

Mining Student Behavioral Concern Through Referrals Using K-Means Clustering

Janeth G. Saren^a, Hidear Talirongan^b,

Florence Jean B. Talirongan^c, Charies L. Malicay^d

^a Janethsaren822@gmail.com, ^b hidear@mu.edu.ph, ^c badilles.fj@gmail.com, ^d labanza.c@gmail.com
^{abcd} College of Computer Studies, Misamis University, Ozamiz City, 7200. Philippines

Abstract

Monitoring students' behavior is one the main concerns faced by higher education institutions nowadays. Several procedures were taking into consideration to do the former statement and one of it is through referrals. The research aimed to cluster student behavior based on the referrals using K-Means algorithm. In this paper, students were clustered into two groups according to gender and year level. The result showed that the primary reasons of referrals among the male students were absences, tardiness, poor academic performance and on probation. On the other hand, for the female was misconduct. In terms of the year level of the students, it showed that the primary reasons of referrals among the first and second year students were absences, tardiness, poor academic performance and on probation. While for the third and fourth year students was misconduct.

Keywords: k-means; clustering; students' behavior; referrals; data mining

1. Introduction

Student behavior remains a leading concern for school as it interferes student achievement and the school climate (Blank & Shavit, 2016; Martinez et. al, 2016). It will help students to appreciate learning opportunities to find them meaningful and worthwhile. To establish and maintain desired behavior patterns, students will gain access to reinforcement. Necessary punishment or referrals will be given to the students if they were not able to perform the requirements. Behavioral views reflected in the school community include report card systems, conduct codes, honor rolls and award ceremonies (Solometo, 2018).

School nowadays takes responsibility of managing student behavior which has become a heightened concern for teachers as a result of increased accountability for a student (Gbollie & Keamu, 2017). It has become a usual practice that the students go to education with the aim of achieving a better life through employment. For producing brilliant results, students regularly attend classes attentively. But in the classroom, it has always been observed that some students have been identified for absenteeism, tardiness, poor academic performance and creating disturbances or misconduct and these students could be referred to the Guidance Center for counseling. However, there are reasons why these students get involved for the said demeanor (Kammerer, 2018).

In the quest to conduct data mining on student behavior, it is important for the researcher to probe data relating to behavioral concern. Data mining refers to extracting or "mining" knowledge from large amounts of data (Asanbe et al., 2016). It becomes one of the most important tools used for solving most of the today's problems that are related to different sectors of our life (Zainuddin & Halili, 2016). Recent procedure

practiced in the Misamis University as a private non-sectarian institution of learning, teachers refer students to the Guidance Office on typical reasons that include absences, tardiness, poor academic performance, on probation and misconduct. The teacher will fill up a form that indicates the date of referral, student's information like id number, name, course and year, reasons for referral, subject and teacher's name with signature. For the absences, the instructor will specify the number of absences a student incurred including the date of absent. For the tardiness, the teacher will list the date and times the student is tardy in coming to class. For the poor academic performance, the teacher will write the grade of the student on the given term of examination. For on probation, the teacher usually does referrals for students who are on probation are that student who has the low general weighted average for the first year student during their high school achievement; and the low general weighted average for the continuing students during the previous semester. For misconduct, the teacher refers student especially those who are not following classroom instructions like not wearing their id and other class policies. However, data on student referrals were not utilized to help designing intervention of the university that can address the different reasons for the improvement of student's behavior. Hence, the study will make use of the existing data by conducting data mining on the students' referrals about absences, tardiness, poor academic performance, on probation and misconduct.

The extraction of student abilities and behavior patterns draw conclusions about human learning (Klingler et.al, 2016). The discovered knowledge can be used to better understand students' behavior (Singh et. al, 2016). The main objective of this paper is to use data mining methodology with the implementation of K-means clustering algorithm for the purpose of determining typical student behavioral concern. K-means is a non-hierarchical clustering method that seeks to partition the data into the form of one or more clusters (Ardiada et. al, 2018). Data mining provides many tasks that could be used to study the student referrals. In this research, the clustering task is used to group student reason of referrals like absences, tardiness, poor academic performance, on probation and Misconduct as to gender (Male or Female) and as to year level (first and second-year group as one and third year and fourth for another group). The study is only focused on determining the number of referred cases through clustering common reasons of referrals of the students.

2. Theoretical Framework

2.1. Review of Related Literature

This literature review provides the brief examination of some of the factors that relate to this study.

A study of Juhanak et. al (2019) explored process mining in identifying the different student behavior patterns. Korpershoek and colleagues (2016) applied a meta-analysis approach on the classroom management approaches and programs on students' academic performance, emotional, behavioral and motivational arenas. It revealed the significant contribution of student conduct. Stone (2017) provided positive behavior supports which aims to support student behavior in select situations.

Further, Johnson and company (2019) examined the conceptions of student misbehaviors. It showed that the most frequent and disruptive problem behavior was talking out of turn, followed by inattentiveness, daydreaming, and idleness and the most unacceptable problem behavior was disrespecting teachers regarding disobedience and rudeness, followed by talking out of turn and verbal aggression.

Clustering analysis can automatically divide data into different categories, it neither needs the prior knowledge about the sample nor needs samples training. Therefore, it is usable in a target recognition, data fusion and even in student behavioral patterns (Park et. al, 2016).

However, how to determine the first clustering center of K-means is a difficulty. In the past, the primary clustering center is randomly assigned in K-means clustering algorithm, result in different k values will get different grouping. By the hierarchical clustering, the similar data are gathered together to get the best fit category (multicategory). Thus, the K-means clustering's first center is determined, then K-means clustering could be implemented. Hierarchical clustering calculates k. When K is determined, K-means

clustering algorithm is applied. Improved K-means clustering analysis identifies fault symptom data. Make for the calculation of the fault symptom data more accurately (Syakur et. al, 2018).

The research done by Singh et al. (2016) and Chang et al. (2020) used k-means clustering algorithm based on student attendance and academic performance. The clustering method formulates the performance of the student on clustered groups with small distance among the group members. Another study by Nelson (2015) revealed the usefulness of k-means clustering for better understanding students' Learning Management Systems participation in the 5-cluster solution of course grade and frequency variables. Further, Klingler et al. (2016) illustrated that student similar behavior patterns could be clustered using sequential data. It resulted that the proposed pipeline can detect interesting student behavior and properties of learning environments. One of the most commonly used clustering algorithm to solve problems is k-means (Kavya & Desai, 2016; Yuan & Yang, 2019). This algorithm works on distance measure, as it is going to calculate the distance between each data set and centroid, by distance calculation if that dataset is near to that centroid then that dataset belongs to that cluster. It will loop until there is no change in any dataset. The performance evaluation was carried out with respect to time, using k means both sequentially and parallel on same datasets.

The study conducted by Dever et al. (2016) utilized k-means cluster analysis to determine the subtypes of risk captured by a screening instrument where the final solution produced four groups: Well-Adapted, Internalizing/Adjustment Problems, Mild Externalizing Problems, and General Problems-Severe.

The study of Pikoula and friends (2019) also identified four subgroups of smokers using k-means clustering. It resulted on data verification in the validation set of COPDGene study. Chao (2016) and Phirangee et al. (2016) identified student activity patterns, shared student characteristics and described differences in how those patterns manifest themselves based on the facilitation method used. Clustering analysis was applied to reveal patterns, characteristics, and differences in the students' performance and satisfaction levels toward instructor and students.

The study of Lazuras and company (2017) used K-Means cluster analysis to create a bully-victim subtype to reflect the importance of considering that many students have been victimized and have engaged in the aggressive and bullying behavior.

An exploratory data mining of Fratamico and colleagues (2017) applied a framework for student modeling in exploratory learning environments. It provided adaptive support with great potential to help improve student learning with the different rich activities.

An efficient approximation to the K-means clustering for massive data of Capo and friends (2017) used recursive partitioning. It described the full dataset by this representation, which ultimately leads to a reduction of distance computations (Xia et. al, 2020).

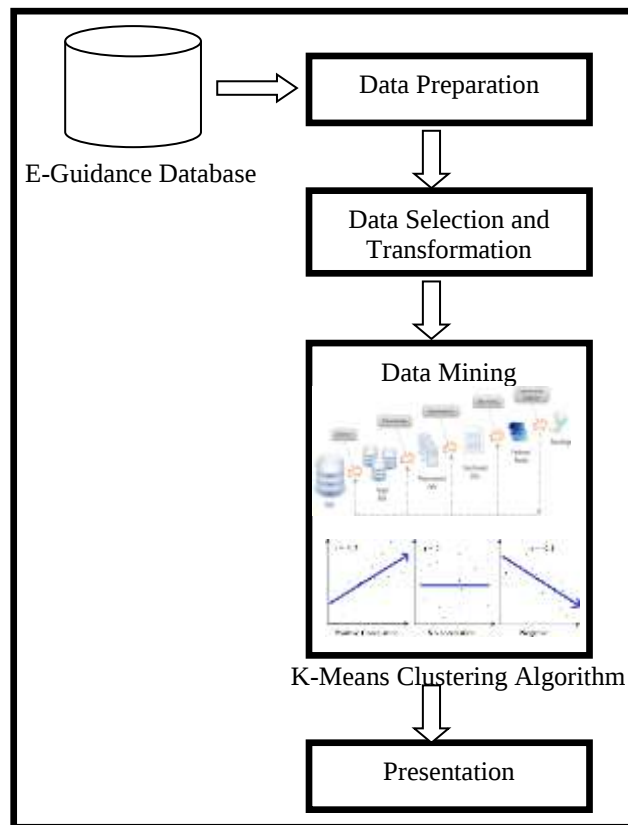
Kia and company (2020) used data mining technique to discover which various sequences of actions characterize high-achieving and low-achieving learners who help institution find those students who are performing well and improve the low-achieving performer.

The same study conducted in the Philippines by Obon and colleagues (2019) of University in Cavite used unsupervised learning in identifying high-performing and low-performing students. It showed that students with the highest number of correct answers experienced flow the most while students with the lowest number of correct answers experienced confusion and boredom the most.

3. Operational Framework

The study utilized the following architectural design which follows several steps that include Data Preparation, Data Selection and Transformation, Data Mining and Presentation. It is presented in Figure 1 for the architectural design of the study.

Figure 1. Architectural Design of the Study



4. Methods

4.1. Data Preparation

In this step, the data was collected from the E-Guidance database which was stored in the Management Information System. The data that was utilized by the researchers covers the 2016-2017 school year referrals. The researcher exported the files to Microsoft Excel and performs Data Filter of selected cells.

4.2. Data Selection and Transformation

After the preparation, the data selection and transformation process was performed. In this step, the researcher performs first the gender clustering by male and female. The researcher counts the number of students who are referred on absences, tardiness, poor academic performance, on probation and misconduct. Another clustering process is the year level where the researcher group first and second year level as one and third and fourth year level as another group. The first group is named junior and the latter is named senior.

4.3. Data Mining

Clustering method partitions data into clusters so that the data having the same characteristics are grouped into one cluster and the data that have different characteristics grouped into another cluster (Arora & Varshney, 2016). In general, K-Means algorithm is as follows:

The first category is a clustering method using the gender (male and female). Table 1 presents the datasets on student referrals in Misamis University, Ozamiz City for the School Year 2016-2017. The first column lists the reasons of referral and the second and third column counts the number of referrals for Male and Female.

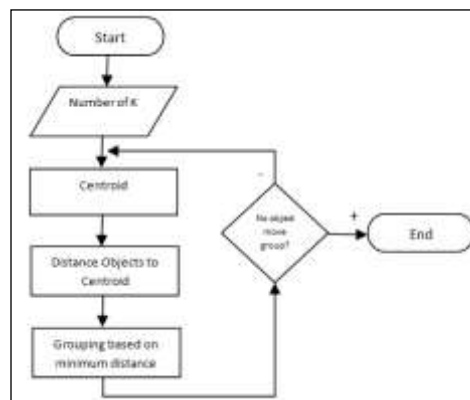
Table 1. Reasons of Referral of Misamis University by Gender

Reasons of Referral	Male	Female
Absences	137	131
Tardiness	63	98
Poor Academic Performance	60	58
On-Probation	45	55
Misconduct	17	8

The K-Means Nearest Neighbor algorithm is a machine-learning instance-based technique. This method does not construct models but stores the training instances. For each new instance, the algorithm compares the distance feature-vectors to the training set. The nearest neighbors are selected based on the distance of the features of the new instance i.e., the similarity between the new instance and the training set vectors. “K” in the algorithm defines the number of nearest neighbors. The classification of the new object is based on the distance between K-clusters and the object. The object is assigned to the cluster with the minimum distance.

The algorithm requires the user to specify the number of clusters k that are desired. When there is no prior knowledge of the number of underlying clusters in the data set, Larose and colleagues (2020) suggest to cycle through various promising values of k and compare clustering solutions for each value of k using some measure of cluster validity as well as domain expertise. K-means because it is simple, fast, and efficient if the number of clusters is known beforehand and the basic steps it follows are: 1. Number of clusters, K , is determined. 2. Assume a centroid or center of the K clusters. Any object can be randomly chosen and initialized as an initial centroid, or the first K objects can also serve as the initial centroids. 3. The distance of each object from each of the centroids is calculated. 4. Group the objects based on minimum distance (find the closest centroid for each object) (Yuan & Yang, 2019). Figure 2 presents the flowchart of k-means Nearest Neighbor algorithm.

Figure 2. K-Means Nearest Neighbor Flowchart



4.4. Presentation

This step presented the results of computing the value of centroids and its distances following several iterations in order to attain similar groups. It also illustrated the results in tables and graphs.

5. Results and Discussions

Monitoring behavioral problems among the students is a primary concern of an institution that contributes to the success of the students. In this study, it showed that the primary reasons of referrals among the male students are absences, tardiness, poor academic performance and on probation. On the other hand, for the female is misconduct. In terms of the year level of the students, it showed that the primary reasons of referrals among the first and second year students are absences, tardiness, poor academic performance and on probation. While for the third and fourth year students is misconduct.

The results of the data mining step are presented in this step.

Step 1: Initial value of centroids: absences and tardiness

Let C_1 and C_2 denote the coordinate of the centroids, then $C_1 = (137, 131)$ and $C_2 = (63, 98)$

Step 2: Objects-Centroids distance denoted as D^0 : The Euclidian distance is used to obtain the distance. The distance matrix at iteration 0 is:

Table 2. Reasons of Referral of Misamis University by Gender

Group	Reasons of Referrals				
	Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct
Male	0	81.0246876	106.1037228	119.3314711	171.8400419
Female	81.024688	0	40.11234224	46.61544808	101.0742301

Each column in the distance matrix symbolizes the reasons of referrals data. The first row corresponds to the distance of each object to the first centroid and the second row is the distance of each object to the second centroid.

The answer for reasons of referrals row (Male) is derived from the formula
 $=\text{SQRT}((\text{raw data of absences(Male)} - C_1)^2 + (\text{raw data of absences(Female)} - C_1)^2)$.

The answer for reasons of referrals row (Female) is derived from the formula
 $=\text{SQRT}((\text{raw data of absences(Male)} - C_2)^2 + (\text{raw data of absences(Female)} - C_2)^2)$.

Step 3: Objects clustering denoted as G^0 : Each reason of referral was assigned based on the minimum distance. Thus, absences is assigned to group 1. On the other hand, tardiness, poor academic performance, on probation and misconduct are assigned to group 2. The element of group matrix below is 1 if and only if the reason is assigned to that group.

Table 3. Objects Clustering: Iteration 0 denoted by G^0 :

Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct	Group
1	0	0	0	0	1
0	1	1	1	1	2

Step 4: Iteration – 1, determine centroids: Compute the new centroid of each group based on the new memberships. Group 1 has 1 member, thus the centroid is: $C_1 = (0, 81.02)$. On the other hand, group 2 has 4 members, hence the centroid is the average coordinated among the four members: $C_2 = (46.95, 119.58)$.

Step 5: Iteration 1, Objects-Centroid distances denoted as D^1 : Repeat the process in step number 2 in order to obtain the new distance matrix based on the new groupings. The results of the process are shown below:

Table 4. Objects-Centroids distances: Iteration 1 denoted D^1 :

Group	Reasons of Referrals				
	Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct
Male	145.83049	65.2469251	64.26613602	51.98350089	74.97736324
Female	90.771375	26.8898882	62.94257371	64.60443192	115.5249285

Step 6: Iteration-1, Objects Clustering denoted as G^1 . Assign each object based on the minimum distance. Based on the new distance matrix, the new group matrix is presented in table 5.

Table 5. Objects Clustering: Iteration 1 denoted by G^1 :

Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct	Group
0	0	0	1	1	1
1	1	1	0	0	2

Step 7: Iteration-2, determine centroids: Compute the new centroid of each group based on the new memberships. Group 1 has 2 members, thus the centroid is the average coordinated among the two members: $C^1 = (63.48, 90.06)$. On the other hand, group 2 has three members, hence the centroid is the average coordinated among the three members: $C^2 = (60.20, 91.78)$.

Step 8: Iteration 2, Objects-Centroid distances denoted as D^2 : Repeat the process in step number 2 in order to obtain the new distance matrix based on the new groupings. The results of the process are shown below:

Table 6. Objects-Centroids distances: Iteration 2 denoted D^2 :

Group	Reasons of Referrals				
	Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct
Male	84.147652	7.94984999	32.5230713	39.63657614	94.31353193
Female	86.233167	6.81956849	33.78178338	39.79867288	94.26365823

Step 9: Iteration-2, Objects Clustering denoted as G^2 :

Assign each object based on the minimum distance. Based on the new distance matrix, the new group matrix is presented below:

Table 7. Objects Clustering: Iteration 2 denoted by G^2 :

Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct	Group
1	0	1	1	0	1
0	1	0	0	1	2

Step 10: Iteration-3, determine centroids: Compute the new centroid of each group based on the new memberships. Group 1 has 3 members; thus the centroid is the average coordinated among the three members. Therefore, $C^1 = (52.01, 53.27)$. On the other hand, group 2 has 2 members, hence the centroid is the average coordinated among the two members: $C^2 = (50.54, 51.13)$.

Step 11: Iteration 3, Objects-Centroid distances denoted as D^3 : Repeat the process in step number 2 in order to obtain the new distance matrix based on the new groupings. The results of the process are shown below:

Table 8. Objects-Centroids distances: Iteration 3 denoted D^3 :

Group	Reasons of Referrals				
	Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct
Male	115.17228	46.0585701	9.282402167	7.222374069	57.23068631
Female	117.70301	48.4958739	11.68908665	6.758201942	54.63865474

Step 12: Iteration-3, Objects Clustering denoted as G^3 :

Assign each object based on the minimum distance. Based on the new distance matrix, the new group matrix is presented below:

Table 9. Objects Clustering: Iteration 3 denoted by G^3 :

Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct	Group
1	1	1	0	0	1
0	0	0	1	1	2

Step 13: Iteration-4, determine centroids: Compute the new centroid of each group based on the new memberships. Group 1 has 3 members, thus the average coordinated among the three members is: $C^1 = (56.84, 59.30)$. On the other hand, group 2 has 2 members, hence the centroid is the average coordinated among the two members: $C^2 = (30.70, 32.23)$.

Step 14: Iteration 4, Objects-Centroid distances denoted as D^4 : Repeat the process in step number 2 in order to obtain the new distance matrix based on the new groupings. The results of the process are shown below:

Table 10. Objects-Centroids distances: Iteration 4 denoted D^4 :

Group	Reasons of Referrals				
	Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct
Male	107.55208	39.1914989	3.417514866	12.5931691	64.94863506
Female	145.10762	73.277151	39.02376004	26.89117437	27.83112833

Step 15: Iteration-4, Objects Clustering denoted as G^4 :

Assign each object based on the minimum distance. Based on the new distance matrix, the new group matrix is presented below:

Table 11. Objects Clustering: Iteration 4 denoted by G^4 :

Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct	Group
1	1	1	1	0	1
0	0	0	0	1	2

Step 16: Iteration-5, determine centroids: Compute the new centroid of each group based on the new memberships. Group 1 has 4 members; thus the centroid is equal to the average coordinated among the four members. Therefore, $C^1 = (40.69, 71.08)$. On the other hand, group 2 has 1 member, hence the centroid is: $C^2 = (27.83, 64.95)$.

Step 17: Iteration 5, Objects-Centroid distances denoted as D^5 : Repeat the process in step number 2 in order to obtain the new distance matrix based on the new groupings. The results of the process are shown below:

Table 12. Objects-Centroids distances: Iteration 5 denoted D^5 :

Group	Reasons of Referrals				
	Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct
Male	113.43231	34.6978699	23.32142568	16.6432064	67.37664736
Female	127.59555	48.2622239	32.91078598	19.84302129	57.96947797

Step 18: Iteration-5, Objects Clustering denoted as G^5 :

Assign each object based on the minimum distance. Based on the new distance matrix, the new group matrix is presented below:

Table 13. Objects Clustering: Iteration 5 denoted by G^5 :

Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct	Group
1	1	1	1	0	1
0	0	0	0	1	2

The results showed that G^4 is equal to G^5 . Comparing the groups of last iteration reveals that the objects does not move group anymore. Thus, the computation of the k-mean clustering has reached its stability and no more iteration is needed. The final grouping is presented in table 14 below.

Table 14. Clustered Reasons of Referral by Gender:

Reasons of Referrals of Misamis University by Gender			
Reasons of Referral	Male	Female	Group
Absences	137	131	1
Tardiness	63	98	1
Poor Academic Performance	60	58	1
On-Probation	45	55	1
Misconduct	17	8	2

The second category is a clustering method using the year level. first and second year level are grouped as one and named as junior while third and fourth year level are another group and named as Senior. Table 15 presents the datasets on student referrals in Misamis University, Ozamiz City for the School Year 2016-2017. The first column lists the reasons of referral and the second and third column counts the number of referrals for junior and senior year level.

Table 15. Reasons of Referral of Misamis University by Grouped Year Level

Reasons of Referral	Junior	Senior
Absences	146	122
Tardiness	82	79
Poor Academic Performance	53	65
On-Probation	62	38
Misconduct	13	12

Step 1: Initial value of centroids: Absences and Tardiness

Let C_1 and C_2 denote the coordinate of the centroids, then $C_1 = (146, 122)$ and $C_2 = (82, 79)$

Step 2: Objects-Centroids distance denoted as D^0 : The Euclidian distance is used to obtain the distance. The distance matrix at iteration 0 is:

Table 16. Objects-Centroids distances: Iteration 0 denoted D^0 :

Group	Reasons of Referrals				
	Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct
Junior	0	77.1038261	109.0779538	118.7939392	172.5949014
Senior	77.1038261	0	32.20248438	45.61797891	96.17692031

Each column in the distance matrix symbolizes the reasons of referrals data. The first row corresponds to the distance of each object to the first centroid and the second row is the distance of each object to the second centroid.

The answer for reasons of referrals row (Junior) is derived from the formula

$$= \text{SQRT}((\text{raw data of absences}(\text{Junior}) - C_1)^2 + (\text{raw data of absences}(\text{Senior}) - C_1)^2).$$

The answer for reasons of referrals row (Senior) is derived from the formula

$$= \text{SQRT}((\text{raw data of absences}(\text{Junior}) - C_2)^2 + (\text{raw data of absences}(\text{Senior}) - C_2)^2).$$

Step 3: Objects clustering denoted as G^0 : Each reason of referral was assigned based on the minimum distance. Thus, absences is assigned to group 1. On the other hand, tardiness, poor academic performance, on probation and misconduct are assigned to group 2. The element of group matrix below is 1 if and only if the reason is assigned to that group.

Table 17. Objects Clustering: Iteration 0 denoted by G^0 :

Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct	Group
1	0	0	0	0	1
0	1	1	1	1	2

Step 4: Iteration – 1, determine centroids: Compute the new centroid of each group based on the new memberships. Group 1 has 1 member; thus the centroid is equal to 0 since it has only one member. Therefore, $C_1 = (0, 77.10)$. On the other hand, group 2 has 4 members, hence the centroid is the average coordinated among the four members: $C_2 = (43.50, 119.39)$.

Step 5: Iteration 1, Objects-Centroid distances denoted as D^1 : Repeat the process in step number 2 in order to obtain the new distance matrix based on the new groupings. The results of the process are shown below:

Table 18. Objects-Centroids distances: Iteration 1 denoted D^1 :

Group	Reasons of Referrals				
	Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct
Junior	152.74707	82.0219207	54.36453445	73.30149532	66.38906667
Senior	102.53381	55.8020336	55.21615127	83.46878765	111.639565

Step 6: Iteration-1, Objects Clustering denoted as G^1 . Assign each object based on the minimum distance. Based on the new distance matrix, the new group matrix is presented in table 19.

Table 19. Objects Clustering: Iteration 1 denoted by G^1 :

Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct	Group
0	0	1	1	1	1
1	1	0	0	0	2

Step 7: Iteration-2, determine centroids: Compute the new centroid of each group based on the new memberships. Group 1 has 3 members, thus the centroid is the average coordinated among the three members: Therefore, $C^1 = (64.69, 83.44)$. On the other hand, group 2 has 2 members, hence the centroid is the average coordinated among the two members: $C^2 = (79.17, 117.38)$.

Step 8: Iteration 2, Objects-Centroid distances denoted as D^2 : Repeat the process in step number 2 in order to obtain the new distance matrix based on the new groupings. The results of the process are shown below:

Table 20. Objects-Centroids distances: Iteration 2 denoted D^2 :

Group	Reasons of Referrals				
	Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct
Junior	89.993788	17.8755432	21.83183334	45.52075834	88.17726838
Senior	66.991265	38.4888299	58.55677022	81.21967374	124.4350652

Step 9: Iteration-2, Objects Clustering denoted as G^2 :

Assign each object based on the minimum distance. Based on the new distance matrix, the new group matrix is presented below:

Table 21. Objects Clustering: Iteration 2 denoted by G^2 :

Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct	Group
0	1	1	1	1	1
1	0	0	0	0	2

Step 10: Iteration-3, determine centroids: Compute the new centroid of each group based on the new memberships. Group 1 has 4 members; thus the centroid is the average coordinated among the four members. Therefore, $C^1 = (43.35, 75.67)$. On the other hand, group 1 has 1 member, hence the centroid is: $C^2 = (66.99, 89.99)$.

Step 11: Iteration 3, Objects-Centroid distances denoted as D^3 : Repeat the process in step number 2 in order to obtain the new distance matrix based on the new groupings. The results of the process are shown below:

Table 22. Objects-Centroids distances: Iteration 3 denoted D^3 :

Group	Reasons of Referrals				
	Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct
Junior	112.61768	38.7914056	14.38936641	42.0378892	70.53879015
Senior	85.245398	18.6044486	28.64340946	52.23281227	94.85824983

Step 12: Iteration-3, Objects Clustering denoted as G^3 :

Assign each object based on the minimum distance. Based on the new distance matrix, the new group matrix is presented below:

Table 23. Objects Clustering: Iteration 3 denoted by G^3 :

Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct	Group
0	0	1	1	1	1
1	1	0	0	0	2

Step 13: Iteration-4, determine centroids: Compute the new centroid of each group based on the new memberships. Group 1 has 3 members; thus the centroid is the average coordinated among the three members. Therefore, $C^1 = (42.32, 58.58)$. On the other hand, group 2 has 2 members, hence the centroid is the average coordinated among the two members: $C^2 = (51.92, 75.70)$.

Step 14: Iteration 4, Objects-Centroid distances denoted as D^4 : Repeat the process in step number 2 in order to obtain the new distance matrix based on the new groupings. The results of the process are shown below:

Table 24. Objects-Centroids distances: Iteration 4 denoted D^4 :

Group	Reasons of Referrals				
	Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct
Junior	121.53787	44.6250394	12.46031393	28.47250668	55.03912522
Senior	104.84936	30.2550865	10.75839571	39.02742497	74.65533297

Step 15: Iteration-4, Objects Clustering denoted as G^4 :

Assign each object based on the minimum distance. Based on the new distance matrix, the new group matrix is presented below:

Table 25. Objects Clustering: Iteration 4 denoted by G^4 :

Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct	Group
0	0	0	1	1	1
1	1	1	0	0	2

Step 16: Iteration-5, determine centroids: Compute the new centroid of each group based on the new memberships. Group 1 has 2 members; thus the centroid is the average coordinated among the two members. Therefore, $C^1 = (41.75, 56.84)$. On the other hand, group 2 has 3 members, hence the centroid is the average coordinated among the three members: $C^2 = (48.62, 59.54)$.

Step 17: Iteration 5, Objects-Centroid distances denoted as D^5 : Repeat the process in step number 2 in order to obtain the new distance matrix based on the new groupings. The results of the process are shown below:

Table 26. Objects-Centroids distances: Iteration 5 denoted D^5 :

Group	Reasons of Referrals				
	Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct
Junior	122.93289	45.9412542	13.89225583	27.65546147	53.26956184
Senior	115.68836	38.6369098	6.998281611	25.35777883	59.40543627

Step 18: Iteration-5, Objects Clustering denoted as G^5 :

Assign each object based on the minimum distance. Based on the new distance matrix, the new group matrix is presented below:

Table 27. Objects Clustering: Iteration 5 denoted by G^5 :

Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct	Group
0	0	0	0	1	1
1	1	1	1	0	2

Step 19: Iteration-6, determine centroids: Compute the new centroid of each group based on the new memberships. Group 1 has 1 member, thus the centroid is: $C^1 = (53.27, 59.40)$. On the other hand, group 2 has 4 members, hence the centroid is the average coordinated among the four members: $C^2 = (46.67, 52.60)$.

Step 20: Iteration 6, Objects-Centroid distances denoted as D^5 : Repeat the process in step number 2 in order to obtain the new distance matrix based on the new groupings. The results of the process are shown below:

Table 28. Objects-Centroids distances: Iteration 6 denoted D^6 :

Group	Reasons of Referrals				
	Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct
Junior	111.87946	34.7762132	5.601054091	23.11737988	62.2005868
Senior	121.16924	44.1005309	13.91722553	21.17352889	52.74936169

Step 21: Iteration-6, Objects Clustering denoted as G^6 :

Assign each object based on the minimum distance. Based on the new distance matrix, the new group matrix is presented below:

Table 29. Objects Clustering: Iteration 6 denoted by G^6 :

Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct	Group
1	1	1	0	0	1
0	0	0	1	1	2

Step 22: Iteration-7, determine centroids: Compute the new centroid of each group based on the new memberships. Group 1 has 3 members; thus the centroid is the average coordinated among the three members. Therefore, $C^1 = (50.75, 59.73)$. On the other hand, group 2 has 2 members, hence the centroid is the average coordinated among the two members: $C^2 = (36.96, 42.66)$.

Step 23: Iteration 7, Objects-Centroid distances denoted as D^5 : Repeat the process in step number 2 in order to obtain the new distance matrix based on the new groupings. The results of the process are shown below:

Table 30. Objects-Centroids distances: Iteration 7 denoted D^7 :

Group	Reasons of Referrals				
	Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct
Junior	113.79725	36.7123122	5.730259652	24.46755825	60.85465523
Senior	134.84956	57.8717626	27.50193198	25.46832047	38.91174784

Step 24: Iteration-7, Objects Clustering denoted as G^7 :

Assign each object based on the minimum distance. Based on the new distance matrix, the new group matrix is presented below:

Table 31. Objects Clustering: Iteration 7 denoted by G^7 :

Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct	Group
1	1	1	1	0	1
0	0	0	0	1	2

Step 25: Iteration-8, determine centroids: Compute the new centroid of each group based on the new memberships. Group 1 has 4 members; thus the centroid is the average coordinated among the four members. Therefore, $C^1 = (45.18, 61.42)$. On the other hand, group 2 has 1 member, hence the centroid is: $C^2 = (38.91, 60.85)$.

Step 26: Iteration 8, Objects-Centroid distances denoted as D^5 : Repeat the process in step number 2 in order to obtain the new distance matrix based on the new groupings. The results of the process are shown below:

Table 32. Objects-Centroids distances: Iteration 7 denoted D^7 :

Group	Reasons of Referrals				
	Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct
Junior	117.62183	40.8031802	8.602177919	28.83835145	58.97433052
Senior	123.31523	46.7530856	14.68545989	32.48696129	55.30095852

Step 27: Iteration-8, Objects Clustering denoted as G^8 :

Assign each object based on the minimum distance. Based on the new distance matrix, the new group matrix is presented below:

Table 33. Objects Clustering: Iteration 8 denoted by G^8 :

Absences	Tardiness	Poor Academic Performance	On-Probation	Misconduct	Group
1	1	1	1	0	1
0	0	0	0	1	2

The results showed that G^7 is equal to G^8 . Comparing the groups of last iteration reveals that the objects does not move group anymore. Thus, the computation of the k-mean clustering has reached its stability and no more iteration is needed. The final grouping is presented in table 34 below.

Table 34. Clustered Reasons of Referral by Grouped Year Level:

Reasons of Referrals of Misamis University by Grouped Year Level			
Reasons of Referral	Male	Female	Group
Absences	146	122	1
Tardiness	82	79	1
Poor Academic Performance	53	65	1
On-Probation	62	38	1
Misconduct	13	12	2

To give the summary of the two categories clustering, Figure 3 presents the scattered plot of clustering by gender and Figure 4 presents the scattered plot of clustering by year level. It showed that the four highest reasons of referrals among the male students are absences, tardiness, poor academic performance and on probation which is the Group 1. For the Group 2, it only consists one reason for female which is misconduct which is the Group 2. In terms of the year level of the students, Group 1 showed that the four highest reasons of referrals among the first and second year students are absences, tardiness, poor academic performance and on probation. While for the Group 2 is misconduct for the third and fourth year level. The two categories yield the same result where Group 1 has the highest number of referrals which include absences, tardiness, poor academic performance and on probation. The second group only consist one reason of referral which is misconduct.

Figure 3. Scattered plot by Gender

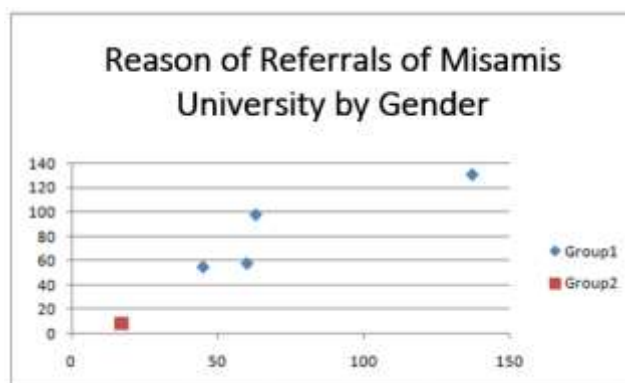
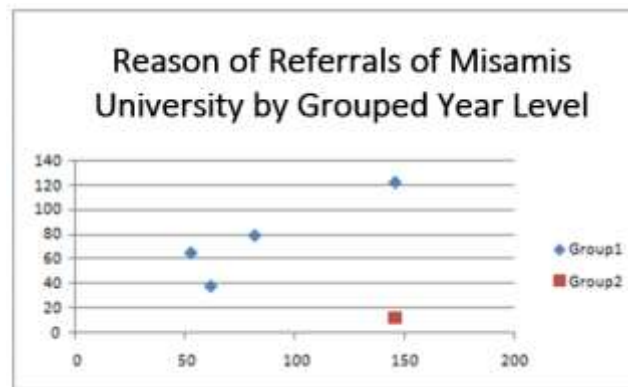


Figure 4. Scattered plot by Year Level



6. Conclusions and Recommendations

The study on determining the number of referred cases through clustering typical reasons of referrals of the students of Misamis University into two categories: gender and year level. The reasons include absences, tardiness, poor academic performance and on probation. The results showed that the primary reasons of referrals among the male students are absences, tardiness, poor academic performance and on probation. On the other hand, for the female is misconduct. In terms of the year level of the students, it showed that the primary reasons of referrals among the first and second year students are absences, tardiness, poor academic performance and on probation. While for the third and fourth year students is misconduct.

For a future work, the result of this research can be used as a reference to study on designing preventive measures of the different reasons of referrals. Furthermore, it is also recommended to use other clustering methods aside from the k-means clustering.

References

- Ardiada, I. M. D., Ariawan, M. P. A., & Sudarma, M. (2018). Data Mining, Evaluation, K-means Evaluation of Supporting Work Quality Using K-Means Algorithm. *International Journal of Engineering and Emerging Technology*, 3(1), 52-55.
- Arora, P., & Varshney, S. (2016). Analysis of k-means and k-medoids algorithm for big data. *Procedia Computer Science*, 78, 507-512.
- Asanbe, M. O., Osofisan, A. O., & William, W. F. (2016). Teachers' Performance Evaluation in Higher Educational Institution using Data Mining Technique. *International Journals of Applied Information System (IJAIS) Volume*, 10, 10-15.
- Blank, C., & Shavit, Y. (2016). The association between student reports of classmates' disruptive behavior and student achievement. *Aera Open*, 2(3), 2332858416653921.
- Capó, M., Pérez, A., & Lozano, J. A. (2017). An efficient approximation to the K-means clustering for massive data. *Knowledge-Based Systems*, 117, 56-69.
- Chang, W., Ji, X., Liu, Y., Xiao, Y., Chen, B., Liu, H., & Zhou, S. (2020). Analysis of University Students' Behavior Based on a Fusion K-Means Clustering Algorithm. *Applied Sciences*, 10(18), 6566.
- Chao, P. Y. (2016). Exploring students' computational practice, design and performance of problem-solving through a visual programming environment. *Computers & Education*, 95, 202-215.
- Dever, B. V., Gallagher, E. K., Hochbein, C. D., Loukas, A., & Dai, C. (2016). Examining Subtypes of Behavioral/Emotional Risk Using Cluster Analysis. *Journal of Psychoeducational Assessment*, 0734282916657646.
- Fratamico, L., Conati, C., Kardan, S., & Roll, I. (2017). Applying a framework for student modeling in exploratory learning environments: Comparing data representation granularity to handle environment complexity. *International Journal of Artificial*

- Intelligence in Education*, 27(2), 320-352.
- Gbolliie, C., & Keamu, H. P. (2017). Student academic performance: The role of motivation, strategies, and perceived factors hindering Liberian junior and senior high school students learning. *Education Research International*, 2017.
- Johnson, Z. D., Goldman, Z. W., & Claus, C. J. (2019). Why do students misbehave? An initial examination of antecedents to student misbehavior. *Communication Quarterly*, 67(1), 1-20.
- Juhaňák, L., Zounek, J., & Rohlíková, L. (2019). Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Computers in Human Behavior*, 92, 496-506.
- Kammerer, E. F. (2018). Undergraduate Moot Court: Student Expectations and Perspectives. *PS: Political Science & Politics*, 51(1), 190-193.
- Kavya, D. S., & Desai, C. D. (2016). Comparative Analysis of K means Clustering Sequentially And Parallely. *nature*, 3(04).
- Kia, F. S., Teasley, S. D., Hatala, M., Karabenick, S. A., & Kay, M. (2020, March). How patterns of students dashboard use are related to their achievement and self-regulatory engagement. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 340-349).
- Klingler, S., Käser, T., Solenthaler, B., & Gross, M. (2016). Temporally Coherent Clustering of Student Data. In *Proceedings of the 9th International Conference on Educational Data Mining, International Educational Data Mining Society* (pp. 102-109).
- Korpershoek, H., Harms, T., de Boer, H., van Kuijk, M., & Doolaard, S. (2016). A meta-analysis of the effects of classroom management strategies and classroom management programs on students' academic, behavioral, emotional, and motivational outcomes. *Review of Educational Research*, 86(3), 643-680.
- Larose, H., Prokoph, N., Matthews, J. D., Schleder, M., Högl, S., Alsulami, A. F., ... & Turner, S. D. (2020). Whole Exome Sequencing reveals NOTCH1 mutations in anaplastic large cell lymphoma and points to Notch both as a key pathway and a potential therapeutic target. *Haematologica*.
- Lazuras, L., Barkoukis, V., & Tsohatzoudis, H. (2017). Face-to-face bullying and cyberbullying in adolescents: Trans-contextual effects and role overlap. *Technology in Society*, 48, 97-101.
- Martinez, A., McMahon, S. D., Coker, C., & Keys, C. B. (2016). Teacher behavioral practices: Relations to student risk behaviors, learning barriers, and school climate. *Psychology in the Schools*, 53(8), 817-830.
- Nelson, K. (2015). Using k-means clustering to model students' participation in traditional courses. *Issues in Information Systems*, 16(4).
- Park, Y., Yu, J. H., & Jo, I. H. (2016). Clustering blended learning courses by online behavior data: A case study in a Korean higher education institute. *The Internet and Higher Education*, 29, 1-11.
- Phirangee, K., Demmans Epp, C., & Hewitt, J. (2016). Exploring the relationships between facilitation methods, students' sense of community and their online behaviours. *Special Issue on Online Learning Analytics. Online Learning Journal*, 20(2), 134-154.
- Pikoula, M., Quint, J. K., Nissen, F., Hemingway, H., Smeeth, L., & Denaxas, S. (2019). Identifying clinically important COPD subtypes using data-driven approaches in primary care population based electronic health records. *BMC medical informatics and decision making*, 19(1), 1-14.
- Singh, M., Nagar, H., & Sant, A. (2016). K-mean and EM Clustering algorithm using attendance performance improvement Primary school Student.
- Solometo, T. E. (2018). *Elementary Parent Reactions to the Motivation-and Goalsetting-Related Aspects of the Standards-Based Report Card*. Gwynedd Mercy University
- Stone, B. (2017). Positive Behavior Support Strategies for Classroom Teachers and Paraeducators.
- Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018, April). Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In *IOP Conference Series: Materials Science and Engineering* (Vol. 336, No. 1, p. 012017). IOP Publishing.
- Xia, S., Peng, D., Meng, D., Zhang, C., Wang, G., Giem, E., ... & Chen, Z. (2020). A fast adaptive k-means with no bounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Yuan, C., & Yang, H. (2019). Research on K-value selection method of K-means clustering algorithm. *J—Multidisciplinary Scientific Journal*, 2(2), 226-235.
- Zainuddin, Z., & Halili, S. H. (2016). Flipped classroom research and trends from different fields of study. *International review of research in open and distributed learning*, 17(3), 313-340.