

Students Academic Performance Analysis Using Machine Learning

¹Dr. SHANKARAGOWDA B B, ²JEERLA BASAVARAJ NANDITHA

¹HOD & Associate Professor , Department of Master of Computer Application, BIET, Davanagere

²Student, Department of MCA, BIET, Davanagere

Abstract : *To far, a number of tools have been developed for improving and assessing student performance. These technologies' outputs can aid in decision-making, which raises student achievement. The tools and methods that have been developed in this field are surveyed in this paper. In order to analyze and forecast the outcomes based on many variables that can enhance student performance, this paper uses a machine learning method. This study also implies that cognitive modeling is a more effective method for enhancing decision-making skills and is valuable for creating high-quality software and performance analysis tools.*

Keywords: Student's performance analysis, Machine learning tools.

I.INTRODUCTION

As we all know, the competition is growing daily and there is a dearth of high-quality education in today's world. Therefore, it is necessary to take excellent action to raise both the educational and student standards. For this, a number of philosophers occasionally offer guidelines and benchmarks for performance enhancement. The systems are still falling behind. Therefore, researchers came to the conclusion that technology can be a significant component in analyzing the defects in the current system and why we are falling behind. Additionally, the use of technology facilitates decision-making by producing reports and graphs for analytical purposes.

The key resource for different universities is its students. With their academic achievement, students and universities play a crucial role in producing graduates of the highest caliber. The degree of educational objective achievement that can be evaluated and tested by exams, assessments, and other forms of

measurements is known as academic performance achievement. However, there are differences in academic performance achievement because different student types may achieve at varying levels.

In a student management system, the academic achievement of students is typically saved in a variety of formats, including files, documents, records, photographs, and other formats. Data from the accessible pupils could be extracted to provide informative data. However, when student data grows, it becomes more difficult to analyze it using database management software and conventional statistical methods . Universities must therefore utilize a technology to retrieve the pertinent data. The performance of the kids could be predicted using this important information.

Although the Intelligent Mining and Decision Support System (InMinds) at Universiti Malaysia Sarawak (UNIMAS) can currently view student performance, due to its confidential setting, it is only available to top management, such as the Deans and Deputy Deans of Undergraduate and Student Development. The teachers, who are not a member of the senior

management, are not allowed to see the performance of the pupils. Without the use of an automated system, professors currently manually search student files and records for data. This makes it difficult for each professor to access their students' data during the course of the semesters.

The suggested system for performance analysis enables instructors to recover students' prior course performance and gain a better knowledge of the factors that affect students' success. In addition, instructors can forecast pupils' performance. This enables the faculty to set higher success rate goals in the future.

II. RELATED WORKS

The performance of pupils using machine learning has been the subject of numerous research in the literature. The majority of this research focuses on identifying the most suitable algorithms for making predictions as well as the features that will be employed in forecasting. There are known scientific publications whose goal is to create an extensive overview of the literature related to the foretelling of academic results.

Various methods and various kinds of knowledge are studied in several researches. Based on the data gathered during this review, supervised learning was found to be the most popular strategy for anticipating students' behavior since it produces precise and trustworthy outcomes. In particular, the authors used the SVM algorithm the most since it produced the most precise predictions. They made predictions using academic data (for example, forecasting course performance aided high-school performance). Even if it is becoming more common, the use of data that describes a student's behavior in a particularly challenging class (log data) is nevertheless still rather uncommon. The study of forecasting student performance is underdeveloped at Bulgarian universities. There aren't many studies of Bulgarian writers related to this problem.

SVM performs best, followed by k-NN, according to a comparison of the classifiers Naive Bayes, Decision Tree, SVM, and k-NN. The kind of information utilized for predicting affects how well the models work. This explains why different classifiers provide the most suitable results in various contexts.

III. PROPOSED WORKS

During the design and implementation stages of the proposed system, a few aspects from the current systems are used. The user interface, student performance prediction, illustration displays, and report generating are some of these features and functionalities. A excellent user interface offers a user-friendly interface since it is simple to use and easy to browse. To ensure that the goals are met, the suggested method incorporates the students' performance forecast. Additionally, it is simpler to analyze student performance thanks to the creation of reports in Portable Document Format (PDF) and visual displays such charts in PDF.

All of the user criteria would be satisfied by these characteristics included in the proposed system.

Following are the user needs that were gathered from professors throughout the system analysis phase:

Capable of autonomously forecasting students' performance for professors.

Able to record and retrieve information about students' performance in specific courses and semesters.

IV. CONTRIBUTION

Students Performance analysis system is developed for predicting the student results based on the previous results stored on student dataset. The main concern of the teachers regarding the performance issue of the students is covered by introducing the analysis tool R. R is a statistical analysis tool that takes a data set as an input and builds a model that is used for predicting the future data set.

V. METHODOLOGY

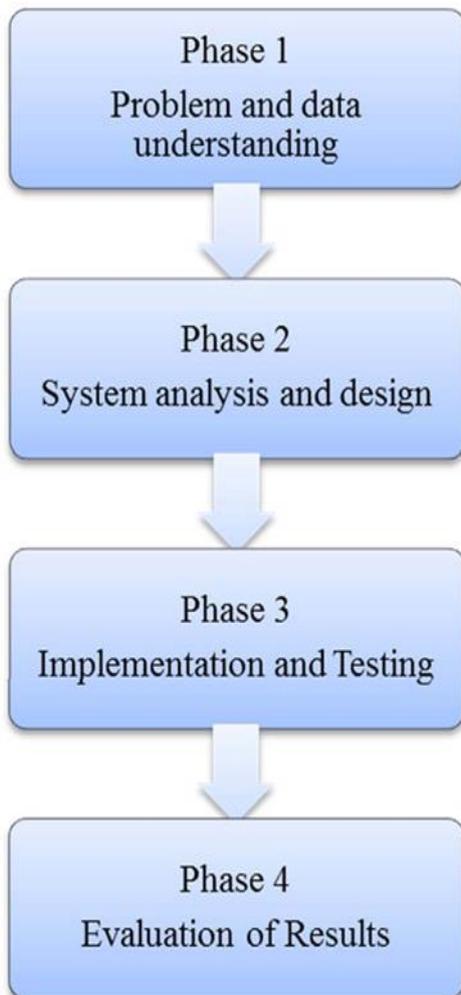


Fig 1:Methodology

Problem and data understanding

Understanding the issue and the data is essential to determine whether the Student Performance Analysis system will be successful. Problems and data understanding are established prior to system development in order to specify the project's goal and objectives. The shortcomings of the current systems are noted and examined for their functionality and efficacy. Following the identification of the issues, each problem's remedies are then located and gathered by additional reading and research on the pertinent research articles.

System analysis and design

The system's overall flow is planned, examined, and designed during this phase. Analyzed and listed in table style are the system and user requirements. The input, operations, and output of the system are represented on a data flow diagram. The context diagram up to the first level's data flow diagram is analyzed and created. Additionally, a logical design of the suggested system is created to guarantee that the finished system performs as planned.

Implementation and Testing

In order to develop IF-THEN rules for the prediction of students' performance, a dataset of student records is gathered and analyzed using data mining techniques during the implementation phase. WEKA is a free software program that is used to generate IF-THEN rules. Training set and test set have been created from the dataset. The training set uses 80% of the dataset, and the test set uses the remaining 20%. The test set is used to evaluate the classification model's accuracy in making predictions whereas the training set is used to train the classification model.

Evaluation of Results

Five end users are asked to rate the Student Analysis Performance system's usability as part of the system evaluation. This is done to guarantee both the accomplishment of the proposed system's goals and the convenience of navigating its interfaces. Additionally, the evaluation is carried out to guarantee that the high efficacy of the suggested system is realized.

VI. ALGORITHM

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.

K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.

K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K-NN algorithm.

K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.

K-NN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data.

It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.

As we can see the 3 nearest neighbors are from category A, hence the new data point belongs to category A

VII. ARCHITECTURE DIAGRAM

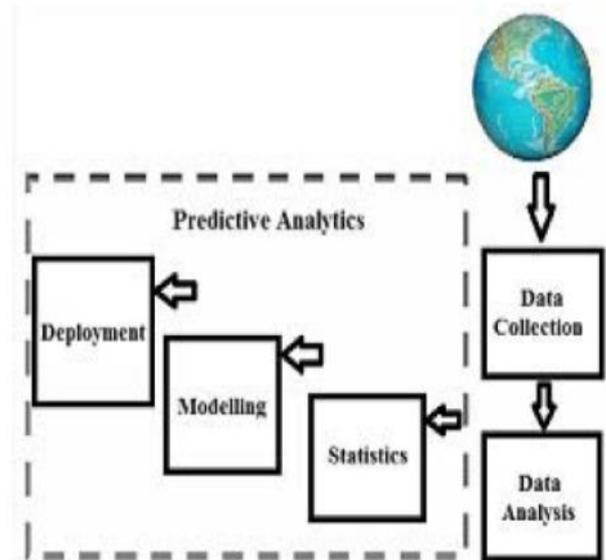


Fig 3: System Architecture



Fig 2: KNN Working Model

VIII. EXPERIMENT AND RESULTS

This paper covers all the objectives discussed above and full all the constraints. In this paper, a sample dataset has taken from the records of College. This dataset has factors (student id, Name, department, semester, result) have been chosen that affects a student’s performance in different semesters. Further a statistical analysis tool R has been chosen, for building a model data set. This data set is further divided into two parts (Test Data-Set, Training Data-Set). Following schematic table represents the sample test data set and training data set respectively. Training and test dataset is divided in the ratio of (75

: 25) respectively. Finally we can shows the result of Naive Bayes approach. It concludes that, error ration is (20 : 5). So the performance of the model is 75%:

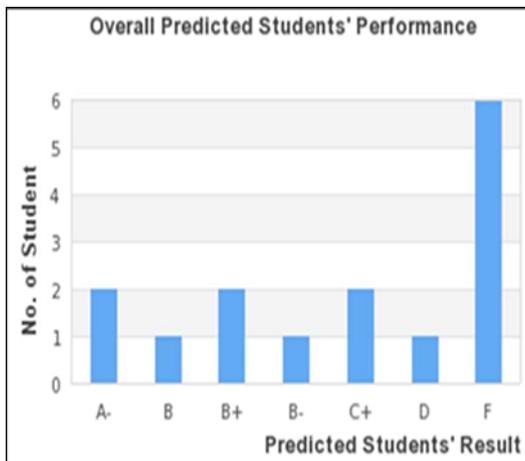


Fig 4:Overall Predicted Students' Performance

IX. CONCLUSION

This paper covers all the objectives discussed above and full all the constraints. The main concern of the teachers regarding the performance issue of the students is covered by introducing the analysis tool R. R is a statistical analysis tool that takes a data set as an input and builds a model that is used for predicting the future data set. This analysis tool R uses machine learning approaches like Naive Bayes, K-nearest neighbour first for building a model. the project concentrates on the development of a system for student performance analysis. A data mining technique, classification algorithm is applied in this project to ensure the prediction of the student performance is possible.Introduction section covers the basic review part, that discusses the tools and approaches that have been designed in this area. Section methodology describes the machine learning tool and the approach for choosing a data set. Section experimental design and result shows the schematic diagrams of tables that are used for performance analysis. Finally, this paper suggests that one should use the cognitive modelling for designing the knowledge-base. So that they can produce better

results.

REFERENCES

- [1] J. Shana, and T. Venkatacalam, "A framework for dynamic Faculty Support System to analyse scourse data", International Journal of Emerging Technology and Advanced Engineering, Vol. 2, No.7,2012, pp.478-482
- [2] Quality Assurance Division (2010). InMinds: Intelligent Mining and Decision Support System. *University of Malaysia Sarawak, UNIMAS* [Online]. Available
- [3] SPA (2013). What is SPA Standard? *SPA Student Performance Analyser* [Online]. Available:
- [4] V. Kumar, and A. Chadha, "An empirical study of the applications of data mining techniques in higher education", International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 2, No..3, 2011, pp. 80-84.
- [5] Ganeshan, Kathiravelu and Li, Xiaosong. "An intelligent student advising system using collaborative filtering", 201" IEEE Frontiers in Education Conference (FIE), pp. 1-8, Oct 2015
- [6] Poza-Lujan, Jose-Luis and Calafate, Carlos T. and Posadas- Yague. "Assessing the Impact of Continuous Evaluation Strategies: Tradeoff Between Student Performance and Instructor Effort", IEEE Transactions on Edu- cation, vol.59, pp.17-23, Feb 2016.
- [7] Elbadrawy, Asmaa and Polyzou, Agoritsa and Ren, Zhiyun and Sweeney. "Predicting Student Performance Using Personalized Analytics", IEEE, vol. 49,pp. 61-69, Apr.2016.