International Research Journal of Computer Science (IRJCS) ISSN: 2393-9842
Issue 06, Volume 6 (June 2019) www.irjcs.com
SPECIAL ISSUE - 5th International Conference - "ACCE – 2019"

CLUSTERING TECHNIQUE FOR STUDENT PERFORMANCE

Mrs .Mary vidya john [1], Akshata police patil[2], Anjali mishra[3], Bindhu reddy G[4], Jamuna N[5]
[1.]Assistant Professor, Department of CSE, Vemana Institute of Technology, Bangalore.
[2, 3, 4, 5.]UG student, Department of CSE, Vemana Institute of Technology, Bangalore.
[2.] policepatilakshata@gmail.com, [3]anjalimishra88@gmail.com,
[4.]reddygbindu@gmail.com, [5.]jamunajamu500@gmail.com.

Manuscript History
Number: IRJCS/RS/Vol.06/Issue06/JNCS10120
Received: 29, May 2019
Final Correction: 30, May 2019
Final Accepted: 02, June 2019
Published: June 2019
doi://10.26562/IRJCS.2019.JNCS10120
Editor: Dr.A.Arul L.S, Chief Editor, IRJCS, AM Publications, India

ABSTRACT - To enhance the quality of education system, student performance analysis plays an important role for decision support. Evaluation of student's performance is an important aspect in every institution. The student's knowledge about the programming concepts can be tested using various forms of assessments. It helps the teacher to identify students who need special attention and allow the teacher to provide appropriate guidance. To achieve this a machine learning approach called the Recursive Clustering technique is used to group the students of the programming course into groups based on their performance in the given assessment. There are three groups: Pass, Fail and Above Average. More attention is given to the students present in lower group as they are prone to fail. Study materials are provided to the students for preparation. Tests are conducted for the students who fail. The process is repeated for three times so that most of the students from the lower group will move to the higher group. The results are compared using k-means algorithm. The results prove that this approach provides an effective way to predict the low performing students from their assessment, thereby enabling the student to be on track.
Keyword: Machine learning, Recursive clustering, K-means algorithm.

I. INTRODUCTION

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. There are three ways for learning: supervised, unsupervised and reinforcement. Today, many institutions are widely using the machine learning techniques for the analysis of student's performance. Ultimate objective is to extract information from large datasets, and to utilize the extracted information in decision making process. Programming skill is a very important aspect for a computer science student. It is important for the students to develop their abilities in order to acquire the skill. Programming language is important because it defines the relationship, semantics and grammar which allow the programmers to effectively communicate with the machines that they program. This paper aims to evaluate the performance of the students by making them into groups (Clusters) according to their performance in the assessment. The clusters are created using Machine Learning Techniques called Recursive Clustering. The best performing machine learning algorithm known as k-means algorithm is used here which provides an easy understanding and implementation. *K*-means clustering is a type of unsupervised learning, which is used in the case of unlabelled data.

The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable *K*. The algorithm works iteratively to assign each data point to one of *K* groups based on the features that are provided. Data points are clustered based on feature similarity.

II. LITERATURE SURVEY

Faculty are analyzing difficulties in programing languages that students don't have knowledge about how program works in various aspects related to storage and execution[3].The number of dropouts are increasing in programming in fact programming courses are difficult to master, at least each person takes more than 10 years to expert in programming languages[2]. Robin at [4] teachers must focus on the application while teaching the programming concepts. Students lack of knowledge in initial stages of the programming courses this leads to poor performance [5]. Mc Carcken et al. [6] noticed that students have serious deficiencies in programming skills at introductory level courses. Traditional teaching method requires more memorization than much needed more critical thinking[15].Critical Thinking can be largely enhanced in peer groups. Therefore, it is observed that the students' performance improves in properly structured cooperative learning environment [16]. A study by Nokes et al. [17] investigating students' learning observed that, the students who practiced traditional ways of learning scored lower than the students who resorted to heuristic techniques of learning in content-based assessments. Students' performance is enhanced when they use blended learning techniques [18].

III. TECHNIQUES AND TOOL

Machine learning (ML) is the technique and study of algorithms and statistical models that computer systems use in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are used in a wide variety of applications, such as email filtering, and computer vision, where it is infeasible to develop an algorithm of specific instructions for performing the task. Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a field of study within machine learning, and focuses on exploratory data analysis through unsupervised learning. In its application across business problems, machine learning is also referred to as predictive analytics.

Recursive Clustering Technique

It is basically a type of *unsupervised learning method* . An unsupervised learning method is a method in which we draw references from datasets consisting of input data without labeled responses. Generally, it is used as a process to find meaningful structure, explanatory underlying processes, generative features, and groupings inherent. Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them. It is the simplest unsupervised learning algorithm that solves clustering problem. K-means algorithm partition n observations into k clusters where each observation belongs to the cluster with the nearest mean serving as a prototype of the cluster **.**

K Means Algorithm

Let X = {x1,x2,x3,……..,xn} be the set of data points and V = {v1,v2,……..,vc} be the
set of centers.
1) Randomly select 'c' cluster centers.
2) Calculate the distance between each data point and cluster centers.
3) Assign the data point to the cluster center whose distance from the cluster center is
minimum of all the cluster centers..
4) Recalculate the new cluster center using:
where, 'ci' represents the number of data points in ith cluster.

III. DESIGN

The architectural configuration procedure is concerned with building up a fundamental basic system for a framework. It includes recognizing the real parts of the framework and interchanges between these segments. The beginning configuration procedure of recognizing these subsystems and building up a structure for subsystem control and correspondence is called construction modeling outline and the yield of this outline procedure is a portrayal of the product structural planning.

The figure .1 shows the system architecture . It shows the way this system is designed and brief working of the system. The DFD is straightforward graphical formalism that can be utilized to speak to a framework as far as the info information to the framework, different preparing did on this information and the yield information created by the framework. A DFD model uses an exceptionally predetermined number of primitive images to speak to the capacities performed by a framework and the information stream among the capacities.

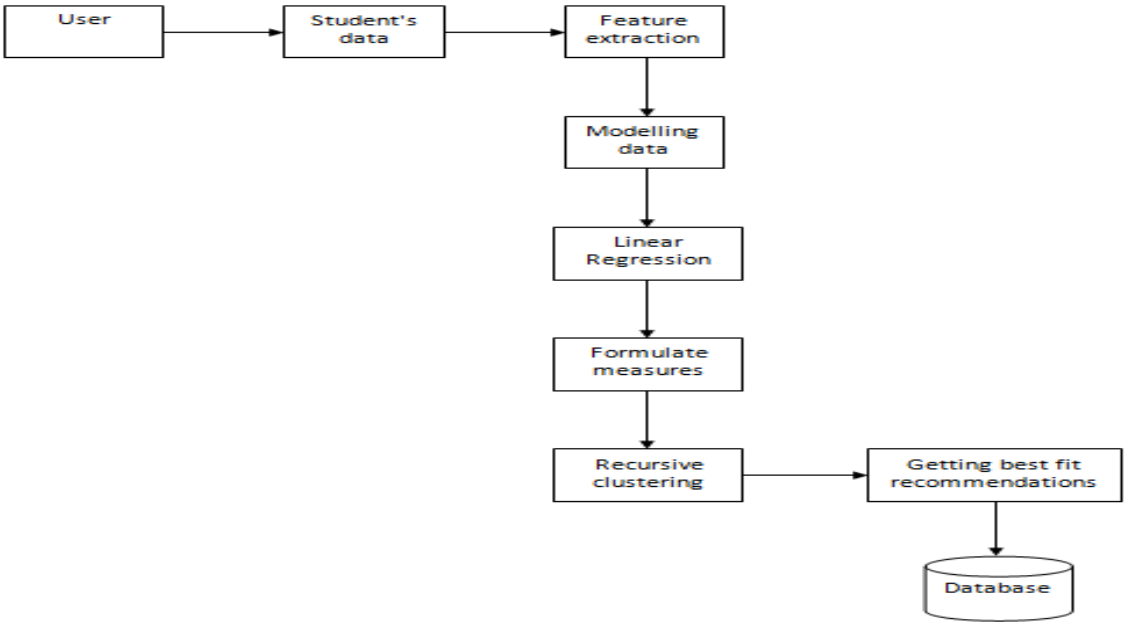

Figure 1. system architecture

The DFD is straightforward graphical formalism that can be utilized to speak to a framework as far as the info information to the framework, different preparing did on this information and the yield information created by the framework. A DFD model uses an exceptionally predetermined number of primitive images to speak to the capacities performed by a framework and the information stream among the capacities. The principle motivation behind why the DFD method is so famous is most likely in light of the way that DFD is an exceptionally basic formalism- It is easy to comprehend and utilization. Beginning with the arrangement of abnormal state works that a framework performs, a DFD display progressively speaks to different sub capacities. Actually, any various leveled model is easy to get it.

IV. IMPLEMENTATION

The implementation phase involves the actual materialization of the ideas, which are expressed in the analysis document and developed in the design phase. Implementation should be perfect mapping of the design document in a suitable programming language in order to achieve the necessary final product. Often the product is ruined due to incorrect programming language chosen for implementation or unsuitable method of programming. It is better for the coding phase to be directly linked to the design phase in the sense if the design is in terms of object oriented terms then implementation should be preferably carried out in a object oriented way. The figure 2shown below depicts a program that's running on the Java platform. As the figure shows, the Java API and the virtual machine insulate the program from the hardware.

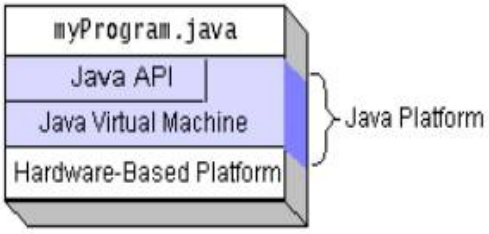

Figure 2 : Java Interpreter Architecture.

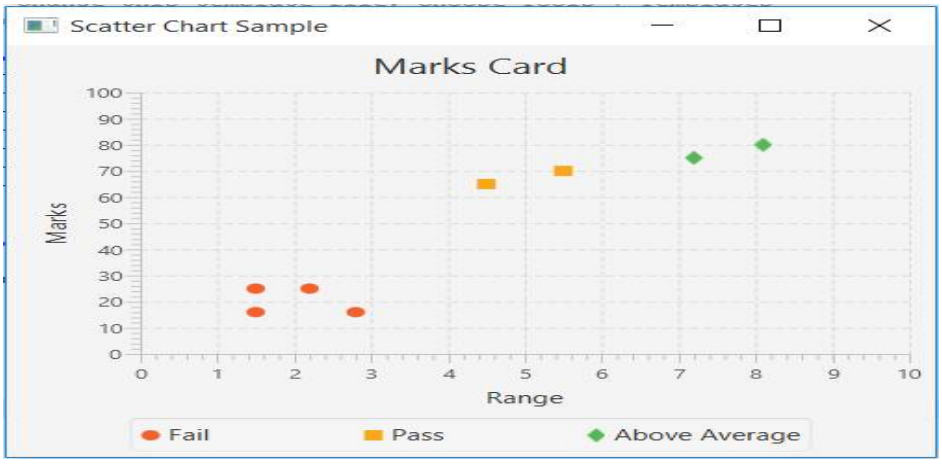

Figure 3: Student clusters c1 based on internal 1

Table 1: Clusters C1 details after internal 1

Range	Number of students	percentage
Fail	4/8	50%
Average	2/8	25%
Above average	2/8	25%

Figure 4: Student cluster C2 based on internal 2 assessment

By analyzing figure 3 and 4 it shows the range and marks of the students in the form of graph. The above figures clusters the results based on their internal performance.
Total students in cluster C1= 8
Total students in cluster C2= 4

Table 2: Cluster C2 details after internal 2

Range	Number of students	Percentage
Fail	2/4	50%
Average	1/4	25%
Above average	1/4	25%

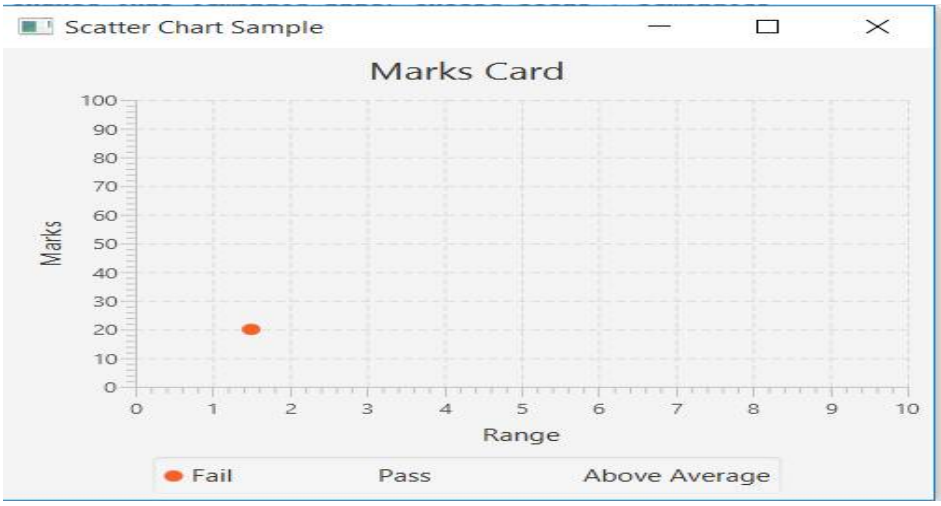

Figure 5: Student cluster C3 based on internal 3 assessment

Total students in cluster C3= 1

Table 3: Cluster C3 details after internal 3

Range	Number of students	Percentage
Fail	1/1	100%
Average	0	0%
Above average	0	0%

By observing figure 4 and 5 it was showed that 50% of the students who are initially in cluster C2 were moved to cluster C3,the number of failure students are more compare to cluster C2.

V. CONCLUSION

Student performance in programming courses will be improved by conducting the assessments and if the students are analysed in earlier stage they can provide a set of programs and notes automatically to improve their performance in exams. This work is done by the clustering technique to group the students into different groups based on their previous performance.

REFERENCES

1. Education in Oman The Drive for Quality, Jointly prepared by The Ministry of Education and The WorldBank, Ministry of Education Publication No. 23/2012
2. E. Soloway and J. Spohrer, Studying the Novice Programmer,Lawrence Erlbaum Associates, Hillsdale, New Jersey. 1989, pp. 497
3. I. Milne and G. Rowe, Difficulties in Learning and teaching Programming - Views of Students and Tutors, Education and Information Technologies, 2002, pp. 55–66.
4. A. Robins, J. Rountree, and N. Rountree. Learning and teaching programming: A review and discussion. Computer Science Education, 13(2):137–172, 2003.
5. M. Kölling and J. Rosenberg, "Blue - A Language for Teaching Object-Oriented Programming", Proc. of the 27th SIGCSE Technical Symposium on Computer Science Education, 1996, pp. 190-194.
6. M. McCracken, V. Almstrum., D. Diaz, M. Guzdial, D. Hagan, Y Kolikant, C. Laxer, L. Thomas, I. Utting and T. Wilusz, A multi-national, multi- institutional study of assessment of programming skills of first-year CS students, SIGCSE Bulletin, 33(4), 2001, pp. 125-180.
7. https://docs.microsoft.com/en-us/azure/machine-learning/machine-learning-algorithm-choice visited on 13/09/17.
8. J. MacQueen, "Some methods for classification and analysis of multivariate observations", Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, 281--297, University of California Press, Berkeley, Calif.1967
9. G. Hamerly, C. Elkan, "Alternatives to the k-means algorithm that find better clusterings" . Proceedings of the eleventh international conference on Information and knowledge management (CIKM), 2002
10. S. Vassilvitskii. "k-means++: the advantages of careful seeding", Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics Philadelphia, PA,USA. 2007, pp. 1027–1035.

