

Using Regression and Clustering Algorithm to Classify Student Attentiveness

Swapnil Gaikwad¹, Abhishek Nagare², Sakina Shaikh³

^{1,2,3} *Bhartiya Vidya Bhavan's Sardar Patel Institute of Technology Munshi Nagar, Andheri(W), Mumbai 400058*

Abstract—There have been multiple studies where researchers have tried to classify student attentiveness. Numerous of these approaches depended on a qualitative examination and lacked any quantitative examination. This study bridges the gap between qualitative and quantitative approaches to classify student attentiveness. The findings from this study can help teachers at all levels improve their teaching methods and implement individualized learning. This study uses video data as input, employing machine learning methods such as linear regression and hierarchical clustering to automatically classify students as attentive or inattentive. Machine learning algorithms are used to train classifiers that estimate time-varying attention levels of individual students. Human-observed attention levels serve as a benchmark for comparison. This activity improves students' ability to perform in their respective fields and contributes to the development of effective educational methodologies.

Index Terms—Linear Regression, Hierarchical Clustering, K- Means, Personalized Learning System

I. INTRODUCTION

Numerous research efforts have been made to determine student attentiveness in classroom situations. Many of these studies focused on qualitative measures to detect and evaluate attentiveness, while others explored quantitative approaches. For instance, the use of biometric bracelets has been explored to measure student attentiveness. Similarly, tracking eye and head positions, as well as analyzing body language, has been employed to infer attentiveness in computer network courses.[4]

Recent advancements in machine learning algorithms and video analytics have enabled automatic monitoring of students' behavior and affective

states, from college to pre- school levels. Visual indicators such as facial expressions, head movements, and body postures can detect affective reactions like interest, boredom, or confusion.[3] In teaching research, monitoring student attention levels provides valuable feedback to instructors, enabling them to tailor their teaching methods to individual students. Students typically fall into categories such as innovators, risk-takers, keepers, or resistant learners.[5] For effective classroom management and personalized learning experiences, it is essential for teachers to evaluate visible signs of attention.[1]

This study discusses a system that uses video data to track and record student behaviors, such as movements, postures, and facial expressions, to determine their attentiveness. Data is analyzed using machine learning methods, including hierarchical clustering and linear regression, to classify students as attentive or inattentive. Hierarchical clustering organizes data into a tree-like structure, providing insights into the relationships between clusters.[9]

II. METHODOLOGY

A. System Overview

The system begins by observing a particular student through video input. The program recognizes and counts behaviors indicative of attentiveness in real-time. This data is stored in a database for analysis. Regression techniques are then used to analyze relationships between variables, while clustering methods group similar data points to classify attentiveness.

Machine Learning Approach Steps:

Data Preparation: Extract relevant features from video data, such as facial expressions and postures.

Data Analysis Model Implementation: Apply regression analysis to establish relationships between behavioral features and attentiveness. Use clustering algorithms like K-Means and hierarchical clustering to group data.

Result Visualization: Represent outcomes through graphs and diagrams for better interpretability.

- 1) Preparing the data
- 2) Implementing a data analysis model
- 3) Visualizing the results

B. Data. Clustering

K-Means Algorithm:

- 1) Select random centroids from the dataset values.
- 2) Calculate the Euclidean distance between data points and centroids.
- 3) Assign points to the nearest centroid.
- 4) Recalculate cluster centroids and repeat until centroids stabilize.

Hierarchical Clustering:

- 1) Organizes clusters into a tree-like structure (dendrogram).
- 2) Unlike K-Means, it doesn't require a pre-determined number of clusters.

3) Regression Techniques

Linear Regression:

- 1) Predicts a dependent variable ('Y') using a linear relationship with an independent variable ('X'):

Formula: $Y = a + bX$

$= a + bX$

- 2) Regression minimizes errors in predictions by calculating the optimal slope and intercept. This method is particularly useful for predicting students' academic performance based on attentiveness metrics extracted from video data.

C. EXPERIMENTS

A. Data Generation To simulate student attentiveness traits, random data was generated within a predefined range (0 to 100). These values represent the frequency of behaviors such as hand raises and eyebrow movements, which serve as indicators of attentiveness. Each data point corresponds to a unique student, with two features—frequency of hand raises and eyebrow movements—being used for visualization and analysis.

D. Generated Data Examples

- Feature 1 (Gaze Shifts): Represents attention shifts

and visual scanning behavior, indicating potential distraction or focus.

- Feature 2 (Head Movements): Signifies active engagement or a response to stimuli, reflecting attentiveness.
- Feature 3 (Eye Blinks): Indicates alertness or drowsiness, correlating with cognitive state.
- Feature 4 (Eyebrow Movements): Suggests cognitive activity, curiosity, or emotional response.
- Feature 5 (Hand Raises): Reflects active participation and engagement in the activity or lesson.

The generated dataset is subsequently analyzed using regression and clustering methods.

E. Results

The analysis demonstrates the effectiveness of clustering and regression in categorizing students into attentive and inattentive groups. Key insights include:

1) Clustering Results:

K-Means effectively grouped students into two distinct clusters based on attentiveness metrics. Hierarchical clustering provided a dendrogram, visualizing the relationships among clusters and confirming the robustness of K-Means results.

1) Regression Results:

Linear regression illustrated a strong relationship between behavioral features and attentiveness. Visualizations such as scatter plots with regression lines highlighted predictive accuracy.

III. FLOW-CHARTS AND OUTPUTS

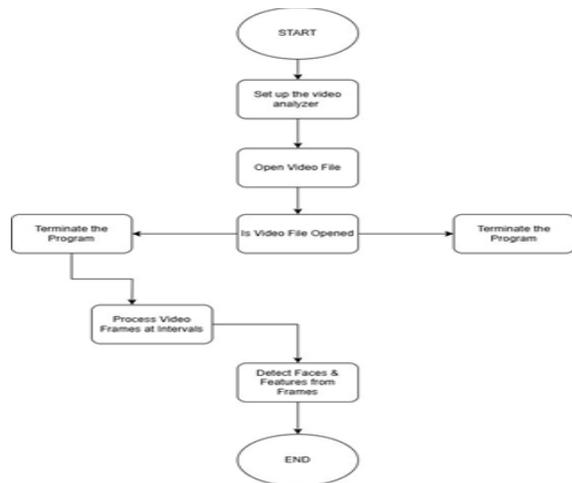


Fig. 1. Work Flow

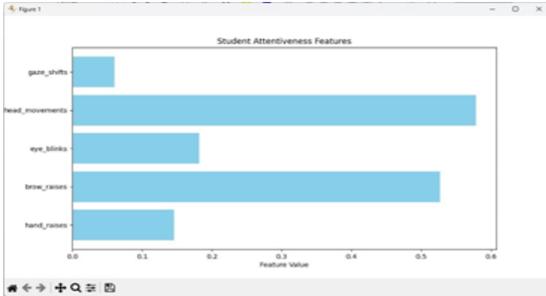


Fig. 2. Output1

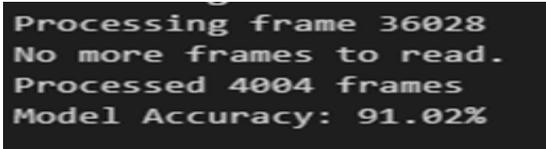


Fig. 3. Output2

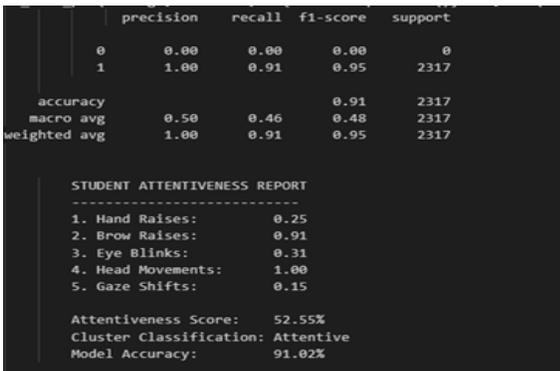


Fig. 4. FinalOutput

IV. CONCLUSION

This research lays the foundation for developing a system capable of automatically classifying students in a learning environment as attentive or inattentive. The use of machine learning methods—specifically clustering and regression—allows for accurate and efficient classification based on visual behavioral features. Key contributions include:

- 1) A framework for assessing attentiveness in real-time using video data.
- 2) Insights into student behavior that can help educators adapt teaching methods for improved engagement and learning outcomes.

REFERENCES

[1] Muhammad Bilal Shaikh and Douglas Chai, "RGB-D Data-Based Action Recognition: A

Review," 2021.

- [2] Janez Zaletelj and Andrej Kosir, "Predicting students' attention in the classroom from Kinect facial and body features," 2017.
- [3] Manus Ross et al., "Using Support Vector Machines to Classify Student Attentiveness for the Development of Personalized Learning Systems," 2013.
- [4] Margaret Martinez, "What is personalized learning? Are we there yet," E-Learning Developer's Journal, 2002.
- [5] Ning Fang and Shaobo Huang, "Regression Models of Predicting Student Academic Performance in an Engineering Dynamics Course," 2010.
- [6] Diti Modi, "Diving into K-Means," 2018.
- [7] Andriy Blokhin, "Linear vs. Multiple Regression: What's the Difference?," 2021.
- [8] "Cluster Analysis: Basic Concepts and Algorithms."
- [9] Jessica McBroom et al., "DETECT: A Hierarchical Clustering Algorithm for Behavioural Trends in Temporal Educational Data," 2020.