



Predicting the Student Performance by using Machine Learning

Challagulla. Lakshmi Saranya¹, Dr. K. V. Sambasiva Rao

¹PG Student, Dept of CSE, NRI Institute of Technology, Vijayawada, A.P

²Professor, Dept of CSE, NRI Institute of Technology, Vijayawada, A.P

Corresponding Authors Email Id: Saranyachallagulla14@gmail.com¹, kvrsrao@nriit.edu.in²

To Cite this Article

Challagulla. Lakshmi Saranya and Dr. K. V. Sambasiva Rao. Predicting the Student Performance by using Machine Learning. International Journal for Modern Trends in Science and Technology 2023, 9(08), pages. 19-23. <https://doi.org/10.46501/IJMTST0908004>

Article Info

Received: 03 July 2023; Accepted: 28 July 2023; Published: 30 July 2023.

Copyright © 2023 Challagulla. Lakshmi Saranya et al. This is an open access article distributed under the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

The student's academic work is often kept in the student administration system in a variety of file types, document types, record types, image types, and other media types. The information gathered from students may be very helpful. However, traditional statistical methods and database management technologies struggle to make sense of the ever-growing student data. Universities need a tool because it helps them collect data. The success of pupils may be predicted with the use of this useful information. The *Leistungs analyze Learning Outcomes* framework is designed to help students achieve their goals across several learning domains. This study suggests a comprehensive framework as a rule-based recommendation technique for not only analyzing and anticipating students' progress, but also for giving their justifications for doing so. Students, educators, and parents may learn as much as possible with the help of the suggested system, which takes into account students' demographic information, academic performance, and psychological characteristics. To improve the precision of academic forecasts using a variety of robust data-mining methods. When a student's restrictions are taken into account, the system is able to provide useful guidance. The case study conducted on 200 students shows that the suggested solution outperforms the existing framework.

Keywords: Recommender System, Performance Analysis, Statistical Techniques, Machine Learning, Classification

INTRODUCTION:

High-quality graduates in the areas of academics, practical knowledge, personal growth, and creative thinking are the goal of any school, university, and other educational institution. The output of pupils at each given university, high school, or other educational establishment must be assessed for this purpose. Assessments of academic achievement may be made using a wide variety of tools. Each student is unique and hence may have a varying level of academic success.

Over the last decade, a surge in new colleges and other institutions of higher learning has occurred. They consistently produce a sizable number of new graduates each year. The issue of dropouts, poor academic performance, and the jobless remains a challenge for universities and institutions, despite their improved ability to pursue teaching. Understanding and analyzing poor performance characteristics is a dynamic and ongoing procedure that obscures insights on past and current student performance in the classroom. Effective

resources are required for accurate scientific assessment and forecasting of student performance. While schools and groups gather this information, it is not utilised to improve students' academic outcomes or inform policy changes.

Research and predictions based on data mining have shown impressive results in a variety of areas, including spotting fraud, predicting consumer behaviour, analysing the financial market, assessing loan applications, predicting loan defaults, gauging the risk of an incursion, and valuing real estate. It has the potential to be as effective in the classroom. It's a powerful way for discovering previously hidden relationships and insights that may otherwise be uncovered and understood via statistical analysis.

2. LITERATURE REVIEW

When it comes to producing accurate findings for the future, Educational Data Mining tends to focus on prediction rather than creating precise numbers. Regular analysis is essential in education to stay up with the ever-evolving trends in curriculum design. In this work, we provide the preliminary findings of a research effort into the use of educational data mining, based on the collection of massive data sets from BMS College of Engineering. Its purpose is to protect students from failing courses and expose the great potential of data mining applications for university administration. Classification and regression issues were analysed using a wide range of machine learning models, all of which were run on the WEKA platform. Classification and regression algorithms were evaluated for their ability to accurately predict a student's grade point average and individual topic grades. In the former case, a voting-based ensemble technique was used to the classification algorithms. The outcomes of classifying and comparing them according to certain criteria are detailed in this work. When applying regression analysis to the latter, the best results were obtained using the Random Forest Algorithm[1]. The academic achievement of a school's students is an important indicator of the school's quality as a whole. The academic outcomes of at-risk pupils may be vastly improved via early identification and the implementation of preventative interventions. In recent years, predictions made using machine learning methods have become more common. While there are many examples of success stories in the

published works, the majority of these methods are only within the reach of teachers who are "computer science" or, more specifically, "artificial intelligence" literate. Indeed, there are numerous choices to be made in the successful and efficient use of data mining technologies, from the definition of student achievement to the selection of student traits to concentrate on to the selection of the most relevant machine learning approach for the task at hand. The purpose of this research is to give a methodical framework for teachers who want to use data mining to foresee their students' achievements. To achieve this goal, we have systematically evaluated the relevant literature and assembled the state-of-the-art, covering all relevant topics and providing detailed explanations and justifications for all relevant judgements and criteria. This research will make it simpler for teachers to use data mining methods, unlocking their full potential in the classroom[2]. There is a growing need for predictive analytics software in universities nowadays. All educational levels may benefit from predictive analytics' utilisation of cutting-edge analytics techniques like machine learning deployment. As most teachers already know, students' grades are one of the most important performance indicators that can be used to track their progress in the classroom. Over the last decade, several variations on machine learning methods have been presented for use in academic settings. Yet, improving the efficiency of forecasting students' grades while dealing with unbalanced datasets presents significant difficulties. As a result, this study provides a thorough evaluation of machine learning algorithms for forecasting students' final marks in first-semester courses. In this study, we will focus on two separate modules. First, we use a dataset consisting of 1282 actual student course grades to evaluate the accuracy performance of six popular machine learning techniques: Decision Tree (J48), Support Vector Machine (SVM), Naive Bayes (NB), K-Nearest Neighbour (kNN), Logistic Regression (LR), and Random Forest (RF). We then introduced a multiclass prediction model that combines the oversampling technique Synthetic Minority Oversampling Technique (SMOTE) with two different feature selection strategies in order to mitigate the overfitting and misclassification outcomes that arise from multi-classification with unequal representations of certain classes. The obtained findings demonstrate a

considerable improvement with an f-measure of 99.5% when the suggested model combines with RF. In order to improve the prediction performance model for unbalanced multi-classification of student grades[3], our suggested model shows similar and promising outcomes. There are a lot of factors that a private institution has to think over before enrolling a student. All enrolled students are expected to remain until they graduate, perform well academically, and finish their degrees on schedule. Starting with the admission of new students, private institutions must make decisions about which applicants to accept in order to meet the quality education objectives of the course of study. This research intends to examine the significance of class and class order variables in predicting students' duration of stay and academic achievement as measured by their ability to graduate. The chosen approach is a kind of "feature extraction," which is a broader category of methods. In this investigation, we compare the 2 and random forest approaches to the rank methods' information gain and gain ratio. Students in a management programme at a private institution in Indonesia provide data for a dataset with 7676 observations covering the years 2010 to 2021. The academic admissions department of the institution is the primary source of information for this research. All methods agreed that cumulative GPA/IP was the most important indicator of whether or not a student would complete their programme. Votes are divided by high school of origin, Selection Test score, and gender. The case specificity of this investigation to Indonesia makes it a one-of-a-kind study[4]. Data is being produced at an unprecedented scale due to the rise of mobile devices and advancements in information and communication technology (ICT). The educational benefits of big data have begun to emerge. Several advances in the field of Big data in research on higher education and educational technology are presented in this study. There has not yet been a systematic study or survey of Big data in education. A small number of writers have published comprehensive literature evaluations that cover just one or two of these areas: educational mining or learning analytics, with narrow explorations of topics like Big data technologies without an educational emphasis or social media data in education. These literature evaluations tend to be too brief to give comprehensive analyses of Big Data in education. In this study, we provide a systematic

overview of the literature concerning the existing and developing Big data paradigms in the field of education. This survey is broken down into five sections: Part one provides a categorization and overview of Big education research to demonstrate the breadth of this field; Part two discusses the various data sources from education platforms or systems, such as learning management systems (LMS), massive open online courses (MOOC), learning object repositories (LOR), OpenCourseWare (OCW), and open educational resources (OER); Part three provides a brief summary of the paper's central arguments and findings.

3. PROPOSED METHOD

Any school, university, or other educational establishment may greatly benefit from its students by generating graduates who are outstanding in academics, practical knowledge, self-improvement, and creative ideas. To achieve this goal, it is vital to assess the results achieved by students enrolled in any educational programme. In order to provide students, educators, and parents with comprehensive data, the suggested system analyses pupil demographics, academic performance, and psychological characteristics. As a means of improving the precision of academic forecasts using various robust data-mining methods. In spite of the student's shortcomings, the system is able to provide useful guidance. The results of a case study involving 200 students show that the suggested approach outperforms the existing framework.

Algorithms Used:

Decision Tree:

Step-1: Begin the tree with the root node, says S, which contains the complete dataset.

Step-2: Find the best attribute in the dataset using Attribute Selection Measure (ASM).

Step-3: Divide the S into subsets that contains possible values for the best attributes.

Step-4: Generate the decision tree node, which contains the best attribute.

Step-5: Recursively make new decision trees using the subsets of the dataset created in step -3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

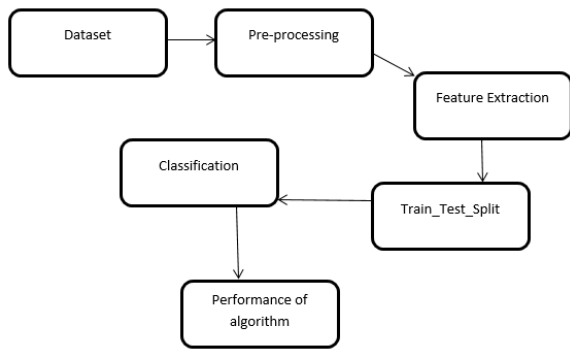
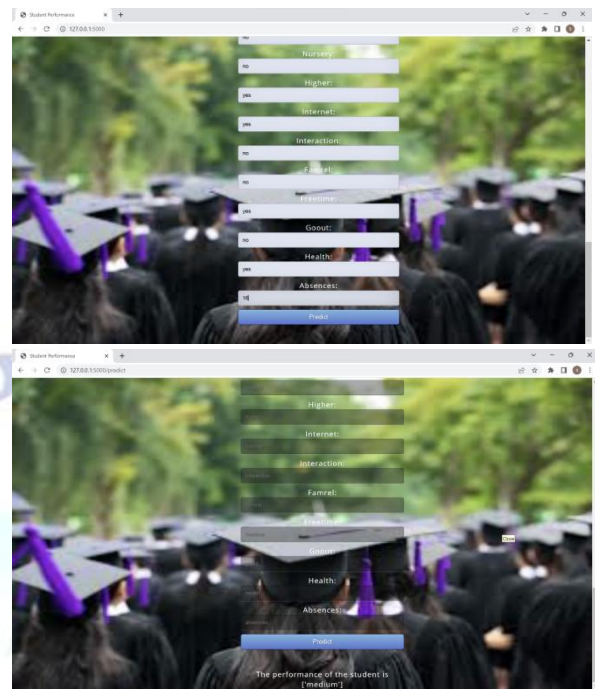


Fig1: System Architecture

4. EXPERIMENTAL RESULTS



5. CONCLUSION

The suggested model outperforms Student Performance throughout a full academic year. The best and most accurate results for a given collection of events are found in the training and validation datasets. We find that the K-Nearest Neighbours, Decision Tree Classifier model achieves the highest levels of perfection (89.74%) and accuracy (94.44%). Basic factors like as presence, homework, and presentation have been ignored. The precision would improve if we took into account every possible factor. When compared to prior efforts, the results obtained using this model are superior. Whatever the outcome, our effort will benefit kids, educators, parents, and educational institutions. If a university's department heads had access to a system that indicated which students need further assistance from them, they might take immediate action to raise those students' performance. The names of all the stragglers in his class would be available to him in a heartbeat. A psychiatrist's assistance would greatly increase the usefulness of our data collection. With their help, we can improve our statistics by include students' psychological information. Next, we'll include other successful prediction algorithms from the fields of machine learning and data mining, such as Naive Bayes, support vector machines, etc., to improve accuracy and add new features. After that, we can make dynamic predictions about the remainder of a student's life. We propose to use a web-based application to put the concept into action,

and in the process, we'll add additional features that will be very useful to kids, educators, and parents.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

- [1] Brijesh Kumar Bhardwaj, Saurabh Pal. "Data Mining: A prediction for performance improvement using classification", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 9, No. 4, April 2011.
- [2] Han, J. and Kamber, M., "Data Mining: Concepts and Techniques", 2nd edition. The Morgan Kaufmann Series in Data Management System, Jim Gray, Series Editor, 2006.
- [3] Pandey, U. K. and Pal, S., "Data Mining: A prediction of performer or underperformer using classification", (IJCSIT) International Journal of Computer Science and Information Technology, Vol. 2(2), 2011, 686- 690, ISSN: 0975-9646.
- [4] Alaa el-Halees, "Mining Students Data to Analyze elearning Behavior: A case Study", 2009.
- [5] Khalid Alkhatib, Hassan Najadat, Ismail Hmeidi and Mohammed L. Ali Shatnawi, "Stock price Prediction Using K-Nearest Neighbor (knn) Algorithm" Vol. 3 No. 3; March 2013
- [6] Nafis Neehal, Book-"Machine Learning Algorithm", 2018(Bangla), Page-132, ISBN: 978984-8042-02-1.
- [7] Varapron P. et al. Using Rough Set theory for Automatic Data Analysis. 29th Congress on Science and Technology of Thailand. 2003.
- [8] Merceron, A. and Yacef, K., "Educational Data Mining: a Case Study" In Proceedings of the 12th International Conference on Artificial Intelligence in Education AIED 2005, Amsterdam , The Netherlands, IOS Press. 2005.
- [9] Romero, C., Ventura, S. and Garcia, E., "Data mining in course management systems: Moodle case study and tutorial". Computers & Education, Vol. 51, No. 1. pp. 368384. 2008.
- [10] Minaei-Bidgoli B., Kashy, D. Kortemeyer G., Punch W., "Predicting Student Performance: An Application of Data Mining Methods with an Educational Web-Based System". In the Processing of 33rd ASEE/IEEE conference of Frontiers in Education. 2003.
- [11] Beikzadeh, M. and Delavari, N., "A New Analysis Model for Data Mining Processes in Higher Educational Systems". On the proceedings of the 6th Information Technology Based Higher Education and Training 7-9 July 2005.
- [12] Waiyamai, K. "Improving Quality of Graduate Students by Data Mining" Department of Computer Engineering, Faculty of Engineering. Kasetsart University , Bangkok, Thailand. 2003.
- [13] International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 13, Number 7 (2017), pp. 1735-1741 © Research India Publications, <http://www.ripublication.com>.
- [14] Amjad Abu Saa. "Educational Data Mining & Students' Performance Prediction". (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 5, 2016.
- [15] Keno C. Piad, Menchita Dumlao, Melvin A. Ballera, Shaneth C. Ambat, "Predicting IT Employability Using Data Mining Techniques," in third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC), 2016.