Predictive Analysis of World University Rankings

Yash Jahagirdar
B.E. Computer Engineering,
Vivekanand Education Society's Institute of Technology,
University of Mumbai,
Mumbai-400074, India

ABSTRACT
The paper compares the top ranked world universities with the Indian universities and tries to explore the reasons behind the success of the top ranked universities. It tries to explore the various factors and the influence they have on rankings. The paper tries to prioritize the various factors influencing the world university rankings. This paper aims to suggest some important measures institutions can undertake to improve their rankings on a global level. It tries to predict the world ranking of a particular university based on certain factors which will save cost and thus provide a good cost-benefit analysis.

General Terms
Data analytics, rankings

Keywords
Prediction, Analysis of world rankings, University Rankings

1. INTRODUCTION
During the past decade, there has been lot of discussion regarding the dismal ranking of Indian universities on the world stage. Many researchers, ranking agencies have tried to cite various reasons for this dismal performance. In the past four years it was seen that various Indian universities are addressing these concerns by first evaluating themselves on global scale by giving information to the ranking agencies. In the past years many Indian universities have tried to improve their world rankings by giving lot of priority to various things, yet when it comes to their world counterparts the Indian universities have performed poorly.

2. LITERATURE REVIEW
Various research papers, news articles have been studied as part of the literature review. It has been found that the opinions, suggestions for improvement of rankings of Indian universities have been inadequate and have not been implemented. The National Institutional Ranking Framework produced by the Ministry of Human Resource Development, Government of India only focuses on the comparison of Indian universities based on various factors. These rankings do not compare the Indian universities with the ranked world universities. Thus, they are only limited to Indian universities and do not offer global view of the rankings. Most of the research papers have focused on government policies and public policies. One paper had focused on practices and profiles of top ranked Indian and world universities on the basis of a ranking agency for only a year [1]. There had been no conclusive research about the improvements or decrements in various factors across the years that influence the rankings. Another paper focuses on the methodology applied to ranking of top 50 Indian universities [2].

Thus, in the literature review, it was observed that the given research papers had either compared rankings of world and Indian universities for only a year or focused on methodology related to ranking of Indian universities only. Additionally, none of the research papers surveyed, predict university rankings based on various factors.

In the past few years, many Indian universities have shown the ability of being better than their world counterparts in some areas. This has not been documented in any research papers till date.

There has been no conclusive research about the priority of factors that are influencing world rankings. There has been no attempt in reducing the number of factors required to predict accurately the ranking of a university while saving time costs.

Hence, by doing this literature review, the various deficiencies have been understood and the paper aims to fill these deficiencies by conducting conclusive research. The paper also tries to predict the rankings of universities on the basis of some important factors than taking all factors into consideration.

3. DATABASE, METHODOLOGY AND ANALYSIS

3.1 Database
The dataset for this research was taken from https://www.kaggle.com/mylesoneill/world-university-rankings and from TimesHigherEducation World University Rankings (THE) website. The veracity of the dataset lies with both the websites. The factors or the features depicted in these rankings and used for the research are as follows:-

1. Research - university score for research (the volume, income and reputation of university).
2. Teaching - university score for teaching (the learning environment in university).
3. Citations - university score for citations (research influence).
4. Student to staff ratio (SSR) - Number of students divided by number of staff for a university.
5. Number of international students (No_Intl) – Number of students who are from foreign countries. This does not take into account the foreign or the international staff in the universities.
6. Female to male ratio (FMR) - Female student to Male student ratio.
7. Income - University score for industry income (knowledge transfer).

The other factors in the dataset such as total score and international staff were not taken into account since they were not available for Indian universities. Some of the universities had not given data to the ranking agency regarding these

factors. Hence, these factors were not conclusive enough for further research.

The No_Intl have been found out by multiplication of the total number of full time enrolled students by percentage of international students.

For the years 2014-2015, 2015-2016 and 2016-2017, the dataset was made comprising of all ranked universities. This dataset consisted of 4 Indian universities for the year 2014-2015. The number of Indian universities for the year 2015-2016 were 17 and for the year 2016-2017 there were 31 Indian universities.

3.2 Methodology

After the dataset was made, the top 10 world ranked universities were taken together and their mean was calculated for all the seven factors. This was done for all the three years in the dataset. They were stored separately as mean of top 10 world ranked universities (MofWorld) for every year. This was done as this would make the calculations easier while also setting a standard to which the comparison of Indian universities would be made.

For years 2015-2016 there were 17 Indian universities and for 2016-2017 there were 31 Indian universities respectively in the dataset. The top 10 ranked Indian universities according to THE rankings were taken for years 2015-2016 and 2016-2017 so as to accurately obtain the difference between the top ranked Indian universities and MofWorld.

The mean of these top 10 ranked Indian universities were taken for each factor and for each year viz. 2014-2015, 2015-2016 and 2016-2017.

In the year 2014-2015 even if there were less than ten ranked Indian universities in the THE dataset still there mean was taken for every factor. The mean of top 10 ranked Indian universities were stored for each year as a separate entry for each year as Mean of top 10 ranked Indian universities (MofIndia).

The top ranked Indian university (TopIn) for each year according to THE ranking viz. 2013-2014, 2014-2015, 2015-2016 and 2016-2017 was obtained. For each year 2014-2015, 2015-2016 and 2016-2017, the factors of MofWorld, MofIndia and TopIn were compared. The percentage decrease of MofIndia and TopIn with respect to MofWorld was calculated for each factor for last three years.

A lot of observations and analysis were done based on this percentage difference. The percentage difference between various factors of TopIn and MofIndia were calculated for every year. Based on the observations, lots of insights were obtained. The percentage increase for every factor of TopIn for every year was calculated and the results were analyzed.

E.g.: In year 2014-2015 and 2015-2016, the TopIn was Indian Institute of Science (IISC). There was a percentage increase of 19.49% in the Research factor in these two years. Likewise many other observations have been found out.

Percentage Increase is defined as the difference between the new value and old value multiplied by 100 and divided by the old value. The converse of percentage increase is percentage decrease. The percentage increase of values of various factors of universities for some years have been taken and analysis has been done.

This comparison between the MofWorld, MofIndia and TopIn has been done and shown using data give in Fig.1.

A dataset was obtained from https://www.kaggle.com/mylesoneill/world-university-rankings comprising of top 200 ranked universities for the year 2015-2016 by the ranking agency THE.

From the dataset obtained, the correlation of the various factors, described in Part 3.1., with respect to world ranking was found. The influence each factor had on the university ranking was obtained from this correlation. It also gave the dependency of world ranking on each of these factors. Also, the dependence of various factors on each other was also found out using correlation. From this a large number of conclusions were derived. This proved useful in understanding the relationships between different factors.

3.3 Analysis

After doing the methodology, at various steps certain observations were found. This helped in doing analysis of the rankings.

By analyzing the percentage decrease between MofWorld and TopIn, it was observed that IISC, the TopIn for last 3 years, has been consistently beating the MofWorld in SSR (student to staff ratio) factor. It is remarkable that IISC which is ranked between 201 and 300 in the last three years, is having more staff per student than even the top 10 world ranked universities. (Refer Fig. 2) This is definitely a positive change for Indian universities.

It has also been analyzed that the TopIn for past three years has been consistently improving its research and teaching which is a positive sign for the years to come in terms of ranking. This has also led to improvement in its rankings from the past years such as in the year 2014-2015 rank of IISC was 276-300 in the year 2015-2016 it was 251-300 and in the year 2016-2017 year it was 201-250. Hence, definitely by increasing its research and teaching, the ranking of IISC has definitely improved. (Refer Fig.3)

On the other hand as shown in Fig. 1, Panjab University was the TopIn in year 2013-2014. Its world ranking was 226-250. In the consequent years the world ranking of Panjab University deteriorated. It was world ranked 276-300 in the year 2014-2015, 501-600 in the year 2015-2016 and 601-800 in the year 2016-2017.

This deterioration in Panjab University's ranking can be attributed to decrease in its research and teaching. From the year 2013-2014 to year 2014-2015 there was a percentage decrease of 25% in the research factor of Panjab University. From year 2014-2015 to 2015-2016, there was a percentage decrease of 16.19% in the research factor of this university. Consequently, due to this decrease in research of Panjab University and subsequent increase in the research and teaching factors of IISC, IISC became the TopIn in year 2014-2015.

It was also analyzed that the ranking of universities also depends upon the No_Intl (number of international students) factor. The percentage decrease of No_Intl between the MofWorld and TopIn and MofIndia has been huge. This has led to sliding in ranking of Indian universities.

After observing the dependency of each factor with every other factor and also the dependency of every factor with the university rank, lot of information was analyzed. The following conclusions were derived from the analysis:

- Research depends strongly upon the teaching environment i.e. teaching factor.

- Research does not depend upon Citations.
- Research does not depend upon industry income i.e. Income factor.
- The learning environment i.e. teaching does not depend on the research influence i.e. Citations.
- The Number of international students does not depend on the industry income.

This shows that research output improves when learning environment improves. This also shows that Research output and quality does not depend upon whether a professor is highly cited or not.

Research also is never dependent on the industry income. This shows that even if the industry income is less still research output is not affected by it. The teaching environment does not depend on whether the professors are highly cited or not.

The correlation between the world ranking and each factor was analyzed. This showed the dependency of the rank with each factor. This led to prioritizing of the factors on the basis of their influence on world ranking. The priority is as follows:

1. Research
2. Teaching
3. Citations
4. No_Intl (number of international students)
5. Income
6. SSR
7. FMR

From this correlation it was analyzed that Research and Teaching factor have the highest influence on university ranking.

This supports the earlier finding that IISC improved its ranking from 2014-2015 to 2016-2017 by increasing its research and teaching. This also supports another finding about Panjab University's fall in the ranking due to decrease in research and teaching factors over the few years.

On the basis of this prioritization of factors, it would also be possible to predict the rankings to a particular group of university rankings using less number of factors. This would lead to saving costs and hence time. This would give rise to a cost benefit analysis.

Based on these priority of factors influencing rankings, the top 200 world ranked universities for year 2015-2016 were obtained from THE ranking agency along with the universities' data about various factors.

When the 200 universities are divided into groups of 50 universities each based on the university rankings are made, observations were done and it was found that there was a clear strong dependence on two factors- Research and Teaching for each group of 50 universities.

Hence, for each group of 50 universities there was strong dependence on the two factors Research and Teaching.

This made it possible to predict the university rankings group such as top 50, 51-100, 101-150 or 151-200 using only two factors- Research and Teaching. These groups were made for prediction of rankings using only some factors rather than using all the factors.

This not only saved time of finding group of university ranking but also gave an estimate of where the ranking of a particular university would be. Also, with this prediction it saved time and costs in obtaining a rough estimate of university's ranking.

This prediction model could be extended to Indian universities as well, when sufficient data is made available by the ranking agencies. This will take time as many universities do not submit data to the ranking agencies. This prediction was done using classification and regression tree.

The analysis of various factors and correlations between various factors was done using R programming.

4. CONCLUSION

From the analysis, lot of conclusions have been derived. They are as follows:

- For Indian universities to improve their rankings they have to concentrate on encouraging research.
- The Indian universities have to also improve teaching as both research and teaching are the major factors responsible for improving world rankings.
- For ranking improvements, the Indian universities have to improve the research output by encouraging more students to take up research as a career option.
- Also, improving research facilities rather than increasing the number of highly cited professors in a few universities is a step in the right direction.
- Students and research associates should be paid more money as stipend. This will increase the number of people opting for research career rather than going for corporate jobs.
- By reducing student to staff ratio like IISC, the Indian universities can improve their rankings.
- Hence, there is a need to employ more staff for which the institutions must make appropriate provisions.
- It is a good news that during the past few years the government as well as the institutions are paying special attention to world ranking of Indian universities. But, more has to be done in order to improve Indian universities' rankings.
- The government should take appropriate steps to improve the rankings of Indian universities.
- The future scope of this idea is to perfect this predictive model and thus to make the predictions accurate while saving time and cost of analysis.
- The available data does not have large number of Indian universities and most of the Indian universities have furnished incomplete data. Hence, it is difficult to predict the rankings of universities based on this incomplete data.
- The future scope of this idea is to collect more data in future which will have more Indian universities with their complete data based on various factors. This will provide more insights into the analysis of universities. It will also produce a better predictive model

5. FIGURES

Fig. 1: The various factors of the universities

	Research	Teaching	Citations	SSR	No_Intl	FMR	Income	year
TopIn (Panjab Uni.)	14	25.8	84.7	22.3	167	0.92304	28.4	2013-2014
MofWorld	94.54	88.79	97.05	9.04	4431	0.8421	65.81	2014-2015
MofIndia	22.225	28.925	60.525	15.275	92	0.4596	46.23	2014-2015
TopIn(IISC)	39.5	32.6	51.6	8.3	34	0.2501	37.9	2014-2015
MofWorld	94.13	86.69	98.13	9.51	4620	0.713	65.22	2015-2016
MofIndia	23.7	32.86	38.12	15.47	64	0.335	48.17	2015-2016
TopIn(IISC)	47.2	42.7	42.4	8.3	34	0.2501	52.4	2015-2016
MofWorld	93.85	88.54	98.5	9.74	4697	0.777	59.7	2016-2017
MofIndia	21.7	33.23	43.33	14.89	74	0.269	56.9	2016-2017
TopIn(IISC)	49.2	50.1	47.3	8.3	34	0.2501	48	2016-2017

Fig. 2: Dot plot of Student to staff ratio vs year

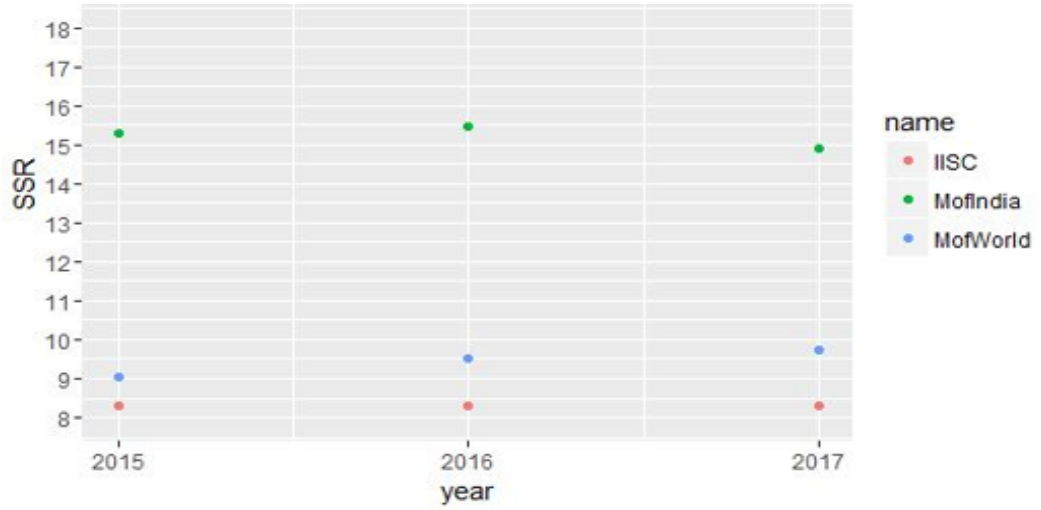
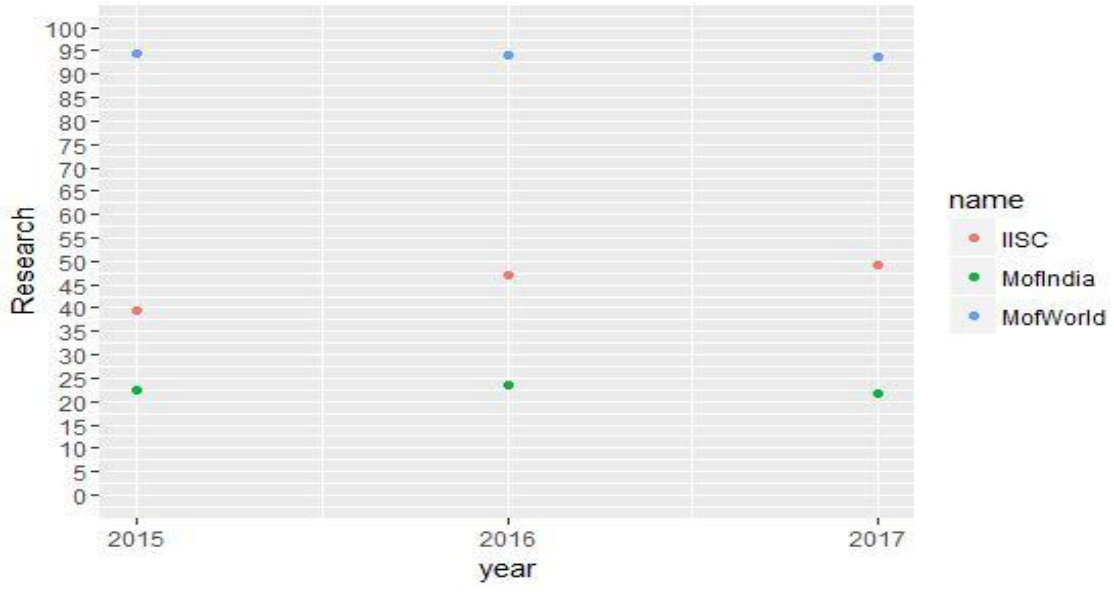


Fig. 3: Dot plot of Research vs year



6. REFERENCES

[1] Shauvik Roy Chowdhury, Koushik Ghosh, "Profile and practices of Indian premier institutes compared to global standards on the basis of QS World Universities Ranking 2013–14" in 2014 2nd International Conference on Business and Information Management (ICBIM).

[2] B.M.Gupta, "Ranking and performance of Indian Universities, based on publication and citation data", Indian Journal of Science and Technology, Vol. 3 No. 7 (July 2010) ISSN: 0974- 6846, pp.837-843.