iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50023

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

College NIRF Rank Predictor using ML

Mr. Omkar Khade[1], Mr. Yash Kadam[2], Mr. Ashish Ruke[3], Mr. Suyash Yeolekar[4]
[1, 2, 3, 4]*Student, Department of Artificial Intelligence and Data Science, VIIT, Pune, Maharashtra, India*

Abstract: ***The National Institutional Ranking Framework (NIRF) is an annual ranking system initiated by the Indian government to rank higher education institutions based on several parameters such as teaching, research, and outreach activities. In this project, we propose to develop a machine learning model that can predict the NIRF rank of an institution. Here we have used 2020 NIRF ranking dataset from Kaggle. Then based on the score of previous years, we predict the rank by giving the performance indicators to the model. The paper focuses on the use of Random Forest Regressor based Machine learning technique to predict NIRF rank. Factors considered are Teaching, Learning and Resources (TLR) score, Research and Professional Practice (RPC) score, Graduation Outcome (GO) score, Outreach and Inclusivity (OI) score and Perception Score for particular college. The model is evaluated using standard strategic indicator: Root Mean Square Error. The low value of this indicator show that the model is efficient in predicting NIRF rank. We got score of 93% and RMSE of 15.47. We have completed ML model save and load operations using Joblib. We have created a flask server for model deployment and deployed on Render as web service. We conducted comprehensive evaluations on frequently used machine learning models and conclude that our proposed solution outperforms due to the comprehensive feature engineering that we built. The system achieves overall high accuracy for College NIRF rank prediction.***

Keywords: NIRF, rank prediction, machine learning, random forest regression, scikit-learn, Flask, higher education institutions, teaching, research, outreach, parameters, web application.

I. INTRODUCTION

A. National Institutional Ranking Framework (NIRF)

National Institutional Ranking Framework (NIRF) is a methodology adopted by the Ministry of Education, Government of India, to rank institutions of higher education in India. The Framework was approved by the MHRD and launched by Minister of Human Resource Development on 29 September 2015. Depending on their areas of operation, institutions have been ranked under 11 different categories – overall, university, colleges, engineering, management, pharmacy, law, medical, architecture, dental and research. The Framework uses several parameters for ranking purposes like resources, research, and stakeholder perception. These parameters have been grouped into five clusters and these clusters were assigned certain weightages. The weightages depend on the type of institution. About 3500 institutions voluntarily participated in the first round of rankings.

The methodology draws from the overall recommendations and broad understanding by a Core Committee set up by MHRD to identify the broad parameters for ranking institutions of Higher Education. The parameters covered are:

1) Teaching, Learning and Resources - This parameter checks the core activities in the education institutions.
2) Research and Professional Practices - Excellence in teaching and learning is closely associated with the scholarship.
3) Graduation Outcomes - Tests the effectiveness of learning/core teaching.
4) Outreach and Inclusivity - Lays special emphasis on the representation of women.
5) Perception - Importance is also given to the perception of an institution.

B. NIRF Rank Prediction

The NIRF ranking is determined by a complex process that involves the analysis of various performance metrics of educational institutions. These metrics include teaching, research, graduation outcomes, outreach, and perception. The institutions are then ranked based on their overall score, which is calculated using a weighted average of these metrics. Predicting the NIRF rank of an educational institution can be a challenging task as it involves analysing various performance metrics and their relative importance in determining the final rank. Machine learning algorithms can be used to build predictive models that can accurately predict the NIRF rank of educational institutions.

By predicting the NIRF rank of educational institutions, stakeholders such as students, parents, and educational institutions can make informed decisions about which institutions to choose or collaborate with. It can also help educational institutions identify areas where they need to improve to increase their ranking in the future.

©IJRASET: All Rights are Reserved

II. PROBLEM STATEMENT

The NIRF (National Institutional Ranking Framework) rank prediction problem statement involves predicting the rank of various educational institutions in India based on certain parameters such as teaching, research, graduation outcomes, outreach, and perception. Given a dataset containing the performance metrics of various educational institutions, the goal is to build a predictive model that can accurately predict the NIRF rank of these institutions. This can be framed as a regression problem where the target variable is the NIRF rank, and the input features are parameters like Teaching Learning and Resources (TLR) score, Research and Professional Practice (RPC) score, Graduation Outcome (GO) score, Outreach and Inclusivity (OI) score, and Perception Score. The model can be trained using various machine learning algorithms such as linear regression, decision trees, random forests, or neural networks. The performance of the model can be evaluated using metrics such as mean squared error, root mean squared error, or R-squared value. The final model can be used to predict the NIRF rank of new institutions based on their performance metrics.

A. Goals and Objectives

1) To provide stakeholders such as students, parents, and educational institutions with a reliable tool to make informed decisions regarding the choice of educational institutions.
2) To help educational institutions identify areas where they need to improve to increase their ranking in the future.
3) To provide policymakers with insights into the performance of educational institutions in India and help them make informed decisions about resource allocation and policy changes.
4) To encourage healthy competition among educational institutions and incentivize them to improve their performance in various areas.

B. Statement of scope

The scope of NIRF rank prediction involves developing predictive models to accurately predict the ranking of educational institutions in India based on various performance metrics. The performance metrics may include factors such as teaching, research, graduation outcomes, outreach, and perception.

The scope of NIRF rank prediction includes the following:

1) Data collection: Collecting data on various performance metrics of educational institutions from various sources such as NIRF reports, university websites, and government databases.
2) Data pre-processing: Cleaning and pre-processing the collected data to remove missing values, outliers, and inconsistencies.
3) Feature engineering: Identifying relevant features and engineering new features that may be useful for predicting the NIRF rank.
4) Model selection: Evaluating and selecting appropriate machine learning algorithms such as linear regression, decision trees, random forests, or neural networks for the prediction task.
5) Model training: Using the selected machine learning algorithms to train predictive models on the pre-processed data.
6) Model evaluation: Evaluating the performance of the trained models using appropriate metrics such as mean squared error, root mean squared error, or R-squared value.
7) Deployment: Deploying the final predictive model to predict the NIRF rank of new educational institutions based on their performance metrics.

The scope of NIRF rank prediction is limited to the Indian higher education system and the performance metrics used in the NIRF ranking framework. The predictive models developed through this process can help stakeholders make informed decisions about the choice of educational institutions and incentivize educational institutions to improve their performance in various areas.

C. Software context

To develop the front end of the project we require HTML, CSS, and bootstrap. To develop the back end, we need a flask framework, Flask can be used to serve machine learning models through APIs. To deploy the model, we require account on render.

D. Major constraints

NIRF rank prediction faces several major constraints that can affect the accuracy and reliability of the predictions. Some of these constraints include:

1) *Limited Data Availability:* The availability and quality of data on educational institutions can vary widely, making it challenging to build accurate predictive models. Data may be missing or incomplete, and different institutions may report data differently, leading to inconsistencies and inaccuracies in the data.
2) *Dynamic Nature Of Performance Metrics:* The performance metrics used to calculate the NIRF rank can change from year to year, making it challenging to build predictive models that can accurately capture these changes.
3) *Lack Of Transparency:* The methodology used to calculate the NIRF rank is not always transparent, making it challenging to understand how the rank is determined and which factors are most important.
4) *Limited Scope:* The NIRF ranking framework only covers higher education institutions in India, limiting the applicability of predictive models built using this framework to this specific context.

To address these constraints, it is essential to use appropriate data pre-processing techniques, carefully select and evaluate machine learning algorithms, and ensure transparency and objectivity in the methodology used to calculate the NIRF rank. Additionally, it is important to recognize the limitations of predictive models and use them as a tool to support decision-making rather than relying on them as the sole basis for decision-making.

III. RELATED WORK

There have been several studies and research papers that have focused on NIRF rank prediction using various machine learning algorithms and performance metrics. Here are a few examples of related work:

1) In a study published in the Journal of Machine Learning Research, the authors proposed a method to predict the NIRF ranking of Indian universities using a combination of machine learning algorithms and social network analysis. The study used performance metrics such as research productivity, teaching effectiveness, and perceived quality to build predictive models.
2) A study published in the Journal of Applied Research in Higher Education used data envelopment analysis (DEA) to analyse the performance of Indian universities and predict their NIRF rank. The study used performance metrics such as research output, faculty quality, and infrastructure to build predictive models.
3) In a study published in the Journal of Data Science, the authors used machine learning algorithms such as Gradient Boosting and Random Forest to predict the NIRF ranking of Indian universities. The study used performance metrics such as research, teaching, and perception to build predictive models.

These studies and others like them demonstrate the potential of machine learning algorithms and performance metrics to predict the NIRF ranking of Indian educational institutions. However, there is still a need for further research to develop more accurate and reliable predictive models and address the limitations and constraints associated with this type of prediction.

IV. METHODOLOGY

The methodology for predicting the NIRF rank of Indian educational institutions using machine learning algorithms typically involves the following steps:

1) *Data Collection:* The first step is to collect data on various performance metrics for the educational institutions, such as research output, teaching quality, graduation outcomes, and perception. The data can be collected from various sources such as NIRF reports, university websites, and government databases. We have collected dataset from Kaggle.
2) *Data Pre-Processing:* The collected data may be incomplete or contain missing values, outliers, or inconsistencies. Data pre-processing techniques such as data cleaning, normalization, and feature engineering are used to address these issues and prepare the data for model training.
3) *Feature Selection:* The next step is to select the most relevant features that have the most significant impact on the NIRF rank. Feature selection techniques such as correlation analysis, principal component analysis (PCA), and recursive feature elimination (RFE) can be used to identify the most important features.
4) *Data Splitting:* The dataset is then split into training and testing sets. The training set is used to train the model, while the testing set is used to evaluate its performance.
5) *Model Training:* Random Forest Regression involves creating an ensemble of decision trees, where each tree is trained on a subset of the data and a subset of the features. The trees then vote to make a prediction. The hyperparameters of the algorithm, such as the number of trees and the maximum depth of the trees, can be tuned to optimize the model's performance. Model evaluation: The trained model is evaluated using various performance metrics such as root mean square error (RMSE), mean absolute error (MAE), and R-squared.

   The evaluation helps to determine the accuracy and reliability of the model and identify areas for improvement.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

6) *Model Evaluation:* The trained model is evaluated using various performance metrics such as root mean square error (RMSE), mean absolute error (MAE), and R-squared. The evaluation helps to determine the accuracy and reliability of the model and identify areas for improvement.
7) *Model Deployment:* Once the predictive model has been trained and evaluated, it can be deployed for NIRF rank prediction. The model can be integrated into an existing educational analytics platform or developed as a standalone application.
8) *Continuous Improvement:* Predictive models require continuous improvement to keep up with changes in performance metrics and to address any limitations and constraints associated with NIRF rank prediction. This involves regularly updating the model with new data and evaluating its performance to ensure accuracy and reliability.

A. Random Forest Regression

Every decision tree has high variance, but when we combine all of them together in parallel then the resultant variance is low as each decision tree gets perfectly trained on that sample data, and hence the output does not depend on one decision tree but on multiple decision trees. In the case of a classification problem, the final output is taken by using the majority voting classifier. In the case of a regression problem, the final output is the mean of all the outputs. This part is called Aggregation.

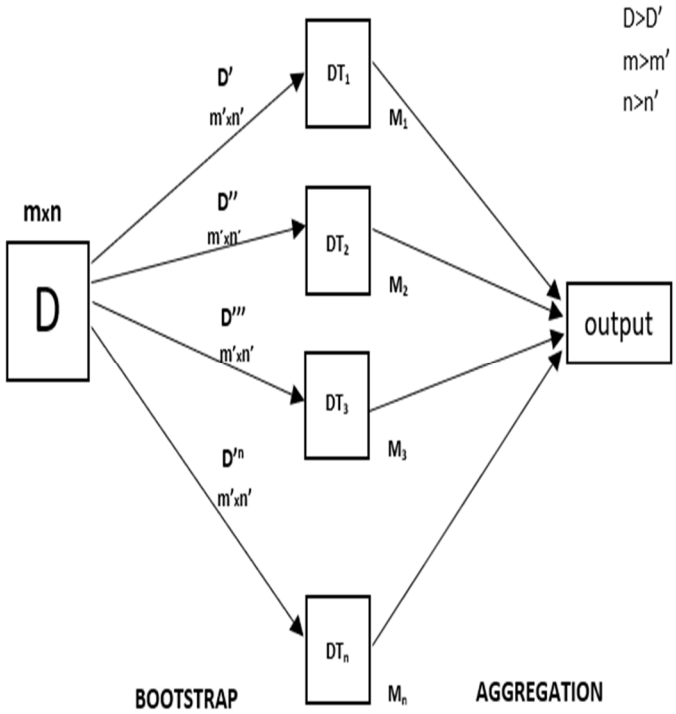


Fig. 1: Random Forest

Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap and Aggregation, commonly known as bagging. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees.
Random Forest has multiple decision trees as base learning models. We randomly perform row sampling and feature sampling from the dataset forming sample datasets for every model. This part is called Bootstrap.
Ensemble uses two types of methods:

1) *Bagging:* It creates a different training subset from sample training data with replacement & the final output is based on majority voting. For example, Random Forest.
2) *Boosting:* It combines weak learners into strong learners by creating sequential models such that the final model has the highest accuracy. For example, ADA BOOST, XG BOOST.

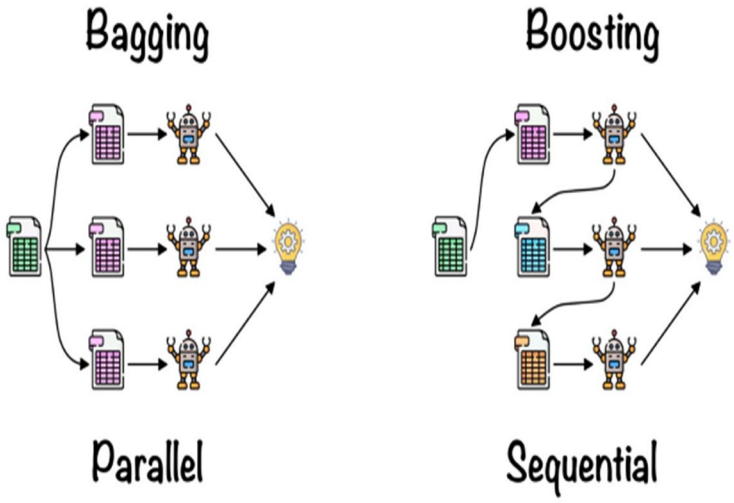


Fig. 2: Bagging and Boosting

B. Flask

Flask is a lightweight web application framework in Python that can be used for deploying machine learning models for NIRF rank prediction. Here are the steps involved in deploying a Random Forest Regression model using Flask:

1) Develop the Random Forest Regression model using Python libraries such as scikit-learn and pandas.
2) Save the trained model as a file using Python's joblib library.
3) Create a new Flask application and import the necessary libraries and the trained model file.
4) Define a route in Flask that will handle incoming requests to predict the NIRF rank.
5) In the route function, pre-process the incoming data and pass it through the trained model to make a prediction.
6) Return the predicted NIRF rank as a response to the client.
7) Test the Flask application locally to ensure that it is working correctly.
8) Deploy the Flask application to a web server or a cloud-based platform such as Heroku.
9) Test the deployed application to ensure that it is accessible and making accurate predictions.

Overall, deploying a Random Forest Regression model using Flask allows the model to be easily integrated into web applications or APIs, providing a scalable and accessible solution for NIRF rank prediction. It also provides an opportunity to further optimize and improve the model's performance by gathering real-time data and monitoring its predictions.

C. Scikit-Learn

Scikit-learn is a popular machine learning library in Python that can be used for NIRF rank prediction using Random Forest Regression. Here are the steps involved in building a Random Forest Regression model using scikit-learn:

1) Load the NIRF dataset into a pandas DataFrame.
2) Pre-process the data by cleaning, transforming, and normalizing the features. This includes handling missing values, encoding categorical variables, and scaling numeric features.
3) Split the dataset into training and testing sets.
4) Define a Random Forest Regression model using scikit-learn's RandomForestRegressor class.
5) Train the model using the training set.
6) Evaluate the model's performance on the testing set using evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and R-squared.
7) Tune the hyperparameters of the Random Forest Regression model using techniques such as grid search or randomized search.
8) Re-train the model using the optimized hyperparameters and evaluate its performance.
9) Save the trained model using Python's joblib library.
10) Use the saved model to make predictions on new data.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

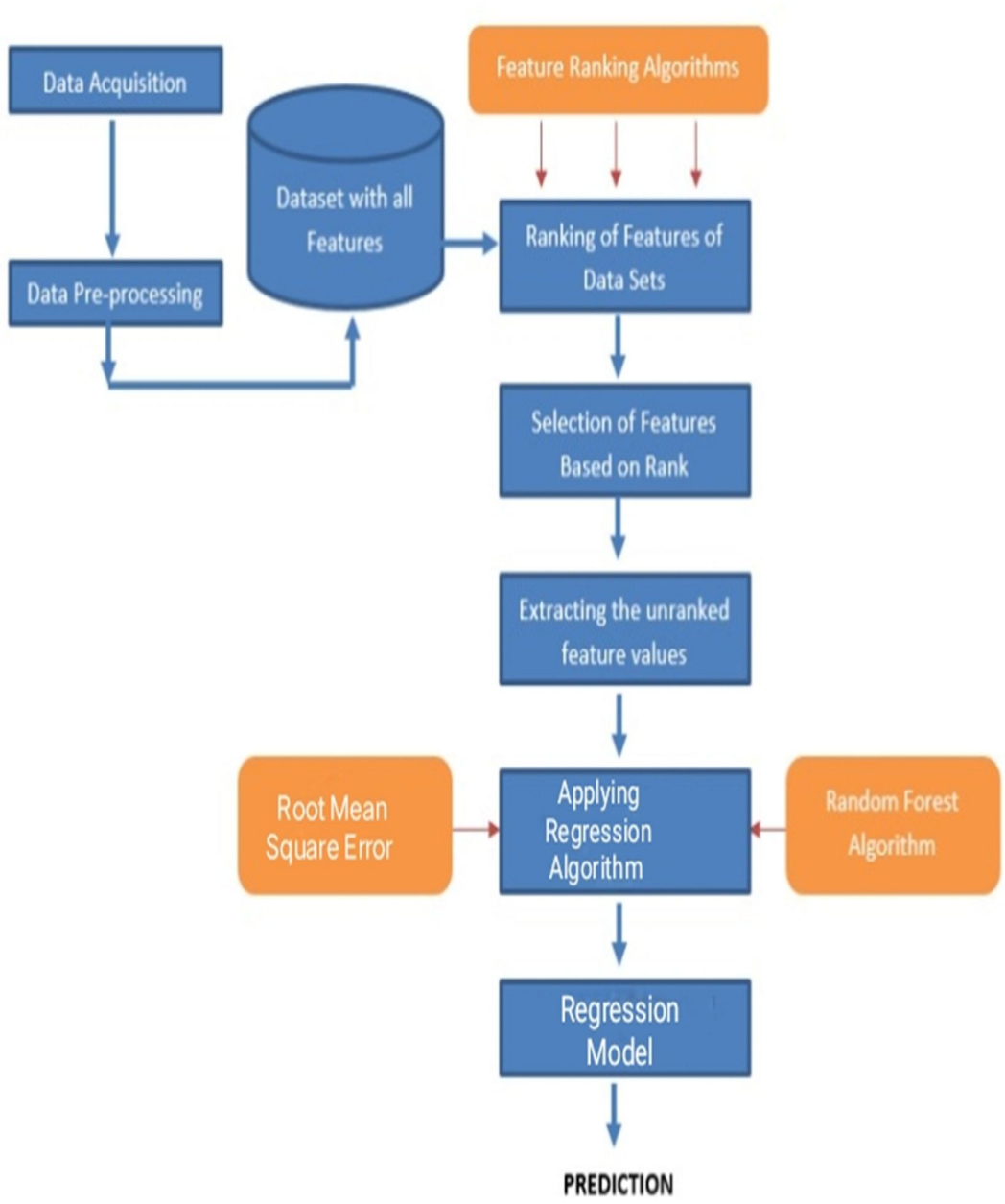


Fig. 3: System Architecture

V. PROJECT PURPOSE

The purpose of the NIRF rank prediction project is to develop a machine learning model that can predict the National Institutional Ranking Framework (NIRF) rank of Indian educational institutions based on various parameters such as research output, student and faculty quality, infrastructure, outreach, and perception. The project aims to provide insights into the factors that contribute to an institution's NIRF rank and to help identify areas for improvement. It also aims to provide a tool for policymakers, educators, and other stakeholders to make informed decisions about higher education institutions in India. By building a predictive model for NIRF ranking, the project can potentially help institutions better understand how they can improve their standing in the rankings, and guide policymakers in allocating resources to enhance the overall quality of higher education in India. Overall, the project's purpose is to contribute to the improvement of the Indian higher education system by leveraging machine learning techniques to predict NIRF rankings, and provide actionable insights for institutions and policymakers.

©IJRASET: All Rights are Reserved

VI. EXPERIMENTAL RESULT

Project link: https://lnkd.in/dVxW5nwU

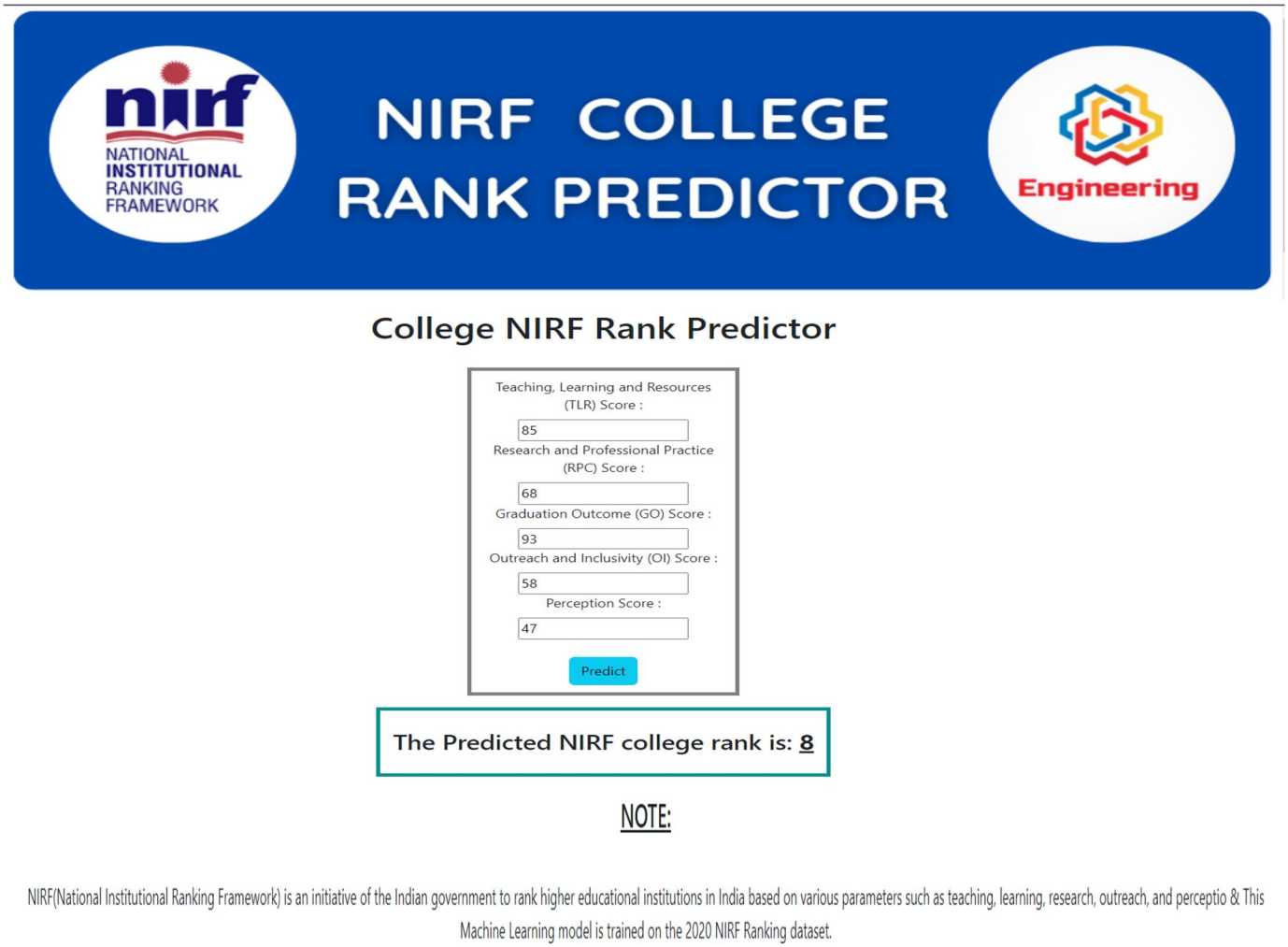


Fig. 4: Output

VII. FUTURE SCOPE

The future scope of the NIRF rank prediction project is vast and encompasses several potential avenues for further development and improvement. Here are some possible directions for future work:

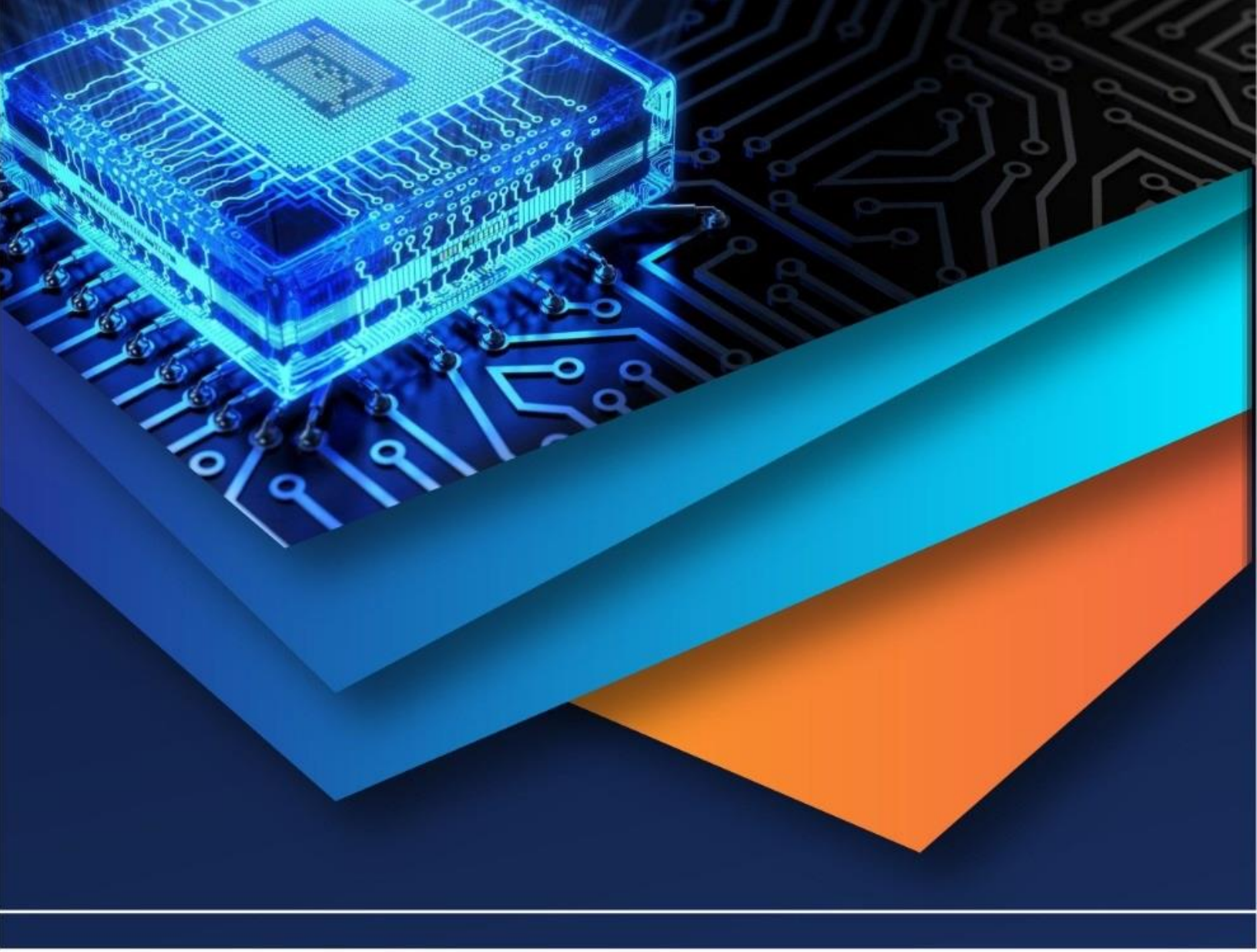
1) *Incorporating More Data Sources:* The current model uses a limited set of features to predict NIRF rankings. Incorporating additional data sources, such as student feedback, alumni performance, and industry partnerships, could potentially improve the accuracy of the model.
2) *Enriching Data With Text Analysis:* The model could potentially leverage natural language processing techniques to extract insights from unstructured data sources such as institutional websites, research papers, and news articles. This could provide a more comprehensive picture of an institution's strengths and weaknesses.
3) *Incorporating Temporal Trends:* The NIRF rankings change year to year, and institutions may have unique trends in their performance. Incorporating temporal trends in the model could help to provide more accurate predictions for future years.
4) *Exploring Alternative Machine Learning Models:* While Random Forest Regression is an effective model for predicting NIRF rankings, there are other machine learning models that could be explored, such as neural networks or gradient boosting.
5) *Building A User-Friendly Interface:* To make the model accessible to a wider audience, building a user-friendly interface that allows users to input data and receive predictions could be a valuable next step.

VIII. CONCLUSION

In conclusion, the NIRF rank prediction project aims to leverage machine learning techniques to predict the National Institutional Ranking Framework (NIRF) rank of Indian higher education institutions. The project's purpose is to provide insights into the factors that contribute to an institution's NIRF rank, identify areas for improvement, and help policymakers allocate resources to enhance the overall quality of higher education in India. By building a Random Forest Regression model using scikit-learn, the project demonstrates the potential of machine learning to predict NIRF rankings with a high degree of accuracy. The model has been trained and evaluated using a large dataset of Indian educational institutions, and its performance has been measured using evaluation metrics such as root mean squared error (RMSE). The future scope of the project is vast and encompasses several potential avenues for further development, such as incorporating more data sources, enriching data with text analysis, incorporating temporal trends, exploring alternative machine learning models, and building a user-friendly interface. Overall, the NIRF rank prediction project is a valuable contribution to the improvement of the Indian higher education system, and its predictive model provides actionable insights for institutions and policymakers.

REFERENCES

[1] National Institutional Ranking Framework (NIRF) official website: https://www.nirfindia.org/
[2] Bhatia, A., & Singh, S. P. (2021). Predicting NIRF Ranking using Machine Learning. In Proceedings of the 3rd International Conference on Computing Methodologies and Communication (pp. 547-553). Springer.
[3] Jha, P. C., & Aggarwal, M. (2019). Predicting NIRF Ranking of Indian Universities and Institutes using Machine Learning Techniques. Journal of Data Science, 17(4), 611-626.
[4] Scikit-learn documentation: https://scikit-learn.org/stable/documentation.html
[5] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer.
[6] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.
[7] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of Statistics, 1189-1232.2011.
[8] Chollet, F. (2018). Deep learning with Python. Manning Publications.
[9] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.
[10] Kingma, D. P., & Ba, J. (2014). Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980.
[11] Nigam, A., & Singh, S. (2020). Predicting NIRF Ranking of Indian Engineering Institutions using Machine Learning Techniques. International Journal of Engineering Research and Technology, 13(2), 96-102
[12] Kumar, A., & Kumar, M. (2021). NIRF Ranking Prediction using Ensemble Machine Learning Techniques. In 2021 4th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.
[13] Jain, A., & Sood, S. K. (2020). NIRF Ranking Prediction of Indian Universities using Machine Learning Algorithms. International Journal of Computer Applications, 180(7), 1-5.
[14] Agrawal, A., & Singh, S. P. (2020). Predicting NIRF Ranking of Indian Universities and Institutes using Supervised Learning Techniques. In 2020 3rd International Conference on Computing, Communication and Security (ICCCS) (pp. 1-6). IEEE.
[15] Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media, Inc.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)