

Analyzing the Student Performance, Forecasting and Generating a fully detailed report Based on Machine Learning.

Nethsara Liyanage
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
tharusharb1012@gmail.com

Abstract—E-learning has shown rapid growth in the last decade, specially in the last pandemic. With the competition, it has focused on teaching and measuring student performance in different and more efficient ways.
... to be completed

I. INTRODUCTION

With the impact of Sars-Cov-2, over 160 countries enacted nationwide shutdown, affecting over 87 percent of the world's student population and several other countries have implemented limited school closures, according to UNESCO's monitoring. In this scenario, a considerable number of schools, universities and other higher educational institutes have transformed to the online education techniques. For the online deliveries, such institutions have motivated to apply software solutions like Learning Management Systems (LMS) in their courses.

When the deliveries of the course modules conducted through the online platforms rather than delivering them in the physical manner, the institutions and the lecturers/teachers would like to have a personal analyze for each student who are they teaching. In physical lectures or the teachers can go to each student and get a rough idea about the student. But in the online lectures they are unable to continue the same procedure. Most educators were unable to assess their students' progress due to their incapacity to evaluate and monitor their actions in a virtual environment. In addition, students are also facing some issues such as lack of focus. Considering the problem that how to obtain more productivity by analyzing and forecasting the progress of the students?, first there should be a method to identified that the productivity levels of the students in both individual and overall group manner. That would help the tutors as well as lecturers to reach out each student and help them to improve their productivity in each subject module. For achieve this scenario, the proposed solution of this paper comes up with overall and individual forecasting result dashboards. When having a forecasting dashboard, not only the teachers or lecturers, but also the students/undergraduates will be highly benefited. They could refer the forecasting dashboard and level up their

abilities and the skills. And they could recognize the level that they are in the batch or group.

II. LITERATURE REVIEW

In the formation of a more efficient generation of corporate personnel, career orientation is a critical issue that should not be underestimated. Thus, we are supposed to analyze the performance of a student and conduct a time series forecasting for the student. Hence, in the consideration of performance of a student, we must rely on every possible data of a student. On the other hand, for analyzing and the determination of the prediction, we must exercise more powerful and more accurate techniques and algorithms. Thus, in this literature review we will review some already existing performing analyzing approaches and algorithms and the existing approaches that can be found to execute the career recommendation. Furthermore, for generate the report of the overall learner classification of the system was supposed to utilize third-party services for data visualizations. Thus, in this literature review the possible third-party applications will also be discussed.

In the consideration of the performance of the students, since we have the gathered data as the inputs of this system, we have the advantage of using data mining techniques as well as the supervised and unsupervised learning techniques. Using data mining techniques, we could gain the advantage of determining the relevant information from the data set. Moreover, these techniques obtain the knowledge from the given data. When considering the supervised and unsupervised learning techniques and from their subcategory's classification, prediction, association, clustering we have the ability of applying the classification technique for analyzing the performance of the student. Furthermore, we can apply the data mining technique for the conduct the student performance.

According to the Lim Pek Choo and Jane Labadin in their study on student performance and forecasting, accuracy of the projected baseline grading is considerably higher when considering accuracy of the both standard grading and the projected baseline grading.

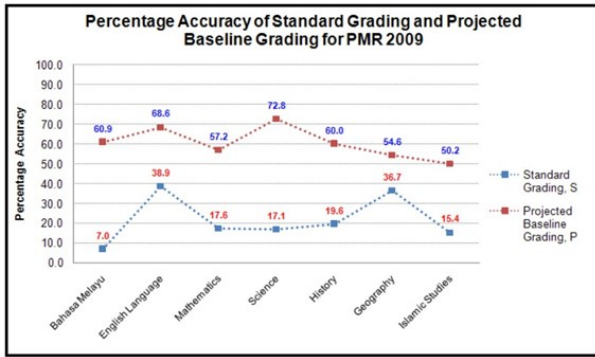


Figure 1 - Deviance of S and P

Student achievement in an online course is linked to their previous session performance as well as their level of interest. Literature has paid little attention to determining whether student performance and participation in previous tests may influence student accomplishment in subsequent examinations. So that, considering the above scenario, the requirement for the relevant solution is more confirmed. Researchers have done many case studies on the mentioned main objective performance progress analysis and forecasting as well as on career recommendation. Considering the conclusions and the methodologies of the [5], [8] and [9], they have followed some basic methods such as,

- Student's exam marks analysis
- Graph theory to check relationship between student and subject

Moreover, considering the existing real-world applications such as Blackboard, Jotform are also have some features like result analysis as well as progress analysis. But they are more focusing on the content delivering to the students. The following table contains the comparison between those existing application with the proposed solution.

Platform / Tool	Time-based learner analyzer	
	Result Analysis	Progress Analysis
Blackboard	Available	Not Available
Jotform	Not Available	Available
AI-Tor	Available	Available

III. METHODOLOGY

There are several phases of methodology used throughout the project development, which is as follows,

- Problem understanding and data understanding
- System analysis and design
- Implementation and testing
- Reporting the results

A. Problem understanding and data understanding

Understanding the problem 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.

Student information is gathered in this step aside from that. Data about the students, including their semester-to-date performance on the pertinent topic module, is gathered. The characteristics of the data set gathered for data mining categorization are shown in the below table.

Attributes	Value
Quiz for week 01	Discrete
Quiz for week 02	Discrete
Quiz for week 03	Discrete
Quiz for week 04	Discrete
Quiz for week 05	Discrete
Quiz for week 06	Discrete
Quiz for week 07	Discrete
Mid Term Examination	Discrete, Grade
Quiz for week 09	Discrete
Quiz for week 10	Discrete
Quiz for week 11	Discrete
Quiz for week 12	Discrete
Quiz for week 13	Discrete
Quiz for week 14	Discrete

Other than these data inputs, the identified skills of the students are taken as another input for the proposed system.

B. 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.

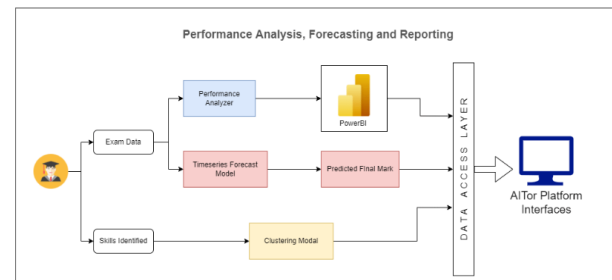


Figure 2 - System Design

C. Implementation and testing

In order to produce IF-THEN rules for the prediction of students' results in the course "Software Process Modeling" a dataset of almost 500 student records from the course is

gathered and evaluated throughout the implementation phase using data mining techniques. 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 requires 80 percent of the dataset, while the test set requires the remaining 20 percent. 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. To ensure that the highest forecast of accuracy could be reached, various decision tree classification techniques are compared for accuracy. Table 2 compares the effectiveness of five distinct decision tree categorization methods that can be found in WEKA.

Technique used	Correctly Classified Instances
J48	58.3%
Simple CART	59.1%
BFTree	61.4%
Random Tree	45.7%
J48graft	58.3%

Figure 3 - Comparison between classification techniques

The BFTree is the decision tree classification technique used in the proposed system since it has the best accuracy (61.4 percent) of the decision tree techniques given in the table above. The WEKA best-first decision tree serves as the basis for the IF-THEN rules. The proposed system's PHP language IF-ELSE condition will incorporate these rules. Before the final test, the rules are used to forecast the students' grades in the "Software Process Modeling" course. The forecast helps the instructors identify students who are likely to fail the "Software Process Modeling" course.

The other main functional component that conducting the time series forecasting for the progress of the student will be more extracted in this part of the paper. As per discussed before in the background and literature review, it could identify that most of the case studies were considered only some of the features that we discussed on the research gap section. Most of the cases, for the time-series forecasting they did consider only in the result analysis. But to get more accurate results regarding the student's performance, the progress analysis should also be considered in the forecasting.

In the proposed application, both the result and the progress analysis will be considered. There would be three main phases in the proposed solution to continue the process of conducting the time-series forecasting for the progress of the student, presenting the detailed report.

Before the system is deployed and utilized by IS professors, it must first undergo unit testing, system testing, and user acceptance testing to identify any flaws. This is done to make sure the system is operating at its best. Additionally, the flaws and defects found in the suggested system during testing can be rectified. To make sure the system's functionalities are operating as planned, the developer will test the system and unit code, while a small number of end users will test the user acceptance code.

IV. CONCLUSION

The conclusion goes here.

REFERENCES

- [1] H. Kopka and P. W. Daly, *A Guide to L^AT_EX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.
- [2] H. Kopka and P. W. Daly, *A Guide to L^AT_EX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.