Unveiling Patterns in Student Achievement Using K-Medoids Clustering

[1]P.Devendra Babu, [2]R.Manthru Nayak, [3]K.Sudhakar Reddy, [4] Y.Mahendra Reddy
Assistant Professor CMRCET

Abstract— In this research, the K-Medoids clustering algorithm is used to investigate student performance from information about grades, attendance, and participation in extracurricular activities. Three different clusters were identifier: Cluster 1 of high performing academics who also have good attendance, Cluster 2 of average performing students with poor attendance, and Cluster 3 low performing academics with high participation in extracurricular activities. The K-Medoids algorithm was used due to its outlier robustness to form accurate clusters.

Keywords-: Student interest model; K-medoids clustering algorithm: Student performance model; segmentation of patterns

I. INTRODUCTION

Student Performance Analysis

Analysis of student performance has grown all the more crucial in the contemporary educational environment where comprehension of the various influences on academic success is imperative in creating efficient learning strategies and support mechanisms. The traditional measure of student performance solely through grades leaves much to be desired, but it fails to capture the intricate dynamics of a student's life. Attendance and participation in outside activities and even behavioral facets, also impact academic performance considerably. Since learning institutions aim at offering all rounded support, methodologies that are able to detect intricate patterns in the behavior and performance of students become increasingly important.

Clustering Algorithms in Educational Data:

Clustering methods have been found to be useful tools in revealing patterns in educational data and segmentation of students into groups based on common properties of these, the K-Medoids clustering algorithm has the benefits of being resilient against outliers and capable of utilizing real data points as cluster centers (medoids). This makes K-Medoids very appropriate for educational data, which tend to include outliers or noise as a result of the diversity of student performance and activities. Through the use of K-Medoids clustering, this research seeks to pinpoint and examine unique patterns in student performance, with the purpose of giving actionable recommendations for.

Research Focus and Methodology:

We use in this study a dataset that has important measures of student achievement, such as grades, attendance history, and extracurricular activity. The data undergoes K-Medoids clustering to uncover latent patterns that cannot be easily observed using conventional analytical techniques. In particular, our analysis suggests three distinct clusters: high achieving students with high attendance, middle achieving students with lower attendance, and low achieving students who are highly active in extracurricular activities. These clusters provide a richer perspective on student performance, implying that academic performance is determined by a range of variables beyond grades alone.

Implications for Educational Practice

Implications of the present study's findings for educational practice are important. By identifying diverse student performance profiles, teachers are able to address the distict requirements of each student group. As an high achievers with good attendance may need gifted academic challenges, whereas low-achieving but extracurricular active students may need to be helped with balancing their non-academic and academic responsibilities. This study highlights the need for an integrated student development approach in which academic and nonacademic variables are examined simultaneously to create a more supportive and efficient learning environment.

Problem Statement: Student success is a function of a number of factors including social economic status, study habits, learning styles, and past academic achievement. The patterns and relationships between these factors and student success can prove difficult to identify because of the diversity and complexity of student information. The conventional approaches to analyzing student performance are not always successful in accounting for these patterns. The purpose of this research is to utilize the K-Medoids clustering algorithm in order to uncover patterns of hidden achievement among students. K-Medoids clustering, a powerful clustering method which partitions data into groups of similar elements, represents a variable option for the detection of significant patterns within student performance data. Using this method the research can categorize students into separate clusters by their achievement level, enabling teachers and policy makers to inform more targeted educational strategies, interventions, and personalized.

Objective:

- To Apply K-Medoids Clustering
- To Identify Distinct Student Profiles
- To Assess the Impact of Non-Academic Factor
- To Provide Actionable Insight

Research Gap: The problem addressed in this research is the need for a strong analytical method to measure the performance of students by aggregating multiple dimensions like academic marks, attendance and extracurricular activities. The prevailing clustering models may not be capable of handling the intricacy of educational data or removing outliers and noise. Therefore, there is a need to adopt a strong clustering model like K-Medoids to identify apparent trends in student performance and comprehend determinants of academic performance in a better way. This research aims to develop a more advanced understanding of the student profile to facilitate effective and targeted educational interventions. There is a need for research on how K-Medoids can be used effectively to treat large educational datasets and how its results can be used to create actionable plans. Additionally, looking at the intersection of different student attributes and how they influence educational outcomes will give further insight into what contributes to student success.

LITERATURE REVIEW

Application of K-Medoids in Student Data:

Fewer studies have utilized K-Medoids due to its computational complexity, but it has been used effectively in scenarios where outliers are present and where the dataset contains categorical or ordinal data.
Research examples:

- Studies analyzing student performance data based on various attributes such as attendance, grades, and participation using K-Medoids.
- Comparative studies showing how K-Medoids outperforms K-Means in certain educational datasets.

Case Studies and Applications

Real-world Applications:

- Case studies where K-Medoids clustering has been applied to analyze student performance, identify learning patterns, or group students for personalized learning plans.
- Examples from primary education, higher education, and online learning platforms.

Success Stories and Challenges:

- Successful implementation of clustering in improving student outcomes.
- Challenges faced in using clustering techniques, such as choosing the right number of clusters, interpreting cluster results, and dealing with large and complex datasets.

EXISTING SYSTEM

The traditional method of regression analysis, descriptive analysis and hypothesis testing is prevalent in student performance analysis. Certain of the prevalent clustering algorithms such as K-Means are often used for clustering students based on their behavior or performance. Hierarchical clustering creates a cluster tree that is often used in finding relationships between small sets of data. Collaborative Filtering Techniques are used in recommendation systems for making student performance prediction based on similarity to others.

Disadvantages:

- Over fitting & Poor Generalization
- Inflexibility with Mixed Data Types
- Interpretability
- Scalability
- Handling Imbalanced Data

PROPOSED SYSTEM

This work introduces a new approach for analyzing student performance using the K-Medoids clustering algorithm that overcomes the drawbacks of earlier clustering techniques through more stable and interpretable clusters. The suggested system will be able to achieve identifying useful patterns in student performance, effective outlier handling, and the generation of useful insights for instructors.

Data preprocessing:

- Data Collection:- Gathering Data
- Data Cleaning: Handling Missing Values (IMPUTATION & DELETION)
- Data Transformation: (Normalization, Standardization)
- Encoding Categorical Data
- Feature Engineering
- Initialization of Medoids
- Clustering Analysis

K-MEDIODS NUMERICAL:

STUDENT	MATHS	SCIENCE
A	85	78
B	92	95
C	70	65
D	88	85
E	80	70
F	60	55
G	75	80
H	95	92
I	65	60
J	78	75

STEP 1:-Let's assume we want to create 2 clusters, so k=2k = 2k=2.
My clusters are B (92, 95), F ((60, 55) :
Use Manhattan distance FROM centroids

- Distance to Student B(92,95):|85-92|+|78-95|=7+17=24|85-92|+|78-95|=7+17=24
- Distance to Student F(60,55):|85-60|+|78-55|=25+23=48|85-60|+|78-55|= 25+23=48

The initial clusters are:

- Cluster B: A, B, D, G, H, J
- Cluster F: C, E, F, I

STEP 2:- Update Medoids: Suppose Student E (80, 70) becomes the new medoid for Cluster F
Repeat Until Convergence

Final Clusters

After convergence, let's say the final clusters are:
- Cluster 1 (Medoid: D (88, 85)): A, B, D, G, H,J
- Cluster 2 (Medoid: E (80, 70)): C, E, F, I
- Cluster 1: This cluster might represent students with consistently higher scores in both subjects
- Cluster 2: This cluster might represent students with relatively lower or average scores.

Key advantages of K-Medoids:

1. Flexibility with Different Distance Metrics
2. Suitability for Mixed Data Types
3. Improved Cluster Stability
4. Applicability to Real-World Educational Data
5. Enhanced Decision Making

REAL TIME WORKING SCENARIO WITH STUDENT DATA

➢ A Data Set has been given as input with subject scores and extracurricular activities

STUDENT ID	NAME	CLUSTER	MATH-SCORE	ENGLISH -SCORE	ATTENDANCE	EXTRA CIRICULAR ACTIVITIES	% HW COMPLETION
1	Alice	1	92	90	95	Debate Club	95
2	Bob	1	88	91	98	Science Club	92
3	Carol	1	85	87	97	Math Club	90
4	Dave	2	75	70	60	No Extracurricular	65
5	Eva	2	72	68	55	No Extracurricular	60
6	Frank	2	70	66	58	No Extracurricular	55
7	Grace	3	60	55	85	Art Club	70
8	Henry	3	58	60	90	Sports Team	75
9	Iris	3	65	62	88	Choir	80
10	Jack	3	64	61	87	Debate Club	70

K-MEDIODS IMPLEMENTED(CODE)

```
Iimportnumpy as np
import pandas as pd
fromsklearn.preprocessing import MinMaxScaler
fromsklearn_extra.cluster import K Medoids

# Create the dataset
data = {
    'Student ID': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10],
  'Name': ['Alice', 'Bob', 'Carol', 'Dave', 'Eva', 'Frank', 
'Grace', 'Henry', 'Iris', 'Jack'],
'Math Score': [92, 88, 85, 75, 72, 70, 60, 58, 65, 64],
'English Score': [90, 91, 87, 70, 68, 66, 55, 60, 62, 
61],
```

```
'Attendance (%)': [95, 98, 97, 60, 55, 58, 85, 90, 88, 87],
'Extracurricular Activities': [1, 1, 1, 0, 0, 0, 1, 1, 1, 1],
'Homework Completion (%)': [95, 92, 90, 65, 60, 55, 70, 75, 80, 70]
}

# Load data into a DataFrame
df = pd.DataFrame(data)

# Select features for clustering
features = ['Math Score', 'English Score', 'Attendance (%)', 'Extracurricular Activities', 'Homework Completion (%)']
X = df[features]

# Normalize the features
scaler = MinMaxScaler()
Normalized = scaler.fit_transform(X)

# Perform K-Medoids clustering
kmedoids = KMedoids(n_clusters=3, random_state=0).fit(X_normalized)

# Adjust cluster labels to match input numbering (1, 2, 3)
labels = kmedoids.labels_
unique_labels = sorted(set(labels))
label_mapping = {old_label: new_label + 1 for new_label,
old_label in enumerate(unique_labels)}

df['Cluster'] = [label_mapping[label] for label in labels]

# Display the Data Frame with adjusted cluster assignments
print(df)
```

At output dataset with change of cluster with respect to extra circular activities

STUDENT ID	NAME	MATH-SCORE	ENGLISH-SCORE	ATTENDANCE	EXTRA CIRCULAR ACTIVITIES	% HW COMPLETION	CLUSTER OUTPUT
1	Alice	92	90	95	Debate Club	95	1
2	Bob	88	91	98	Science Club	92	1
3	Carol	85	87	97	Math Club	90	2
4	Dave	75	70	60	No Extracurricular	65	3
5	Eva	72	68	55	No Extracurricular	60	3
6	Frank	70	66	58	No Extracurricular	55	3
7	Grace	60	55	85	Art Club	70	2
8	Henry	58	60	90	Sports Team	75	2
9	Iris	65	62	88	Choir	80	2
10	Jack	64	61	87	Debate Club	70	2

FUTURE DIRECTIONS

One development area is to improve the scalability of the K-Medoids algorithm to handle bigger and more intricate datasets hybrid clustering approaches that combine K-Medoids with other algorithms or machine learning models can make clustering results more robust and adaptable. Combining clustering resulting results with predictive models can allow researchers to provide more precise predictions of student performance and tailor learning approaches to unique needs. Combining multimodal data, such as qualitative feedback, social interactions, and physiological signals, can also be used to improve analysis, providing a more detailed picture of student performance.

CONCLUSION

K-Medoids algorithm for grouping student's performance into effective groups according to their achievement and other desired traits. Scalability of the K-Medoids algorithm with big data and high dimensional data is a problem. Generally, this research offers important contributions to the application of K-Medoids in educational data analysis, offering a foundation for further research that can improve student performance and educational attainment.

DATASET REFERENCES:
Kaggle:https://archive.ics.uci.edu/ml/datasets/Student+Performance
Source: http://www3.dsi.uminho.pt/pcortez

REFERENCES

[1] Spatial-temporal correlation analysis of mountain wind speed information in western Sichuan plateau based on improved K-Mediods clustering method, 24-26 November 2023, 10.1109/SGEE60678.2023.10481696

[2] VLC System Employing Clustering Algorithm, 31 May 2021, 10.1109/ICTC51749.2021.9441663

[3] Improved K-means Based on Density Parameters and Normalized Distance, 05-08 March 2021, 10.1109/ICBDA51983.2021.9403172

[4] Analysis of Students Behavior Characteristics Based on K-Mediods + Éclat, 05-07 May2021,10.1109/CSCWD49262.2021.943763

8
[5] A SURVEY OF K-MEDOIDS ALGORITHMS CLUSTERING TECHNIQUES – A REVIEW AND COMPARISON OF K-MEANS, K-MEDOIDS, DBSCAN
[6] K-MEDOIDS CLUSTERING ALGORITHM EXPLAINED
[7] UNDERSTANDING K-MEDOIODS