Global Journal of Computer and Engineering Technology (GJCET)

Vol. 3, No. 2, July, 2023 ISSN: 2767-1933

EVALUATION OF ACADEMIC PERFORMANCE OF STUDENTS USING K-MEANS CLUSTERING ALGORITHM

S. P. Siddique Ibrahim, VIT-AP University, India(siddique4x@gmail.com)
Addanki Dedeepya, VIT-AP University, India(dedeepya.21bce9017@vitapstudent.ac.in)

ABSTRACT

In the field of analysis of data and machine learning data clustering is a fundamental technique that entails assembling comparable data points according to their inherent characteristics. To facilitate more effective data exploration, organization, and decision-making, data clustering aims to reveal hidden patterns, structure and correlations within datasets. In this study, the k-means clustering algorithm is used to evaluate the academic achievement of students. Faculty can assess how well a student will perform in the upcoming semesters by evaluating the student's academic performance based on quiz, assignment, midterm, attendance, and lab performance.

Keywords: Academic performance, K-Means Clustering, Education data analytics, Machine learning, Student performance prediction.

1. INTRODUCTION

In the context of education, clustering is the process of organizing learners, courses, or educational materials according to specific traits, qualities, or learning styles. It is important for education since it facilitates individualized Learning, which implies that clustering can assist teachers in customizing instruction to meet the requirements of each student. In order to maximize student engagement and comprehension, staff might tailor instruction and provide focused support by grouping students with comparable learning styles or skills together (siddique et al., 2019). In the educational context, clustering plays a vital role in structuring learners, courses, or educational materials based on particular traits, qualities, or learning styles. This process is particularly significant in education because it enables individualized learning, a pedagogical approach that tailors instruction to meet the unique needs and preferences of each student. The essence of clustering in education lies in its ability to enhance the personalization of the learning experience.

The application of clustering in education is instrumental for teachers seeking to customize their instructional methods. By categorizing students with similar learning styles or skill sets into groups, educators can create more targeted and effective teaching strategies. This approach allows teachers to adapt their methods to the specific characteristics of each cluster, maximizing student engagement and comprehension(Christy et al., 2019).

For instance, if a group of students exhibits a preference for visual learning, educators can incorporate more visual aids, diagrams, or videos into their lessons for that particular cluster. On the other hand, a cluster with a preference for hands-on activities might benefit from more interactive and practical learning experiences. By recognizing and leveraging these diverse learning styles within clusters, teachers can provide a more tailored and engaging educational environment(SPS et al., 2023).

Furthermore, clustering in education extends beyond just students. Courses and educational materials can also be organized into clusters based on content, difficulty level, or teaching approach. This ensures that the learning resources align with the specific needs and proficiency levels of different student groups(Ibrahim et al., 2020).
In the realm of data analysis and machine learning, data clustering stands out as a pivotal technique, integral for organizing and understanding large datasets. This method involves the grouping of similar data points based on their inherent characteristics. The primary objective of data clustering is to unveil concealed patterns, structures, and correlations within datasets, fostering more efficient data exploration, organization, and decision-making processes (Selva et al., 2023).

In the specific context of academic assessment, this study employs the k-means clustering algorithm to evaluate students' academic achievements. The k-means algorithm is a widely utilized clustering technique that partitions data points into k clusters, with each cluster having a centroid representing its mean. This approach is particularly valuable in assessing and predicting academic performance.

By leveraging various academic metrics such as quiz scores, assignment completion, midterm exam results, attendance records, and lab performance, faculty members can gain comprehensive insights into a student's overall academic prowess. The k-means clustering algorithm allows for the segmentation of students into distinct groups based on these diverse factors, enabling educators to discern patterns and trends in academic achievement.

The ultimate goal of applying the k-means clustering algorithm to academic data is to provide educators with a predictive tool. By analyzing historical data and clustering students with similar performance attributes, faculty members can make informed predictions about how well a student is likely to perform in upcoming semesters. This predictive capability can be invaluable for designing personalized interventions and support systems tailored to individual students' needs, ultimately enhancing the overall educational experience and outcomes.

2. LITERATURE REVIEW

A paper used clustering and decision tree for the evaluation of student's academic performance.They considered some of the student's attributes and applied k-means clustering algorithm on it(Ibrahim et al., 2019) . After clustering the students are grouped into High, Medium and Low. Later a decision tree is used to make decision based on student categories. The data used in that project is a sample dataset.

Another paper focused on investigating the development of an expert system using fuzzy logic and a fuzzy C-means. They developed an innovative technique for student allocation based on Bayesian approach.

Another paper explains the difference in outcomes when assessing results derived from a fuzzy expert system, a clear distinction emerges between traditional and new ways. While the traditional approach sticks on to mathematical rules, the actual process of evaluation using fuzzy logic is more flexible. The suggested fuzzy C-means based Fuzzy Expert System FCM seamlessly transformed the precise data into fuzzy sets and simultaneously computed.(S Devi et al., 2021)

Another paper performed association rule mining for evaluating the performance.

Another paper used different algorithms like FCM, K-Means clustering, Classification, Decision Tree and association Rule Mining while determining the improvement in performance of students and compared efficiency of those algorithms(Priyanka et al., 2015).

In literature review different sources related to the topic should be presented. The literature should be related with either domain or method/technique/algorithm used in the correspondence research. The previous research (Loi et al., 2019)works should be described in the form of title, problem statement, objectives. Shatty et al. (2015) like all the cited literature must be written in references(Ibrahim et al., 2023).

3. EXPERIENTIAL WORK

3.1 Clustering of Data

A cluster is a grouping of data objects based on shared characteristics, determined through cluster analysis. This process finds out similarities among data points and organizes them into clusters without predefined class labels, making it an unsupervised learning technique. It's primary applications are in pattern recognition and spatial data analysis(Galit et al., 2007). The effectiveness of a clustering method is assessed by its capacity to unveil patterns which are hidden.

3.2 Role of Clustering Techniques in Education

The problem of improving students' academic performance is difficult for any university or institution. The scholastic achievement of first-year engineering students holds significant importance and typically has a direct impact on their cumulative grade point average, or CGPA. Clustering techniques can be used to group first year engineering students based on various characteristics and academic indicators, such as academic performance, attendance, coursework completion, and study habits. This allows educators to find out students who may be weak at performing or dropping out from their academics. By recognizing such clusters, educational institutions can provide timely interventions and support to help these students succeed.

3.3 Proposed Methodology

The main objective is to evaluate student's academic performance in upcoming semester based on previous semester's CGPA along with the internal assessments like quiz, assignment, mid-Test, attendance and lab performance using K-Means Clustering algorithm.

Requirements Analysis:

1.Real time Student Data Set: The real time data will be collected from Engineering students in Andhra Pradesh state.
Attributes in Student Data Set:

1. Student ID
2. Student Name
3. Branch Name
4. CGPA
5. Quiz
6. Assignment
7. Mid-test
8. Attendance
9. Lab Performance

2. K-Means Clustering Algorithm
3. Python for implementing k-Means clustering in determining performance of students.

K-Means clustering is performed on the student data consisting of 120 students collected across various engineering colleges in Andhra Pradesh state through survey form. K-Means clustering is performed on all the attributes and all students are divided into five clusters.

The dataset likely includes a range of attributes or features related to students, potentially encompassing academic performance metrics, demographic information, responses to survey questions, and other relevant factors. The choice of attributes for clustering plays a pivotal role in delineating meaningful patterns in the data.

Each cluster formed represents a group of students who share similar characteristics based on the attributes considered in the analysis. Interpretation of these clusters involves examining the commonalities among students within each cluster and discerning the distinguishing features that contribute to the clustering.

Insights and Applications:

The outcome of the K-Means clustering provides valuable insights into the diversity and patterns within the student population. These insights can have practical applications in various domains, such as educational policy-making, targeted interventions for specific student groups, or even optimizing resource allocation within educational institutions.
Validation and Refinement:
It's essential to note that the results of the K-Means clustering should be interpreted with care. Validation techniques may have been employed to assess the robustness of the clusters, and iterations of the analysis may have been conducted with adjustments to enhance the clustering efficacy.

Communication of results:

Finally, the findings of the K-Means clustering analysis should be effectively communicated, potentially through visualizations, summary statistics, or detailed reports. The implications of the clusters can be discussed with stakeholders, contributing to informed decision-making processes in the educational context.
In summary, the K-Means clustering on the student data from Andhra Pradesh engineering colleges offers a systematic approach to uncovering patterns, providing a foundation for targeted and data-driven strategies in the educational landscape.

3.4 Design of Proposed Methodology

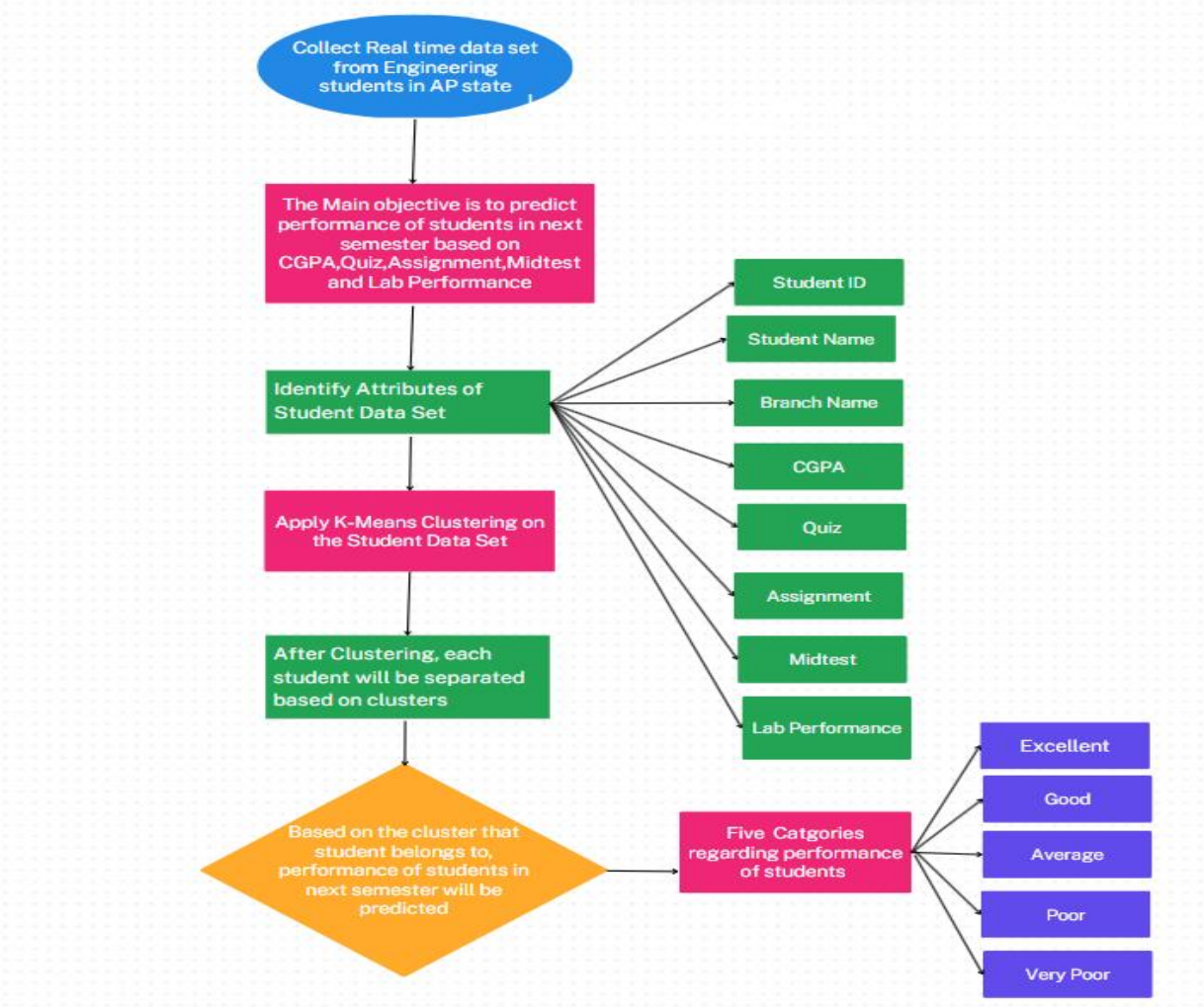

Figure 1: Proposed System working model

3.5 K-Means clustering algorithm

With a specified value of 'k', the k-means algorithm is executed through a sequence of steps:

STEP1.Segregate objects into k distinct and nonempty subsets
STEP2.Calculate initial seed points by determining the centroids of the current cluster partition(the centroid being the central or mean point of each cluster).
STEP3.Associate each and every with the cluster that has the nearest seed point.

STEP4.Return to step-2, repeating the process until no new changes are made.

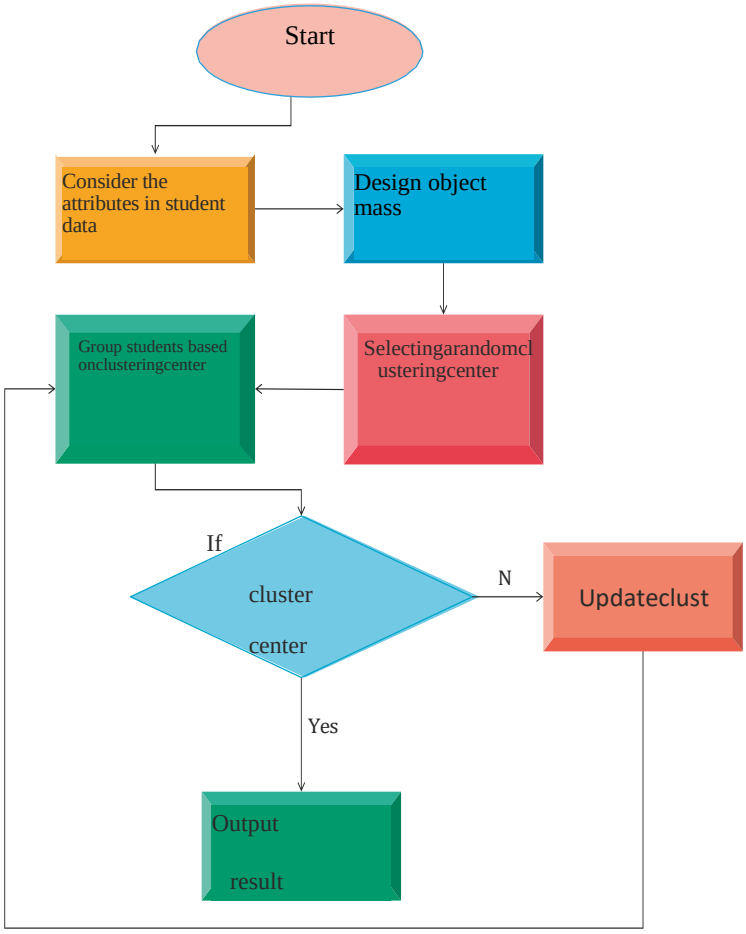

Figure 2: K-means algorithm

Methodology of Clustering Process:

1.Begin the process.
2. Data preprocessing is the first preliminary step before proceeding further because as a part of data preprocessing, data cleaning will be performed like handling out missing values, predicting outliers e.tc.
3.And then distance matrix construction is the next step. A distance matrix is a table that shows the distances between pairs of objects. Distance matrices are commonly used in various fields including clustering, multidimensional scaling and phylogenetics.
4. Students are grouped into clusters based on distance.
5.The farthest student will be selected as the cluster.
6.Student with least distance from the cluster will be made as a new cluster center.
7. Actual clustering would be performed on the data.
8. And then the required result will be obtained.
9.End of the process

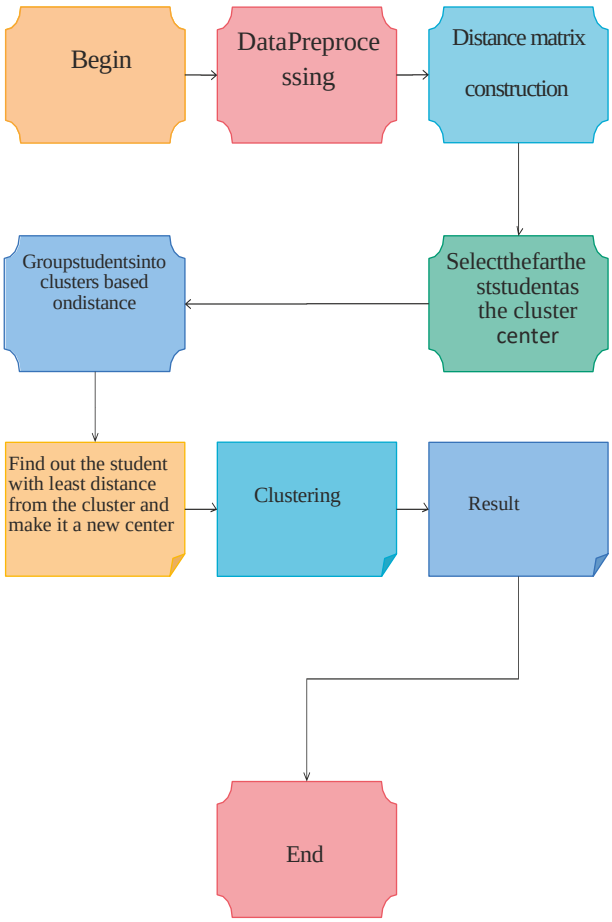

Figure 3: K-means clustering work procedure

4. Result And Discussion

After performing k-means clustering algorithm on all the attributes and on dividing all the students into five clusters, the results are :

Cluster 1 : Center 9.5325
Cluster 2 : Center 6.586
Cluster 3 : Center 8.52
Cluster 4 : Center 8.799807692307692
Cluster 5 : Center 7.973809523809524
Performance of students in next semester is calculated based on the cluster

Students who belong to cluster 1 are predicted to be performing “Excellent” in next semester.
Students who belong to cluster 2 are predicted to be performing “Very Poor” in next semester Students who belong to cluster 3 are predicted to be performing “Average” in next semester
Students who belong to cluster 4 are predicted to be performing “Good” in next semester
Students who belong to cluster 5 are predicted to be performing “Poor” in next semester.
https://github.com/dedeepya29/DataMiningProject.git

REFERENCES

Christy A., Malar J., Deva P., Siddique M., Kalpana P., Selvi, P. T., & Kirubakaran, R. (2019). A Novel Cluster Based Scheme for Node Positioning .In Indoor Environment An Efficient Traditional And Ontological-Based Distance Measures For Text Clustering Approaches With Its Performance, CSE001.

Connolly T., C. Begg & A. Strachan (1999), Database Systems: A Practical Approach to Design, Implementation, and Management (3rd Ed.). *Harlow: Addison-Wesley*.687.

Sathya D., Ibrahim S. & Jagadeesan D., (2023), Wearable Sensors and AI Algorithms for Monitoring Maternal Health, www.igi-global.com,. https://www.igi-global.com/chapter/wearable-sensors-and-ai-algorithms-for-monitoring-maternal-health/332209.

Galit (2007), Examining online learning processes based on log files analysis: a case study. Research, Reflection and Innovations in Integrating ICT in Education.

Ibrahim & M. Sivabalakrishnan, (2024), An Enhanced Weighted Associative Classification Algorithm Without Preassigned Weight based on Ranking Hubs, *International Journal of Advanced Computer Science and Applications*, Volume 10.

Ibrahim, S., & Sivabalakrishnan, M. (2019). An Enhanced Weighted Associative Classification Algorithm Without Preassigned Weight based on Ranking Hubs, *International Journal of Advanced Computer Science and Applications*, 10(10).

Kifaya(2009) , Mining student evaluation using associative classification and clustering, Communications of the *IBIMA* ,Volume 11, IISN 1943-7765.

Priyanka, R., & Ibrahim, S.(2015). A Survey on Infrequent Weighted Item set Mining Approaches. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 4(1), 2278-1323.

Devi S. & Anto, S. (2021), An efficient document clustering using hybridised harmony search K-means algorithm with multi-view point, *International Journal of Cloud Computing*, 10, 29 – 143.

Siddique I. S. P. , Sivabalakrishnan M. (2020),Rare Lazy Learning Associative Classification using Cogency Measure for Heart Disease Prediction, Intelligent Computing in Engineering, *Springer*,1125, 681-691. .

Sivabalakrishnan., Mohamed S P,(2018), Lazy Learning Associative Classification in Map Reduce Framework, *International Journal of Recent Technology and Engineering (IJRTE*), 7(4S), 168-172.

Siddique I.,Sivabalakrishnan, M. (2020), An Evolutionary Memetic Weighted Associative Classification Algorithm for Heart Disease Prediction. In: Hemanth, D., Kumar, B., Manavalan, G. Recent Advances on Memetic Algorithms and its Applications in Image Processing. Studies in Computational Intelligence, *Springer*, 8.

Selva K., Kalaivani S. , Siddique I. , S. P., & Swathi, G. (2023). Traffic and fragmentation aware algorithm for routing and spectrum assignment in Elastic Optical Network (EON). *Optical Fiber Technology*, 81, 103480.

Ibrahim P.S. , Rakshitha I., Vasisri ., R. H. Aswitha, M. R. Rao & Krishna D. (2023), Revolutionizing Solar Generation Data Mining through Advanced Machine Learning Algorithms: Novel Insights and Results, *International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, Bangalore, India , 1-8.

Siddique, & Sivabalakrishnan M. (2019). An Enhanced Weighted Associative Classification Algorithm without Preassigned Weight based on Ranking Hubs. *International Journal of Advanced Computer Science and Applications* 10(10).