Analyzing Student Performance in Programming Education Using Classification Techniques

Team Members: CSE A Batch-2

Sneha Teja Sree Reddy Thondapu -AP19110010124
Nandini Thimmireddy Gari -AP19110010128
Likhitha Tadikonda -AP19110010006
Aman Gupta -AP19110010125

I. Abstract

Programming Skills are very important for any computer engineering student to get good marks in exams, apply the concepts to solve any real-world problem, crack any job interview, etc. But the only way they can know about their performances is to analyze and improve their skills regularly by seeing the statistics of their results in that semester regularly. With the help of this project, one can analyze their scores regularly, introspect, and can deliberately practice for better scores. This reduces the students' stress, anxiety, and depression about getting good scores in their academics. This analysis helps even professors to improvise the learning outcomes of students and increase their performance in whatever field they are working in.

Engineering has a vast division of marks in each subject like the internals, externals, and lab components. Every student has his/her comfort zone in each of these three components. But when they step out of their comfort zone there comes the problem. In this research, we aggregated the department of computer science students' data from the Web Technology subject (CSE202) which we took in the 4th Semester from our university SRM AP, Amaravati. We implemented classification algorithms like KNN, Decision tree, Logistic regression, Naive Bayes, and Adaboost algorithms to analyze the data of the students. We compared all the ML algorithms based on 200 classification instances. This analysis helps us understand how many students are performing well and how many could not perform well in all these internal, external, and lab exams and also helps students find their weak areas that have to be focussed on to improve their performance before they take up next tests.

II. Introduction

Data analysis is very important for students, teachers, educationalists, etc to understand, evaluate, introspect and discover useful information. With Data analysis we can solve many problems- Once said by Einstein that if you could find the problem or understand the question, you have the 90% of the solution, and the same applies to data analysis also. When we can sort the data out according to the problem in an understandable way, then we can fix any problem or get any answer related to the data objects. Since computer science subjects give more weightage

to the students' problem-solving skills, the evaluation and division of the subject into sub-components is also done by the faculty such as lab performance, internal, etc.

This data analysis classifies all the huge data and gathers them into groups to make the evaluation simpler using Machine Learning Algorithms, Graphs, Classifiers, Outliers, etc. Since covid-19 has come, there is a necessity to make smart learning or online learning more efficient. Over the last decade, many programmers over the country have used these algorithms to analyze the data.

III. Objectives of this research

- Take a large dataset of student's subjects and analyze that data using classification and ML algorithms
- Choose the best algorithm among all the ML algorithms used above so that we can recommend the best to the students as well as evaluators for getting maximum accuracy scores.

IV. Literature Work

Data mining in higher education is a recent research field and this area of research is gaining popularity because of its potential for educational institutes. Data Mining can be used in the educational field to enhance our understanding of the learning process to focus on identifying, extracting, and evaluating variables related to the learning process of students as described by Alaa el-Halees.

Pandey and Pal conducted a study on student performance by selecting 600 students from different colleges of Dr. R. M. L. Awadh University, Faizabad, India. By means of Bayes Classification on category, language, and background qualification, it was found whether newcomer students will perform or not.

Hijazi and Naqvi conducted a study on student performance by selecting a sample of 300 students (225 males, 75 females) from a group of colleges affiliated with the Punjab university of Pakistan. The hypothesis that was stated "Student's attitude towards attendance in class, hours spent in a study on daily basis after college, students' family income, students' mother's age, and mother's education are significantly related with student performance" was framed. By means of simple linear regression analysis, it was found that the factors like mother’s education and the student’s family income were highly correlated with the student's academic performance.

Khan [7] conducted a performance study on 400 students comprising 200 boys and 200 girls selected from the senior secondary school of Aligarh Muslim University, Aligarh, India with the main objective to establish the prognostic value of different measures of cognition, personality, and demographic variables for success at higher secondary level in science stream. The selection

was based on a cluster sampling technique in which the entire population of interest was divided into groups or clusters, and a random sample of these clusters was selected for further analyses. It was found that girls with high socioeconomic status had relatively higher academic achievement in the science stream and boys with low socioeconomic status had relatively higher academic achievement in general.

Al-Radaideh, et al [9] applied a decision tree model to predict the final grade of students who studied the C++ course at Yarmouk University, Jordan in the year 2005. Three different classification methods namely ID3, C4.5, and the NaïveBayes were used. The outcome of their results indicated that the Decision Tree model had better prediction than other models.

order to uncover buried knowledge, several data mining approaches were used. We utilized the decision tree approach, KNN, Logistic Regression, Naive Bayes, and Adaboost classifier to evaluate the data set and predict the student success or failure rate. Both professors and students will benefit from this research.

V.REFERENCES

1. **Alaa el-Halees**: https://arxiv.org/ftp/arxiv/papers/1201/1201.3417.pdf
2. **Pandey & Pal:**https://www.researchgate.net/publication/51989865_Mining_Educational_Data_to_Analyze_Students'_Performance
3. **Hijazi and Naqvi:**https://www.ijert.org/research/mining-parent-socio-economic-factors-to-predict-students-academic-performance-in-osun-state-college-of-technology-esa-oke-IJERTV2IS120345.pdf
4. **Khan:**https://www.ijert.org/mining-parent-socio-economic-factors-to-predict-students-academic-performance-in-osun-state-college-of-technology-esa-oke-2
5. **Al-Radaideh:**https://repensarlafisica.files.wordpress.com/2015/10/mining-student-data-using-decision-trees.pdf

Thank-you