

4. Jolliffe, L. T. (2010). Principal Component Analysis. New York: Springer.

5. MHRD. (2020). NIRF Indian Ranking,

https://nirfcdn.azureedge.net/2019/framework/Engineering.pdf. New Delhi: MHRD.

6. Rekha Nagar, e. (2019). A literature survey on Machine Learning Algorithms. JETIR ,471-474.

7. Srimathi H. Krishnamoorthy, A. (2020). Review on NIRF. Journal of Critical Review,259-262.

8. Vasiley, I. (2019). Python Deep Learning: Exploring deep learning techniques, nural network architecture and GAN with PyTourch, Keras and TensorFlow. Birmingham, UK:Packt Publishing.

9. Vidal, R. (2016). Generalized Principal Component Analysis. New York: Springer.

10. Zeliko Iyezic, e. (2014). Statistics, Data Mining, and Machine Learning in Astronomy: Apractical Python guide for the Analysis of Survey Data. New Jersy, USA: Princeton University Press.