https://doi.org/10.51153/kjcis.v5i2.122

An Assessment for Understanding Student Behaviour by Applying Machine Learning Technique

Syeda Kainat Ahmed[1] Syed Mushhad M. Gilani*[2] Sidra Sultan[3]

Abdur Rehman Riaz[4] Muhammad Wasim Abbas[5]

Abstract

In the past, research have been carried out to study the behaviour of students as it is an important topic in psychology. Parents and teachers are concerned about their children's actions in class. Learning about students' behaviour in school is essential for teachers for their own development and growth of them. A class is composed of students with distinctive characteristics and capacities, where a few students are sharp and some are dull. In some cases, it becomes inconvenient for the educators to spot who is keeping the pace with them and who is falling behind. The proposed approach will allow the students to groom their personalities and overcome their shortcomings, and it will offer assistance to the instructor to identify which students require more consideration from them. To do that, we chose and applied the K-mean Clustering Algorithm. In other words, this research attempts to discover homogeneous subgroups inside the information. K means calculation is an iterative calculation that tries to tract the dataset into K pre-defined unmistakable non-overlapping subgroups (clusters), where each information point has a place as if it were one bunch. It tries to make the intra-cluster information focuses as comparable as conceivable, whereas also keeping the cluster as diverse (distant) as conceivable.

Keyword: Machine Learning, K-mean Clustering Algorithm, Behaviour Understanding, Student Psychology

1. Introduction

Child education plays an important role in the progression of any country [1]. In the future, they must serve and withhold the country. Thus, it is vital to understand the student's behavior to give them quality education in the best suitable environment, as education is

[1]*PMAS Arid Agriculture University, Rawalpindi | syedakainatahmed844@gmail.com*
[2]*The University of Agriculture, Faisalabad | mushhad@gmail.com (corresponding author)*
[3]*PMAS Arid Agriculture University, Rawalpindi | sidrasultan97@gmail.com*
[4]*PMAS Arid Agriculture University, Rawalpindi | abdurrehman.ar475@gmail.com*
[5]*Computer Science, University of Agriculture Wuhan, China | wasimabbas@whu.edu.cn*

the right of every child [2]. It can only be possible under some expert guidance, which can teach them right from wrong, how to overcome their fears, and shortcomings, improve their grades, skills, effective learning methods, etc. [3]. A single class is composed of students with different characteristics and abilities, where some students are intelligent while some are less able. Sometimes it becomes difficult for the teacher to spot which student is picking up the pace of the teacher and which is falling behind [4].

It is observed that there is an increase in the number of students in educational institutes every year as compared to the previous year. To manage the increasing number of students, they are divided into different sections. The sections are formed generally and are not based on any particular characteristics of students, because, sometimes, it becomes difficult for the teachers to handle the large strength of a different variety of students. Some students are shy while some are confident, some are disorderly while some are obedient, and some are dull while some are intelligent. There is a difference in the interaction of students with other students and teachers as well. So, in this project, a system is designed that predicts the behavior of students as given by [5]. For this, firstly a survey was designed containing several questions that were filled in both by the students and teachers. In this, students with similar characteristics are grouped in one section and a teacher is assigned to them according to their behavior. This helped the teachers to make sure that everybody was moving along at the same pace while delivering lectures, the school (staff, administration) to improve their results, parents to keep an eye on their children's progress, and also the students to groom their personalities and overcome their shortcomings [6].

This research work allows the students to be grouped according to their capabilities, which helps the teachers to learn which students require more attention from them. It helps both teachers and students to deliver their lectures more effectively and to improve their grades and skills, respectively. For this purpose, data was gathered from different schools, then students were divided into different categories according to their behavior through the application of Machine Learning (ML) techniques. Students with similar behavior and attributes were assigned to the same section. Lastly, the teacher whose personality matched with those students was made responsible for their teaching. This removed the communication barriers between the teacher and students and provided a friendly class atmosphere for students to express themselves. Parents also had access to their children's progress reports. Moreover, the school administration had a record of teachers' and students' reports [7]. In section I, Introduction has been discussed. Section II deals with the related research while Section III describes the proposed system and section IV results are shown.

2. Related Work

The research was performed to study the effects of hospitalization on children [8]. The online

survey was taken from the parents and its outcome revealed the types of mental pressure and diseases faced by the children. The studies also suggested some remedies for these types of disorders. Another experiment was conducted in a school to observe the behavior of children while evacuating in non-emergency situations without any guidance using the Cellular Automata CA model [9]. It was perceived that children in groups were taking longer time than those who were walking alone. In the end, they matched the results with their simulated results to ensure the correctness and precision of their designed system [10].

Zheng, Jiang and Shen [11] proposed a system that can spot attitudes with low-resolution and extreme impediments to help teachers, they can improve their teaching quality. They built a large-scale data set. An improved Region-Based Convolutional Neural Network (R-CNN) network was presented to discover student behaviour in the real classroom. Online Hard Example Mining (OHEM) was used for class imbalance issues. Fan et al. [12] describe a system that can remodel the activity occasions of the examinee about signaling data obtained by Kinect. It examines the attributes of activity occasions from period and recurrence proportions. The misconduct or disobedience of the examinee was also found in the studies. They did it by information procurement, pre-preprocessing module, occasion distinguishing module, misconduct observing tool, and track record constitute [13].

The research in [14] detects the behavior of the students while doing the class tasks and activities by using a round-robin coding strategy, regression tree, and observing students with the naked eye [15]. It suggested that if students are looking at the teacher, they are on-task. Otherwise, they are off-task. Some of the reasons for off-task behavior are self-intrusion, peer diversion, natural diversion supplies, strolling, or other self-distractions. Singh [16] diagnosed Attention Deficit Hyperactivity Disorder (ADHD) in children that experienced loss of self-control because of cognitive disabilities and unfriendly habitat and observed their behavior under different circumstances. It was discovered that the children can control themselves if they are not bothered unnecessarily. It was done by grouping children into 3 and then interviewing them for one hour with four different female researchers. Some questions were also asked by the parents because parental behavior also influences children's behavior [17].

In another method, 51 children ages ranging from 4 to 6 years were interviewed (qualitative approach) and asked what happiness is to them and what and who makes them happy [18]. The studies show that, not feeling tired or lazy and having positive affiliations, playing, learning, writing, drawing, helping their mothers, reading comics, watching TV and sports make them happy. Relationships with people in the family, friends, and their toys are also a source of happiness for them [19]. Lima et al. [20] discussed causes and remedies for childhood depression. They did a qualitative analysis of 180 articles, out of which 25 were in their domain. Out of those 25 articles, data was extracted and arranged in compiled form. The outcome of the research was based on etiology, diagnosis, prevention, prognosis, and treatment [21].

Safaei and Youzbashi [22] held a correlated comparative study, a multistage method, and used SPSS Software version 25.0 for data analysis. Their research included 140 people, out of which 80 were girls and 60 were boys, and their ages ranged from 8 to 10 years for the contrast of seriousness of the fixation and working memory in children with the obsessive-compulsive disorder and healthy children. The outcomes indicated that there was a huge connection between working memory and habitual issues in girls and boys [23]. The figures for mean and Standard Deviation (SD) in youngsters and patients were (177.24+-11.02) and (171.11+-8.08) respectively.

Kessels and Heyder [24] inspected the mental convenience of locks engaging in disruptive behavior for low-achieving students from an attributional point of view. In their experiment, 178 ninth-grade students were selected, and targeted the those students, who displayed disruptive behavior. They connected multilevel examination while testing for mediation effects. It was also discovered that those students were more popular but not liked personally. The research also included the comprehension of troublesome behavior in class as an endeavor of children showing unsatisfactory results, to inspire face-saving attributions and improve their peer status [25]. A taxonomy of the literature review is shown in Figure 1. Table I shows the previous literature and also discuss their finding and limitation.

Table 1: Comparison table of related Studies

State of the art Approaches	Findings	Limitations
Online questionnaire for parents about behaviour of children (Jiao et al., 2020).	Parents were instructed to, spend time with their children, and develop an interest in music for mental relief and comfort.	Only suggestions were given no algorithm was applied
Cellular Automata (Chen et al., 2019).	Grouped children took more time for evacuation A path with fewer obstacles regardless of distance was chosen	Behaviour of children under guidance was also not detected
OHEM combined with R-CNN is used to detect student behaviour (Zheng, Jiang and Shen, 2020).	The proposed system detects more behaviors with low-resolution and severe occlusion and helps teachers to improve teaching quality.	Behaviour of Students with selected poses was detected
An approach consists of four components data acquisition and pre-processing module event identification module misbehaviour monitoring engine record module (Fan et al., 2016).	Their approach rebuilds the action events of the examinee in terms of gesture and then these gestures are used to detect misbehaviour	There exists a possibility that the input data may comprise garbage value.
51 kids ranging from 4 to 6 years were interviewed (Izzaty, 2018).	55% of children said not feeling lazy or tired is happiness	No objective analysis was carried out. The sample size was too small (size=51)

Multistage cluster sampling method (Safaei and Youzbashi, 2020).	Healthy groups had higher working memory and lesser severity of obsession compared to patients	A very small number of students were taken for tests and all were of adult age.
Multilevel analysis while testing for mediation effects (Kessels and Heyder, 2020).	Disruptive behaviour causes a lack of effort instead of a lack of interest.	There is no information that whether the disruption was conscious or not.

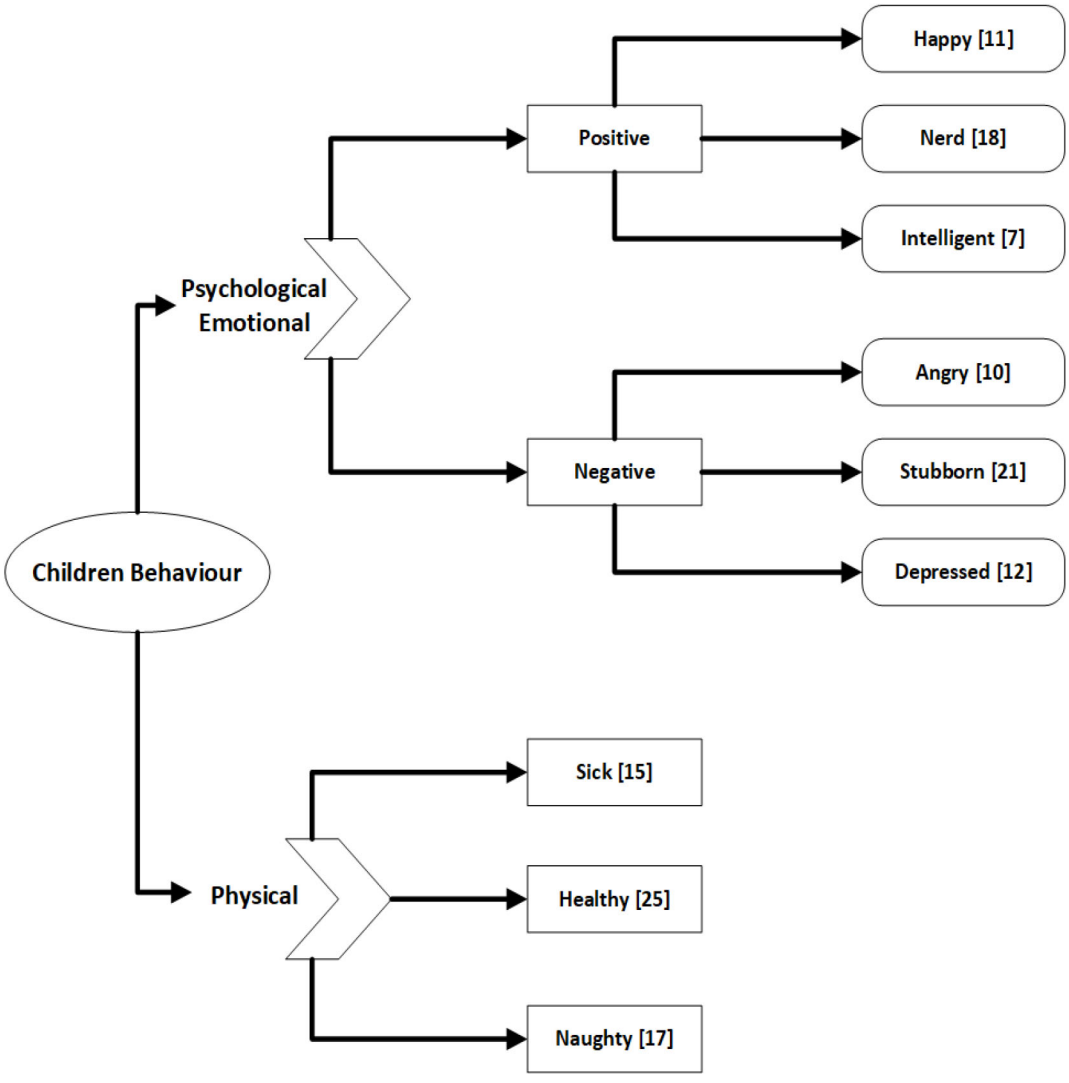


Figure 1: Taxonomy of child behavior interconnected examinations

3. System Model

The proposed model is supported to clarify the kinds of data that can emerge from a thematic approach for analysis that provides simple guidance on the opportunities and limitations of such data. The proposed model consists of five stages, as illustrated in Figure 2.

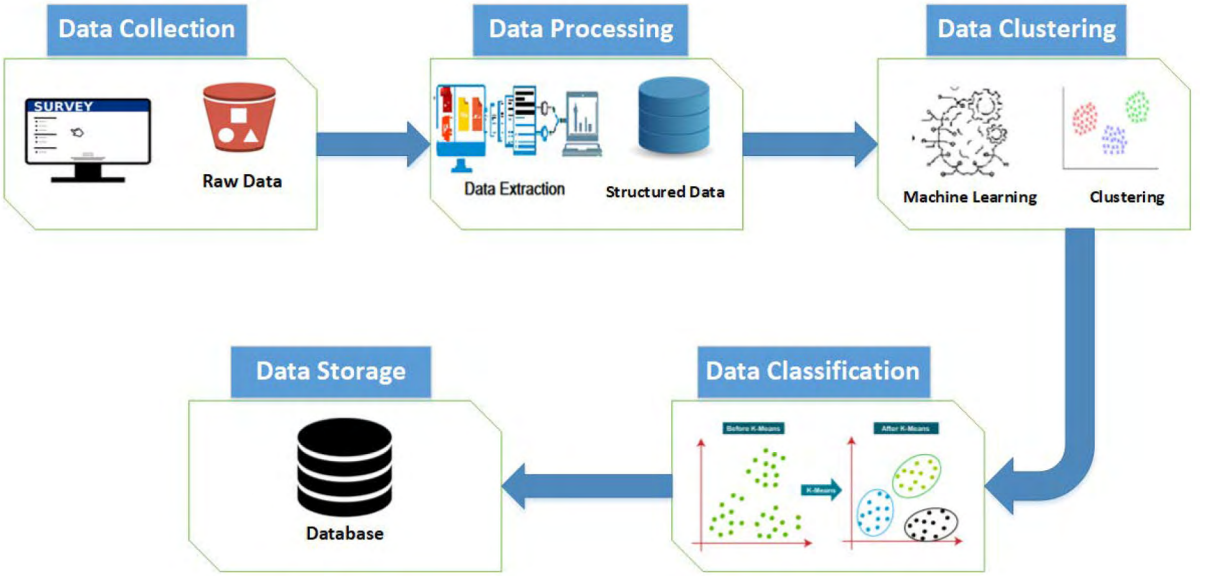
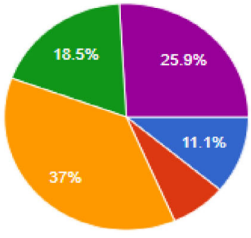


Figure 2: Proposed system

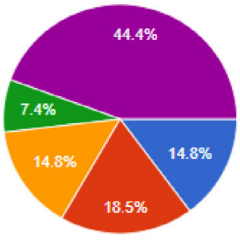
A. Data Collection

The first stage of the system model is data collection, in which a survey is conducted to store initial raw data. We designed diverse kinds of questions for students and asked them to rate themselves from 1 to 5. Where 1 was for the lowest and 5 was for the highest rank. The collected data is arranged in the form of charts and spreadsheet files, as details are presented in Figure 3. The chart contains the following rankings

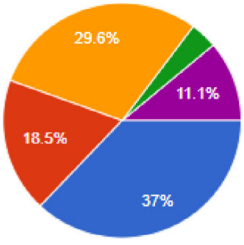
- Blue represents the students who rated themselves with a 1
- Red represents the students who rated themselves with a 2
- Orange represents the students who rated themselves with a 3
- Green represents the students who rated themselves with a 4
- Purple represents the students who rated themselves with a 5



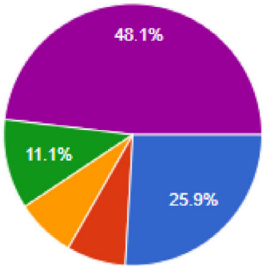
(a): I am talkative.



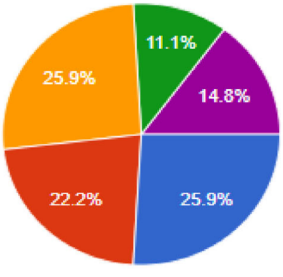
(b): I am original and/or always come up with something original.



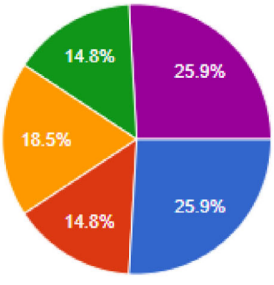
(c): I am reserved.



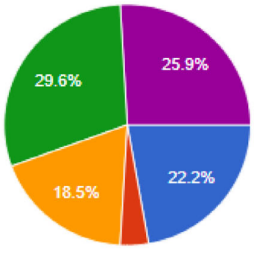
(d): I am helpful and unselfish with others.



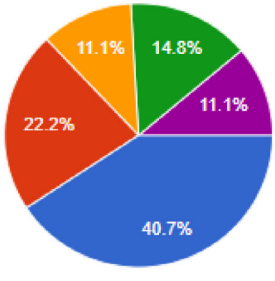
(e) :I am careless.



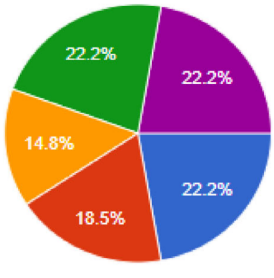
(f): I can handle stress well.



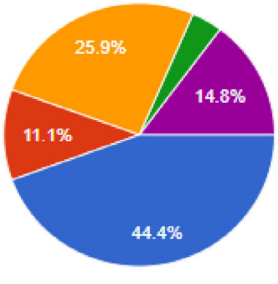
(g): I ask questions.



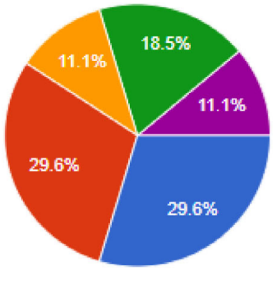
(h): I am lazy.



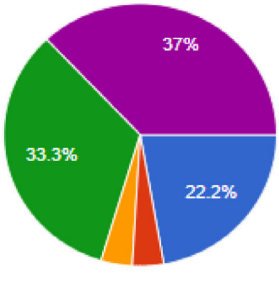
(i): I like to present in front of the public.



(j): I am shy in asking questions.



(k): I take time to understand things.



(l): I finish my work on time.

Figure 3: Results compiled by the proposed approach

B. Data Pre-processing

The second step is data preprocessing, in which a large amount of data is extracted and filtered. The randomly saved data is converted into usable and understandable sets of data. Then transformed it into structured data by storing it in the database. It is a very vital step to generate accurate results. Sometimes, a developer or researcher is tempted to skip this step and move to the next but clustering of unstructured data is not a good approach. When we have a big amount of data, a lot of data is not required and it should be filtered out. For preprocessing the data set, we perform the following steps:

i. We prepared a google form and filled it by experts.
ii. Data is extracted from google Forms.
iii. Data cleaning is performed on extracted data like removing some extra columns.
iv. The refine data has 27 rows and 26 columns
v. K-mean clustering is applied to clean data
vi. By applying the algorithm, we get the results

C. Data Clustering

Data Clustering is the third step. For categorizing, we applied k-means clustering to our structured data and, k as a result, we formed different sections of students. We set the value of k equal to the number of sections in which we want to divide our students, where k is equal to the total number of sections. We used the distance formula to determine the Euclidean distance as given in the following equation (i). The equation, 'd', represents the distance which can handle both ordinal and quantitative values. 'X' and 'Y' are the two coordinates or dimensions. A limit is applied on the dimensions, which is starts from '1' and runs till its value riches to 'p'

$$d(x,y) = \sum_{i=1}^{p} |xi-yi| \qquad \text{(i)}$$

Pseudocode for k-means clustering Algorithm:

Inputs:

D = {t1, t2……, tn} // student's data

K //total number of sections

Output:

K // k sections of students

//initialization

1. K-means algorithm:

2. Assign initial or random values for centroids (m1, m1……, mk)
3. For given values of iterations:
4. Iterate through random values:
5. Find the mean closest to the value
6. Assign a value to mean
7. Update mean

8. Repeat:

9. End

Assign the student each time to the section with which it has a minimum distance until convergence is achieved.

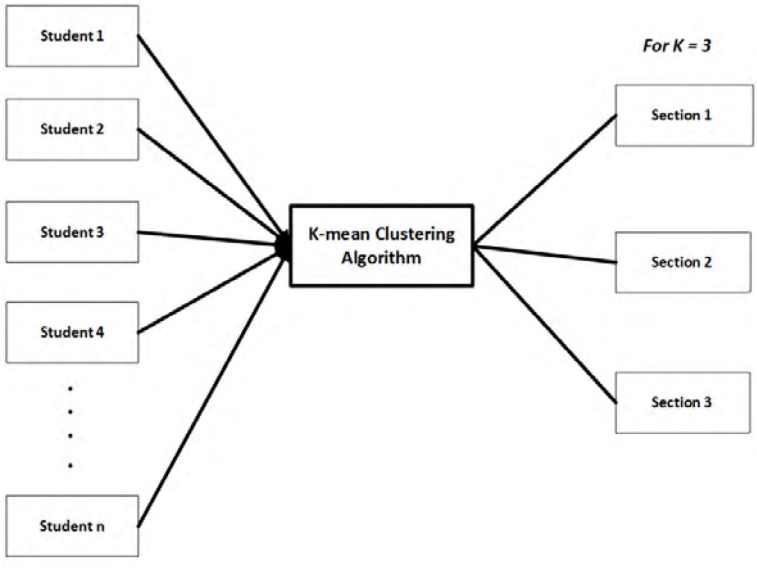


Figure 4: Demonstration of K-mean clustering algorithm for performing classification

D. Data Classification

The next step is data classification, in which the students with the same characteristics and habits are grouped and allotted the same section. Depending on the students' behavior, the most suitable teacher for them was assigned. This impact of the K-mean clustering algorithm is shown in Figure 4. This figure explains that there are a number of students. Every student has their own habits, characteristics, and behavior. When we applied the K-mean clustering algorithm to the students, it is fall in one section. Here we have a value of k is 3 so we have 3 sections, which means 3 types of categories for classification.

E. Data Storage

All the machine learning and artificial intelligence algorithms are based on data. Data is the lifeline in it and for the prediction of anything this data is used. The fifth and last stage is data Storage. All grouped data and information of students and teachers are stored in the database. Once data is collected, it must be stored in some reliable and secure database. The database is scalable thus that it can be increased or decreased when required.

4. Simulation and Results

The algorithm was applied and, as a result, regular subgroups or sections were made through the provided information. The algorithm divided the dataset into K pre-defined unique non-overlapping subgroups (clusters) where each information point has a place as it were one bunch. The following results were obtained after the first attributes were divided into 3 subgroups and the algorithm was applied.

In the first subgroup, the attributes considered were regarding the age and gender of the students, along with the attributes like students' analyzing behavior including talkativeness, innovation, reserved nature, carefulness, deep thinking skills, stress handling, and worrying. While considering all these attributes, k-means clustering (keeping k = 3) was applied as illustrated in Figure 5 (a). It divided students into 3 clusters (0,1,2) in the ratio [11:6:5]. The figure shows that those students having carefulness and deep thinking are less chance of stress and worry.

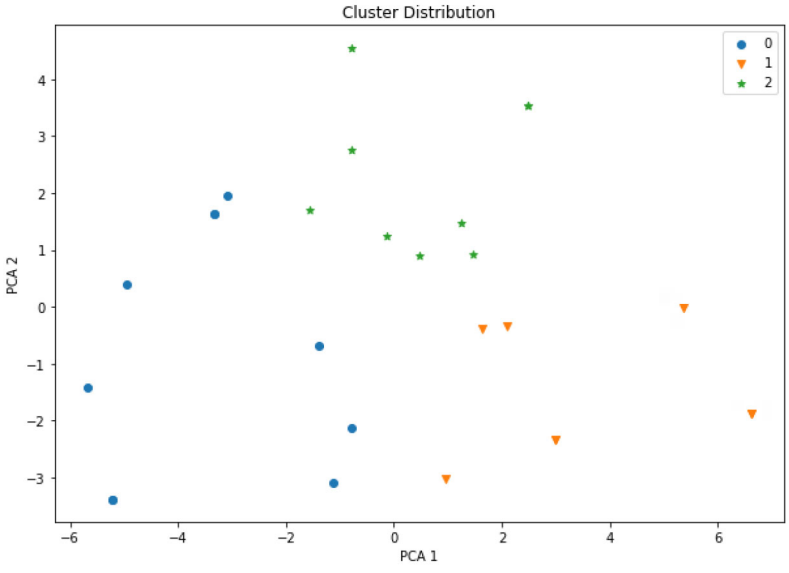


Figure 5 (a): k-means clustering (keeping k=3) on attribute set A

The second subgroup had students based on their age, gender and further attributes including laziness, confidence, shyness, obedience, punctuality, ability to work in groups, following directions while working, and understanding of nature. While considering the above-mentioned attributes, the K-means clustering algorithm (keeping k=3) was applied, as is illustrated in Figure 5 (b). This resulted in the division of students into 3 clusters (0,1,2) in the ratio [8:6:8]. When the graph reaches 0, as it gives the highest values apart from values that have medium or low points.

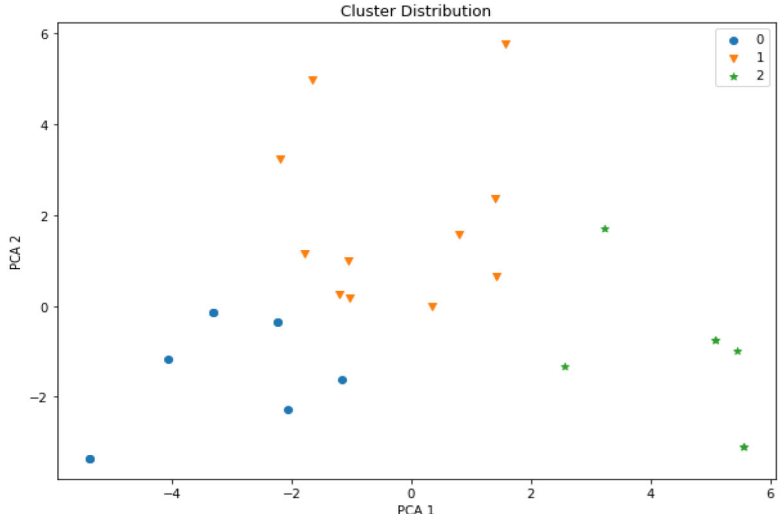


Figure 5 (b): k-means clustering (keeping k=3) on attribute set A

The attributes like age and gender of the students along with behaviors like deep thinking, stress handling, laziness, helping nature, worrying, interest in games and their frequent question asking were placed in the third subgroup. The k-means clustering algorithm (keeping k = 3) was applied by considering all of the above-mentioned attributes as illustrated in figure 5 (c). The students were divided into 3 clusters (0,1,2) in ratio [8:11:4]. Most of the values in the graph are high when they are away from the center value, which is zero. At zero, clusters have low values.

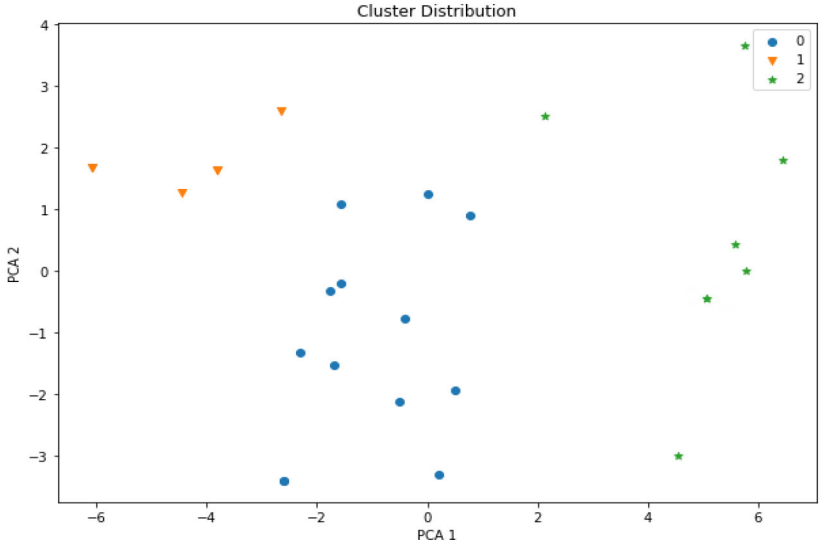


Figure 5 (c): k-means clustering (keeping k=3) on attribute set A

The algorithm was applied and all three sub-groups were merged after obtaining the above results. This was done while considering the attributes illustrated in Figure 5 (d). The division is made based on the students' having the same characteristics and behaviors. These are then placed in the same sub-group and there are a total of 3 sub-sections. The students were divided into 3 clusters (0,1,2) in the ratio [5:6:12]. Most values are running in a synchronized way, and the values have similarly high and low points.

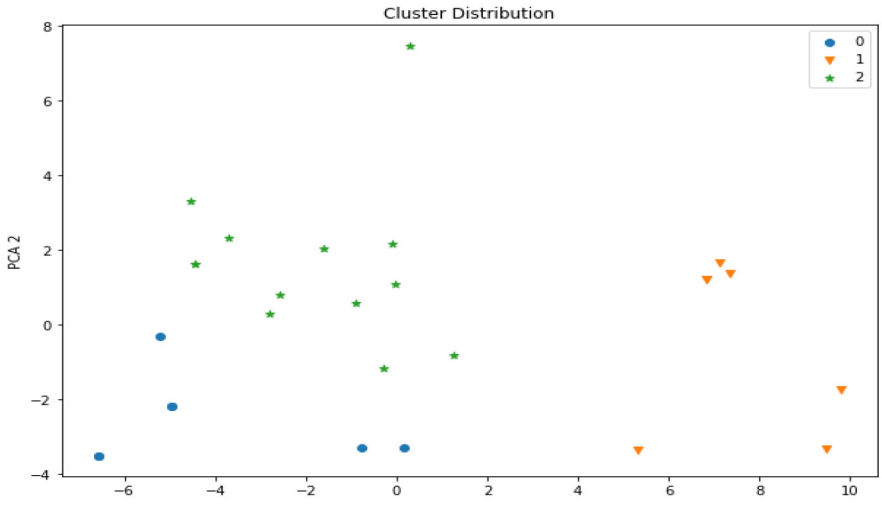


Figure 5 (d): K-means on all three combined

6. Conclusion

The previous studies and research carried out in the field of student behavior understanding and the required advancements were investigated and discussed. For the data collection, an online survey was designed and filled out by students up to class 5, as the proposed work mainly focuses on primary level children. K-means clustering algorithm was applied to the collected data and the sections of pupils were made based on similar characteristics, habits, and behavior. In the future, a new algorithm is being developed that will allow us to divide the student population into sections that are proportionally even as the current algorithm divides the children into an uneven proportion. Further literature review will be enhanced and experiments will be performed for the validation of our work.

Acknowledgment

This work is supported by PMAS-Arid Agriculture University funded project titled "Accurate sensing Information gathering by employing IoT in Agriculture of Pakistan" Project no. PMAS-AAUR/ ORIC/ 2026 Dated 08-03-2021.

References

[1] F. Idris, Z. Hassan, A. Ya'acob, S. K. Gill, and N. A. M. Awal, "The Role of Education in Shaping Youth's National Identity," Procedia - Soc. Behav. Sci., vol. 59, pp. 443–450, Oct. 2012, doi: 10.1016/j.sbspro.2012.09.299.

[2] A. R. Monteiro, "The right of the child to education: What right to what education?," in Procedia - Social and Behavioral Sciences, Jan. 2010, vol. 9, pp. 1988–1992, doi: 10.1016/j.sbspro.2010.12.433.

[3] V. K. Salgong, O. Ngumi, and K. Chege, "The Role of Guidance and Counseling in Enhancing Student Discipline in Secondary Schools in Koibatek District," J. Educ. Pract., vol. 7, no. 13, pp. 142–151, 2016, Accessed: Jun. 21, 2021. [Online]. Available: https://eric.ed.gov/?id=EJ1102862.

[4] K. Van Petegem, A. Aelterman, H. Van Keer, and Y. Rosseel, "The influence of student characteristics and interpersonal teacher behaviour in the classroom on student's wellbeing," Soc. Indic. Res., vol. 85, no. 2, pp. 279–291, Nov. 2007, doi: 10.1007/s11205-007-9093-7.

[5] R. K. Chaffee, A. M. Briesch, R. J. Volpe, A. H. Johnson, and L. Dudley, "Effects of a Class-Wide Positive Peer Reporting Intervention on Middle School Student Behavior:," https://doi.org/10.1177/0198742919881112, vol. 45, no. 4, pp. 224–237, Oct. 2019, doi: 10.1177/0198742919881112.

[6] I. Albluwi, "Using Static Analysis Tools for Analyzing Student Behavior in an Introductory Programming Course," Artic. Jordanian J. Comput. Inf. Technol., vol. 06, no. 03, 2020, doi: 10.5455/jjcit.71-1584234700.

[7] E. Bernaras, J. Jaureguizar, and M. Garaigordobil, "Child and adolescent depression: A review of theories, evaluation instruments, prevention programs, and treatments," Front. Psychol., vol. 10, no. MAR, 2019, doi: 10.3389/FPSYG.2019.00543/FULL.

[8] J. Zong, C. Cui, Y. Ma, L. Yao, M. Chen, and Y. Yin, "Behavior-driven Student Performance Prediction with Tri-branch Convolutional Neural Network," Int. Conf. Inf. Knowl. Manag. Proc., pp. 2353–2356, Oct. 2020, doi: 10.1145/3340531.3412110.

[9] R. K. Orr, P. Caldarella, B. D. Hansen, and H. P. Wills, "Managing Student Behavior in a Middle School Special Education Classroom Using CW-FIT Tier 1," J. Behav. Educ. 2019 291, vol. 29, no. 1, pp. 168–187, Apr. 2019, doi: 10.1007/S10864-019-09325-W.

[10] W. Y. Jiao et al., "Behavioral and Emotional Disorders in Children during the COVID-19 Epidemic," J. Pediatr., vol. 221, pp. 264–266, Jun. 2020, doi: 10.1016/j.jpeds.2020.03.013.

[11] L. Chen, T. Q. Tang, Z. Song, H. J. Huang, and R. Y. Guo, "Child behavior during evacuation under non-emergency situations: Experimental and simulation results," Simul. Model. Pract. Theory, vol. 90, pp. 31–44, Jan. 2019, doi: 10.1016/j.simpat.2018.10.007.

[12] R. Zheng, F. Jiang, and R. Shen, "Intelligent Student Behavior Analysis System for Real Classrooms," in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, May 2020, vol. 2020-May, pp. 9244–9248, doi: 10.1109/ICASSP40776.2020.9053457.

[13] Z. Fan, J. Xu, W. Liu, and W. Cheng, "Gesture based misbehavior detection in online examination," in ICCSE 2016 - 11th International Conference on Computer Science and Education, Oct. 2016, pp. 234–238, doi: 10.1109/ICCSE.2016.7581586.

[14] K. Godwin, V. Almeda, M. Petroccia, R. S. Baker, and A. V Fisher, "Classroom activities and off-task behavior in elementary school children," Proc. Annu. Meet. Cogn. Sci. Soc. Title Classr. Act. off-task Behav. Elem. Sch. Child. Publ. Date 2013 Peer Rev., vol. 35, no. 35, pp. 2428–2433, 2013, Accessed: Jun. 21, 2021. [Online]. Available: https://escholarship.org/content/qt8mx9h5hq/qt8mx9h5hq.pdf.

[15] I. Singh, "A disorder of anger and aggression: Children's perspectives on attention deficit/hyperactivity disorder in the UK," Soc. Sci. Med., vol. 73, no. 6, pp. 889–896, Sep. 2011, doi: 10.1016/j.socscimed.2011.03.049.

[16] S. Sarwar, "Influence of Parenting Style on Children's Behaviour," J. Educ. Educ. Dev., vol. 3, no. 2, Dec. 2016, Accessed: Sep. 12, 2021. [Online]. Available: https://papers.ssrn.com/abstract=2882540.

[17] R. E. Izzaty, "Happiness in early childhood," Psychol. Res. Interv., vol. 1, no. 2, Sep. 2018, doi: 10.21831/pri.v1i2.22024.

[18] N. N. R. Lima et al., "Childhood depression: a systematic review.," Neuropsychiatr. Dis. Treat., vol. 9, pp. 1417–25, Sep. 2013, doi: 10.2147/NDT.S42402.

[19] L. Safaei and M. Youzbashi, "Comparison of the Severity of Obsession and Working Memory in Children with Obsessive Compulsive Disorder and Healthy Children," Int. J. Pediatr., vol. 8, no. 10, pp. 12275–12284, 2020, doi: 10.22038/ijp.2020.50337.4006.

[20] U. Kessels and A. Heyder, "Not stupid, but lazy? Psychological benefits of disruptive classroom behavior from an attributional perspective," Soc. Psychol. Educ., vol. 23, no. 3, pp. 583–613, Jul. 2020, doi: 10.1007/s11218-020-09550-6.

[21] M. C. Levy, W. G. Kronenberger, and B. D. Carter, "Brief Report: Illness Factors and Child Behavior Before and During Pediatric Hospitalization," J. Pediatr. Psychol., vol. 33, no. 8, pp. 905–909, Sep. 2008, doi: 10.1093/JPEPSY/JSN039.

[22] A. D. Faigenbaum, L. A. Milliken, and W. L. Westcott, "Maximal Strength Testing in Healthy Children," Natl. Strength Cond. Assoc. J. Strength Cond. Res, vol. 17, no. 1, pp. 162–166, 2003.

[23] L. Mitchell, "NAUGHTY OR. NEEDY? EXCLUSIONS: A STUDY OF ONE LOCAL EDUCATION AUTHORITY.," 1998. Accessed: Sep. 10, 2021. [Online]. Available: https://etheses.whiterose.ac.uk/4243/1/DX206624.pdf.

[24] M. T. STEIN, A. GRAZIANO, B. HOWARD, and H. DUBOWITZ, "Maria : Stubborn, willful, and always full of energy," J. Dev. Behav. Pediatr., vol. 17, no. 4, 1996.

[25] T. L. Cross, "Nerds and Geeks: Society's Evolving Stereotypes of Our Students with Gifts and Talents:," http://dx.doi.org/10.1177/107621750502800406, vol. 28, no. 4, pp. 26–65, Jul. 2016, doi: 10.1177/107621750502800406.