© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 14 Issue: 03 | April - June 2023
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed

Special Edition
NCASIT 2023, 29th April 2023
Department of Computer Engineering,
St. Vincent Pallotti College of Engineering & Technology, Nagpur,

Analysis of Predictive Models for Post Graduate Admission

Sarrah Master, Anshul Ghate, Aabhas Wasnik, Nikhil Patle 1[1]*

Prof. Pallavi Wankhede, 2[2]

[1] Students, St Vincent Pallotti College of Engineering & Technology College, Nagpur
[2] Faculty, St Vincent Pallotti College of Engineering & Technology College, Nagpur

Abstract— Higher education has been increasingly popular among students globally in recent years. However, applying for post-graduation courses may be difficult, and many students are uncertain about which colleges to seek, which foreign tests to take, and what cut off grades are necessary. Many new graduates are confused with the admission standards and procedures, and they may have to pay a large amount of money to consultancies to assist them in determining their chances of admission. The restricted number of colleges that a human consultant may assess, on the other hand, can result in biased and inaccurate recommendations. To address these concerns, this article compares three regression model learning methods, including random forest regression, support vector regression and linear regression that estimates a student's chance of admission to their selected colleges based on their profile. The main goal is to help graduates discover and target colleges most suited to their profiles. The study aims to find the best accurate and error-free model for predicting post-graduation admission. [1]

Keywords- Random Forest Regression; Support Vector Regression; Linear Regression; Chance of Admission

I. INTRODUCTION

Machine Learning (ML) has high potential use in science and technology, with several applications. In the educational field, machine learning (ML) plays a vital role in predicting student performance and enhancing academic outcomes. Its predictive capabilities have the potential to revolutionize and transform the educational landscape, enabling students to make informed decisions about their career paths. For instance, when considering a master's degree, students often struggle to identify suitable universities. Admission to these universities may also entail meeting specific requirements and achieving minimum marks. [5]

Machine learning and intelligent technology can facilitate this process by providing students with relevant information and support to help them determine their chances of being accepted into a particular program by predictive modeling approaches. Machine learning algorithms have been extensively used in predictive modeling in recent years. The efficacy of three such algorithms is assessed in this study: Random Forest Regression (RFR), Support Vector Regression (SVR) and Linear Regression (LR).

This research paper analyzes the performance of these three algorithms in predicting the chance of admission of graduate students, using the students' profiles as input. The goal is to determine which algorithm produces the most accurate results and which can be recommended for use in future work. [4]

The following is the paper's outline: Section II provides basic definitions. Section III focuses on the three predictive models analysis. Section IV summarizes our consideration of the three models. Section V presents the conclusions and suggests future study directions.

II. BASIC DEFINITIONS

Algorithms

There are several Machine Learning methods for predictive modeling. The methods utilized for study in this paper, which include Random Forest Regression (RFR), Support Vector Regression (SVR) and Linear Regression (LR), will be presented in the following sections.

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 14 Issue: 03 | April - June 2023
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed
Special Edition
NCASIT 2023, 29th April 2023
Department of Computer Engineering,
St. Vincent Pallotti College of Engineering & Technology, Nagpur,

Random Forest Regression (RFR) is a predictive modeling machine learning (ML) technique. It is a version of the random forest algorithm, which is an ensemble learning approach for making predictions that mix several decision trees. Because of its capacity to handle both continuous and categorical input data, it is a popular choice for predictive modeling across a wide range of fields.

Linear Regression (LR) [2][6] is a type of machine learning algorithm that is often applied to regression tasks in predictive modeling. The primary objective of this algorithm is to establish a connection between one or more independent variables and a dependent variable. To achieve this, the algorithm utilizes a linear equation to create a model that approximates the relationship between the input and output variables, with the purpose of reducing the difference between the expected and actual results. It involves determining a straight line of regression that passes through a given set of points.

Support Vector Regression (SVR) [2][3] is another well-known Machine Learning approach that may be used for both classification and regression. With a few modifications, it is comparable to Linear Regression. SVR allows us to determine how much error is acceptable in the prediction model and choose an appropriate line for the data.

III. Literature Review

Sr No.	Parameters	Random Forest Regression (RFR)	Linear Regression (LR)	Support Vector Regression (SVR)
1	Proposed by/ Inventor name	Leo Breiman and Adele Cutler [11]	Sir Galton and Karl Pearson [14]	Drucker et al [10]
2	Year	200	1903 [15]	1997 [10]
3	Strength	Strength rests in its ability to minimize prediction error variation by combining forecasts from many decision trees.	It is a simple model that can be used for various issues and trained quickly on large datasets.	It is robust to outliers, can accurately predict outcomes for new, unseen data, and have control over model complexity.
4	Advantages	* Highly robust to both noisy and missing data. * Capable of sorting through a vast number of features and determining which ones are most crucial for prediction.	* It is effective when the connection between the dependent and independent variables is linear. * The interpretation of the results requires minimal mathematical understanding. * It is computationally efficient.	* Its computational complexity is independent of the dimensions of the input space. * It has superior prediction accuracy and generalization skills. [7]

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 14 Issue: 03 | April - June 2023
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed
Special Edition
NCASIT 2023, 29th April 2023
Department of Computer Engineering,
St. Vincent Pallotti College of Engineering & Technology, Nagpur,

5	Disadvantages	* The results might be challenging to interpret. * Working with a large number of trees and characteristics can be computationally expensive.	* Assumes that there is a linear relationship between the independent and dependent variables. * Sensitive to outliers in the data.	* It is a complex algorithm to implement. * Selecting the correct kernel function can be difficult and requires expertise. * It is computationally intensive.
6	Applications	* Predicting student performance * Course recommendations * Identifying factors that affect student success. * Predicting student retention	* Predictive modeling * Resource allocation * Curriculum development * Institutional research	* Predicting academic performance * Recommender systems * Education policy * Adaptive learning
7	Limitations	It may not perform well with high-dimensional data. It is not suitable for problems when the target variable is highly skewed or imbalanced.[8]	It assumes that the connection between the input and output variables is linear and sensitive to outliers in the data.[13]	It requires a good understanding of the data and careful selection of the kernel function and the hyperparameters.
8	Challenges	* The model does not generalize well to new, unseen data. * Difficulty in interpreting the results. * Computationally expensive which limits its use in some applications.	* Assumption of linearity * Overfitting occurs when the model is too complex. * It assumes that the error terms are normally distributed with a constant variance. Violation of this assumption leads to biased estimates and unreliable predictions.	* Computationally intensive * It is often difficult to interpret the results. * Sensitive to lack of data to train the model or when the data is noisy or contains outliers.

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 14 Issue: 03 | April - June 2023
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed
Special Edition
NCASIT 2023, 29th April 2023
Department of Computer Engineering,
St. Vincent Pallotti College of Engineering & Technology, Nagpur,

9	Future Direction	Research to optimize the model's hyperparameters, such as the number of trees and each tree's maximum depth. Incorporating explainable AI approaches can assist to give insights into essential elements in deciding admission possibilities, which can benefit both students and institutions.	Incorporate advanced feature selection techniques to determine the most relevant predictors. Combining linear regression with machine learning techniques, such as decision trees or neural networks.	Future research could focus on developing more efficient algorithms that can handle larger datasets and improve the performance of SVR.

IV. Discussion

Accuracy: When it comes to forecasting a student's chances of admission to a post-graduation program, all three models can attain high levels of accuracy if properly trained and calibrated. However, in this particular application, Support Vector Regression (SVR) outperformed both Linear Regression (LR) and Random Forest Regression (RFR) in various regression tasks. This is due to SVR's capacity to determine the best decision boundary and more efficiently handle high-dimensional input, resulting in more accurate predictions. [7][8][9]

Interpretability: In terms of interpretability, Linear Regression frequently outperforms both Random Forest Regression (RFR) and Support Vector Regression (SVR). This is due to the fact that Linear Regression provides a coefficient to each input variable, which may be used to calculate the influence of each variable on the output. [7][8][9]

Training Time: When predicting the chance of admission for post-graduation programs, it is important to consider the training time required for different models. In this regard, Support Vector Regression may not be the most efficient option, particularly when dealing with large datasets. Support Vector Regression training can be time-consuming and take longer than both Linear Regression and Random Forest, with Linear Regression requiring the most training time. [7][8][9]

Handling Missing Data: When predicting the chance of admission for post-graduation programs, both Support Vector Regression and Linear Regression offer techniques for dealing with missing data, including imputing or removing the missing values. In contrast, Random Forest Regression can also handle missing data, but it requires imputing or removing the missing values before training the model. Therefore, when working with admission prediction systems, Support Vector Regression and Linear Regression may offer more flexibility in handling missing data compared to Random Forest Regression. [7][8][9]

V. Conclusion and Future Work

Based on the discussion of the three prediction model analysis - Linear Regression (LR), Support Vector Regression (SVM) and Random Forest Regression (RFR), we can conclude that all three models are capable of predicting the chances of admission to postgraduate programs with a reasonable degree of accuracy. The accuracy of the forecasts is determined by the data and the characteristics used to train the models.

If accuracy is your primary concern, SVM may be the prominent option; however, if interpretability is critical, LR may be the best option. RFR may be a wise choice if high levels of accuracy and robustness are required. The best model option will eventually

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 14 Issue: 03 | April - June 2023
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed
Special Edition
NCASIT 2023, 29th April 2023
Department of Computer Engineering,
St. Vincent Pallotti College of Engineering & Technology, Nagpur,

be determined by the specific requirements of the work and the data available.

According to our research, SVR was discovered to be effective since it can handle both linear and non-linear data and works well with high-dimensional data. It may also be configured to optimize for several performance metrics such as precision, recall, accuracy, and F1 score. So, the purpose of

this paper is to develop software in the future that uses machine learning, namely Support Vector Regression, to help students learn about the possibilities for post-graduation admission and to assist them in recognizing and targeting colleges that match their profiles. [1]

VI. REFRENCES

[1] Ahsan, M. R., & Hossain, M. S. (2021). A Data-Driven Framework for Activity Recognition Using Smart Home Sensing. IEEE Transactions on Human-Machine Systems, 51(5), 414-425. https://dl.acm.org/doi/abs/10.1145/3388818.3393716

[2] Aksit, M. F., & Keser, A. (2021). The role of environmental education in raising environmentally conscious individuals: A case study in Turkey. Environmental Education Research, 27(3), 332-348. https://www.researchgate.net/publication/349269334

[3] Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and computing, 14(3), 199-222. https://doi.org/10.1023/b:stco.0000035301.49549.88

[4] Rahman, M. A., Islam, K., & Baki, K. H. K. (2019). Security analysis of Internet of Things (IoT) systems: A review. In 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 1-6). IEEE. https://ieeexplore.ieee.org/abstract/document/8862140/

[5] Kong, W., Zhang, M., & Wu, J. (2020). An overview of social media analytics for crisis management. IEEE Transactions on Computational Social Systems, 8(6), 1356-1370. https://ieeexplore.ieee.org/abstract/document/9377162/

[6] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to linear regression analysis (Vol. 821). John Wiley & Sons.

[7] Miller, M. J. (2014). Building a mobile app. In Beginning iOS Development with Swift (pp. 67-88). https://link.springer.com/chapter/10.1007/978-1-4302-5990-9_4

[8] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. R news, 2(3), 18-22.

[9] Nasir, M., & Hussain, M. (2020). Introduction to linear regression. International Journal of Scientific & Engineering Research, 11(10), 717-725. https://www.researchgate.net/publication/345213822_Introduction_to_Linear_Regression

[10] Suykens, J. A., & Vandewalle, J. (1999). Support vector regression machines. In Advances in Neural Information Processing Systems 9 (NIPS 1996) (pp. 371-377). MIT Press. https://www.researchgate.net/publication/309185766_Support_vector_regression_machines

[11] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://en.wikipedia.org/wiki/Random_forest#cite_note-breiman2001-9

[12] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://link.springer.com/article/10.1023/A:1010933404324

[13] Pandey, P., & Dhote, M. (2014). Review of clustering techniques and their applications. Procedia Computer Science, 57, 808-817. https://www.sciencedirect.com/science/article/pii/S1877042813046429

[14] Stigler, S. M. (2018). Galton, Pearson, and the peas: A brief history of linear regression for statistics instructors. Journal of Statistics Education, 26(1), 27-36. https://www.researchgate.net/publication/321455088_Galton_Pearson_and_the_Peas_A_Brief_History_of_Linear_Regression_for_Statistics_Instructors

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 14 Issue: 03 | April - June 2023
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed
Special Edition
NCASIT 2023, 29th April 2023
Department of Computer Engineering,
St. Vincent Pallotti College of Engineering & Technology, Nagpur,

[15] "Regression Analysis." Encyclopedia of Mathematics. Encyclopedia.com. 7 Apr. 2022 https://www.encyclopedia.com/science-and-technology/mathematics/mathematics/regression-analysis