e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:03/March-2024 Impact Factor- 7.868 www.irjmets.com

A MACHINE LEARNING MODEL FOR PREDICTION OF GRADUATE ADMISSIONS

P Honey Diana*1, Vangipurapu Jyothi Swarup*2, Ammula Pranay*3, Challa Shashidhar*4

*1Assistant Professor, Department Computer Science And Engineering, Malla Reddy College Of Engineering And Technology, Hyderabad, India.

*2,3,4Final Year Students, Department Computer Science And Engineering, Malla Reddy College Of Engineering And Technology, Hyderabad, India.

ABSTRACT

Prospective graduate students always face a dilemma deciding universities of their choice while applying to master's programs. While there are a good number of predictors and consultancies that guide a student, they aren't always reliable since decision is made on the basis of select past admissions. In this project, we present a Machine Learning based method where we use different algorithms, such as Random Forest, SVM, Liner Regression, given the profile of the student to predict colleges based on their profile. We then compute different models and compare their performance to select the best performing model. Results then indicate if the university of choice is can be accepted or rejected. Using this method user can enter Various factors as input like GRE, TOFEL, B.Tech percentage, term applying for , total technical papers published. Based on these features machine learning model can be selected and predictions of which college is possible for applying for MS is calculated and displayed to user.

I. INTRODUCTION

Prospective graduate students embarking on the journey of applying to master's programs often find themselves grappling with the daunting task of selecting the universities that best align with their aspirations and academic pursuits. While there exists a plethora of predictors and consultancies purporting to offer guidance in this decision-making process, their reliability remains a subject of skepticism, primarily due to their reliance on a limited pool of past admissions data. In response to this inherent challenge, this project presents an innovative machine learning-based approach aimed at providing prospective students with a more robust and data-driven means of university selection. Through the utilization of diverse machine learning algorithms such as Random Forest, Support Vector Machines (SVM), and Linear Regression, we endeavor to leverage the profile information provided by students to predict the likelihood of acceptance at various universities. This approach involves the collection and preprocessing of comprehensive datasets encompassing a wide range of variables, including GRE scores, TOEFL scores, B.Tech percentage, the term for application, and total technical papers published. By training and evaluating multiple models using these datasets, we aim to compare their performance metrics meticulously, ultimately identifying the most effective model for predicting university acceptances. The results derived from these models serve as invaluable insights for prospective students, offering clarity on whether their desired universities are likely to extend an acceptance or rejection. Armed with this predictive capability, users can make more informed decisions regarding their applications, thereby maximizing their chances of securing admission to their preferred MS programs. Moreover, the user-friendly interface facilitates the seamless input of various factors, empowering students to tailor the predictions to their specific profiles and preferences. We aim to bring students closer to their university of choice through a robust evaluation of their profiles. A good number of predictors and consultancy services fail in understanding the admission procedure and either suggest extremely ambitious schools or lower ranked ones. In this paper, we have included parameters that are all relevant for graduate admissions. Barring a few exceptional cases in which a student may unexpectedly fetch an admit in a top school, most of the results are as expected and give a fair idea about the selection criteria. In further sections, we explore the different models and try to understand their functioning.

II. LITERATURE REVIEW

There are a number of predictors that evaluate a profile based on past admissions. With the scheme of evaluation changing every year and with stricter guidelines, requirements vary considerably. One significant work observed in the same direction is [2]. The classification algorithms used in [2] uses data from an old format of UCLA graduate dataset. The test scores and other parameters are more suited for US student applications. In our work, we use Regression models that give a definite value between 0 and 1 which is useful in understanding a student's profile. It also helps in analysing how important a particular parameter is for the admission and greatly affects the output value when one parameter is changed. Our dataset was created for the defined problem and is original in the true sense.

There are many examples of the application of machine learning techniques to analyze data and other information in the context of educational settings. This area of study is generally known as "educational data mining" (EDM) and it is a recently emergent field with its own journals [15], conferences [16] and research community [17,18]. A subset of EDM research that focuses on analyzing data in order to allow institutions of higher education better clarity and predictability on the size of their student bodies is often known as enrollment management. Enrollment management is "an organizational concept and systematic set of activities whose purpose is to exert influence over student enrollment" [7].

Below, we provide multiple examples of research by others that combines aspects of enrollment management with applications of data science techniques in various educational settings. These are applications of machine learning to: college admission from the student perspective; supporting the work of a graduate admission committee at a PhD granting institution; predicting student graduation time and dropout; monitoring student progress and performance; evaluating effectiveness of teaching methods by mining non-experimental data of student scores in learning activities; and classifying the acceptance decisions of admitted students.

There are many websites which purport to predict college admission from the perspective of an aspiring student. A few examples are go4ivy.com1 , collegeai.com2 , project.chanceme3 , and niche.com 4 . Websites such as these claim to utilize artificial intelligence to predict a student's likelihood of being admitted to a college of their choice without providing specific details about software used and techniques implemented. Our work differs in that we are predicting the likelihood of a student accepting an admission offer from a college not providing an estimate of the chances of a student's admission to college. Unlike these websites, we provide a complete description of our materials and methods below.

In the work of Waters and Miikkulainen [19], machine learning algorithms were used to predict how likely an admission committee is to admit each of 588 PhD applicants based on the information provided in their application file. Students whose likelihood of admission is high have their files fully reviewed to verify the model's predictions and increase the efficiency of the admissions process by reducing the time spent on applications that are unlikely to be successful. Our research differs from this work in multiple ways; our setting is at the undergraduate level, we classify decisions by the students not decisions by the college; our dataset is an order of magnitude larger; and the feature being optimized is incoming class size not time spent on decision making.

III. METHODOLOGY

Dataset:

Data set is collected form online. At the time of writing this paper, the dataset has over 400 downloads and more than 2000 views. This dataset contains parameters that are considered carefully by the admissions committee. First section contains scores including GRE, TOEFL and Undergraduate GPA. Statement of Purpose and Letter of Recommendation are two other important entities. Research Experience is highlighted in binary form. All the parameters are normalized before training to ensure that values lie between the specified range. A few profiles in the dataset contain values that have been previously obtained by students. A unique feature of this dataset is that it contains equal number of categorical and numerical features. The data has been collected and prepared typically from an Indian student's perspective. However, it can also be used by other grading systems with minor modifications. A second version of the dataset will be released which will have an additional two hundred entries.

Data preprocessing:

In this step data is pre processed by removing unwanted data and NAN values and using features and labels which are useful to fit in to algorithm and then process data for prediction.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:03/March-2024 Impact Factor- 7.868 www.irjmets.com

Data split Test training:

In this stage data is divided in to test and train values using train test split function and store features and labels in to test train values. Train set is 30 percent of test set data which is used for checking accuracy of the dataset.

Model Training:

In this stage different algorithms are used to check which algorithm provides best accuracy and select one algorithm to use that for fitting features and labels and then run algorithm in this way model is trained.

Prediction and accuracy:

In this stage new input or test set is taken as input and given as input to predict function of the algorithm and then result of labels are as output of the algorithm.

SDLC METHDOLOGIES:

This document play a vital role in the development of life cycle (SDLC) as it describes the complete requirement of the system. It means for use by developers and will be the basic during testing phase. Any changes made to the requirements in the future will have to go through formal change approval process.

SPIRAL MODEL was defined by Barry Boehm in his 1988 article, "A spiral Model of Software Development and Enhancement. This model was not the first model to discuss iterative development, but it was the first model to explain why the iteration models.

As originally envisioned, the iterations were typically 6 months to 2 years long. Each phase starts with a design goal and ends with a client reviewing the progress thus far. Analysis and engineering efforts are applied at each phase of the project, with an eye toward the end goal of the project.

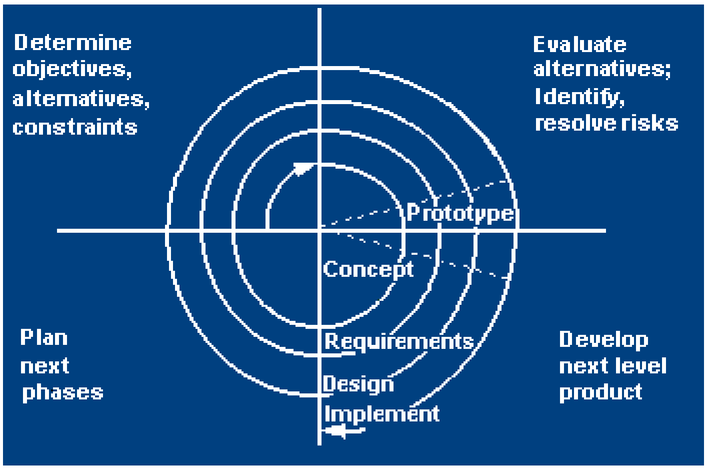

Fig 1: Spiral model

The steps for Spiral Model can be generalized as follows:

- The new system requirements are defined in as much details as possible. This usually involves interviewing a number of usersrepresenting all the external or internal users and other aspects of the existing system.
- A preliminary design is created for the new system.
- A first prototype of the new system is constructed from the preliminary design. This is usually a scaled-down system, and represents an approximation of the characteristics of the final product.
- A second prototype is evolved by a fourfold procedure:
- Evaluating the first prototype in terms of its strengths, weakness, and risks.

1. Defining the requirements of the second prototype.
2. Planning a designing the second prototype.
3. Constructing and testing the second prototype.

- At the customer option, the entire project can be aborted if the risk is deemed too great. Risk factors might involve development cost overruns, operating-cost miscalculation, or any other factor that could, in the customer's judgment, result in a less-than-satisfactory final product.

- The existing prototype is evaluated in the same manner as was the previous prototype, and if necessary, another prototype is developed from it according to the fourfold procedure outlined above.
- The preceding steps are iterated until the customer is satisfied that the refined prototype represents the final product desired.
- The final system is constructed, based on the refined prototype.
- The final system is thoroughly evaluated and tested. Routine maintenance is carried on a continuing basis to prevent large scale failures and to minimize down time.

IV. RESULTS AND DISCUSSION

Table 1: Sample input data

	gre_score	gre_score_quant	gre_score_quant	test_score_toefl	undergraduation_score	work_ex	papers_published	ranking
4603	314	168	146	109.0	3.26	13	0	66
7734	300	151	149	101.0	2.72	32	0	30
828	306	163	143	101.0	2.38	45	0	89
5472	313	164	149	108.0	2.35	21	0	35
6108	309	161	148	106.0	2.10	53	0	118

test_accuracy: 0.6916167664670658

train_accuracy: 0.9868796349115802

test_f1_score: 0.6808322266489598

train_f1_score: 0.9859798841816519

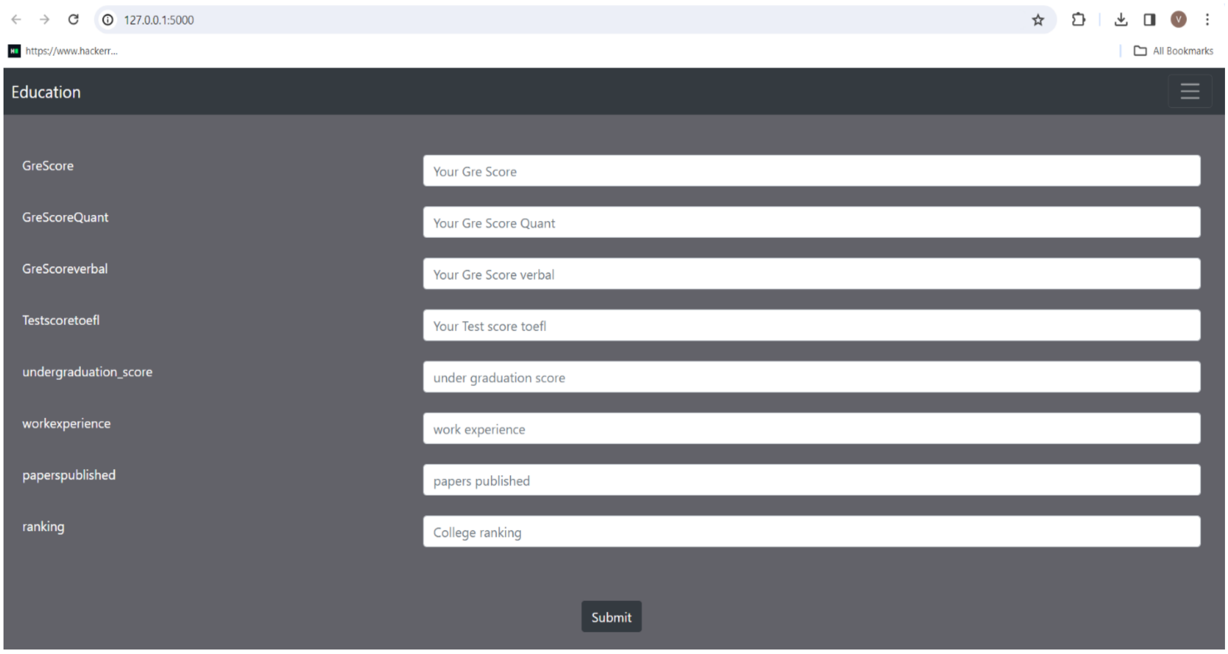

Fig 2: Output screen1

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:03/March-2024 Impact Factor- 7.868 www.irjmets.com

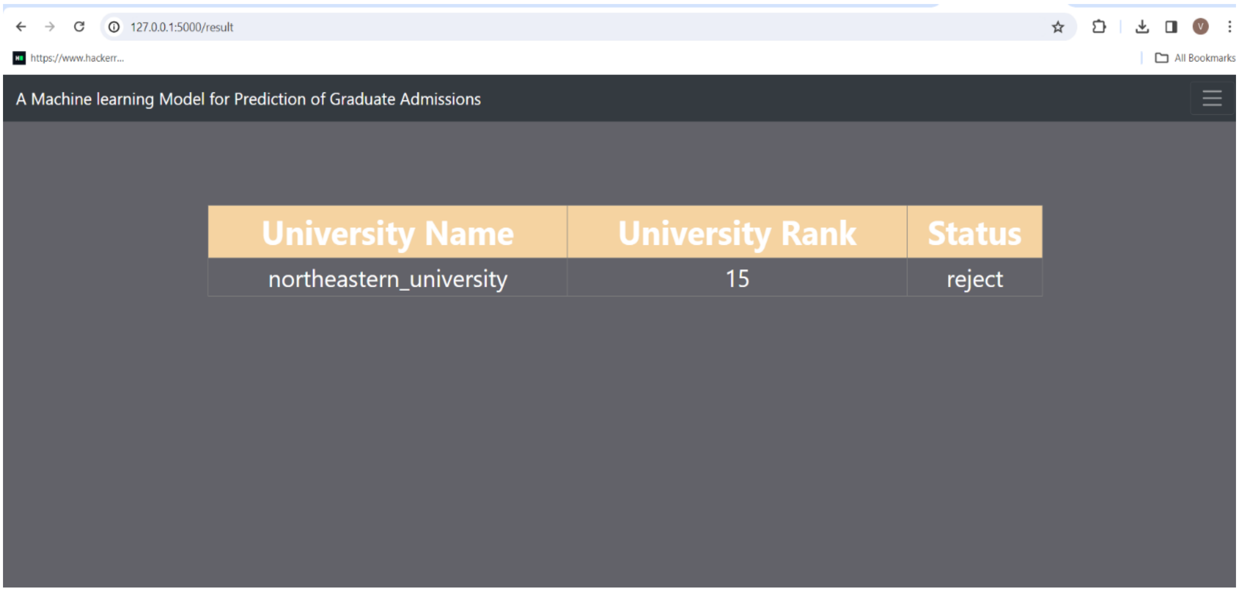

Fig 3: Output screen2

V. CONCLUSION

Even though at educational field Machine Learning is still emerging, its effectiveness to analyze information is notorious. Through the analysis, predictions, and visualizations of information, for higher education' directors obtain a greater understanding of the different variables involved when making a decision. Machine Learning supports this process providing various algorithms suitable to the different kinds of data and the different kinds of predictions required. We employ three supervised classification algorithms: Decision Trees, Random Forests and Logistic Regression, where Random Forest performs the best outcomes

VI. REFERENCES

[1] A. Lašáková, L. Bajzíková, and I. Dedze, "Barriers and drivers of innovation in higher education: Case study-based evidence across ten European Universities," Int. J. Educ. Develop., vol. 55, pp. 69–79, May 2017.

[2] Y. V. N. Acevedo and C. E. M. Marín, "System architecture based on learning analytics to educational decision makers toolkit," Adv. Comput. Sci. Eng., vol. 13, no. 2, pp. 89–105, 2014.

[3] A. R. T. Góes, M. T. A. Steiner, and P. J. S. Neto, "Education quality measured by the classification of school performance using quality labels," Appl. Mech. Mater., vols. 670–671, pp. 1675–1683, Oct. 2014.

[4] H. Lounis and T. Fares, "Using efficient machine-learning models to assess two important quality factors: Maintainability and reusability," in Proc. Joint Conf. 21st Int. Workshop Softw. Meas. 6th Int. Conf. Softw. Process Product Meas., Nov. 2011, pp. 170–177.

[5] Y. Nieto, V. García-Díaz, C. Montenegro, and R. G. Crespo, "Supporting academic decision making at higher educational institutions using machine learning-based algorithms," Soft Comput., vol. 23, no. 12, pp. 4145–4153, 2018.

[6] K. V. Pincus, D. E. Stout, J. E. Sorensen, K. D. Stocks, and R. A. Lawson, "Forces for change in higher education and implications for the accounting academy," J. Accounting Educ., vol. 40, pp. 1–18, Sep. 2016.

[7] A.-P. Pavel, A. Fruth, and M.-N. Neacsu, "ICT and e-learning—Catalysts for innovation and quality in higher education," Procedia Econ. Finance, vol. 23, pp. 704–711, Jan. 2015.