



Grouping of Student's Appeal System using K-Means Algorithm: A Study

Mary Ann C. Paguio, Dr. Menchita F. Dumlao, Dr. Shaneth C. Ambat,
Dr. Ma. Amelia E. Damian, Dr. Dennis B. Gonzales
AMA Computer University, Philippines

Abstract—Clustering Technique is a way of grouping the related data set or any set of objects into groups into nearest similar related group. It is a technique that enables to find specific discriminative factors or attributes for the studied data. Each member of a cluster should be very similar to other members in its cluster and very dissimilar to other clusters. When a new data is introduced, it is classified into the most similar clusters. In order to show the techniques for creating clusters k-means algorithm will used. K- Mean's algorithm mainly used to partition the clusters based on their means. The aimed of this study is to show how K-Means algorithm applied in the Student's Appeal System and analyse the requirements needed in clustering applicable for grouping the student's appeal in terms of re-sit, make-up and appeal against exam result. In addition, this study aimed to provide the web-enhanced clustering solution for student's appeal in HCT-IT Department using K-Means Algorithm. It provides grouping of student's data into re-sit, makeup and against exam result appeal by clustering technique. In addition, this study helps a lot in projecting the number who are eligible in applying student's appeal easily, shorten the time in processing the appeal, increase the level of readiness in preparing the exam questionnaires ahead of time, projects the number of sections and subject to be open for the next semester, serves as the ground in teaching performance because of the number of instances that a student filed an appeal, lessen the Appeal Members to wait for a long queue to get the initial approval from the Head of Department, monitoring the progress of appeal process and provides on-time results.

Keywords— Clustering Technique, K-Means Algorithm, Student's Appeal

I. INTRODUCTION

Nowadays, processing of appeal is becoming a tiring and tedious in any institution. The Appeal Committee has to effectively meet the challenges of processing and checking each student's appeal. Due to voluminous number of students applying an appeal, creation of system for student's appeal is needed. Colleges or universities is the important learning environment for the students to form their sense of well-being, and safety in terms of complaints, problem, issues or concern related to the involvement in the college or university. In order to change such things the student can file a complaint or an appeal. Academic Appeal is a procedure which allows the students in certain circumstances to ask for a review of a decision relating to the grades or academic progress. It comes as a formal request brought by the student to change the grade, to re-sit, make-up, file complaint and other issues of standards of academic industry. Students have the right to receive the clear explanation with valid proof to ensure efficient and orderly investigation is done on their part. In order to address the grouping of student's appeal, clustering technique is used. This technique is used to divide the data into groups that are similar and the main goal is to divide the data sets into a separate datasets (group) and put together of all data sets that has similar to one another (Chun-Wei Sai, 2009). Developing the system that will automatically clustered the student data into re-sit, make-up and appeal against exam is a great help in this study especially to the Appeal Committee.

II. OVERVIEW OF THE STUDY

Clustering of an appeal is a way of preserving knowledge in any Higher Educational Institution (HEI) particularly in Gulf countries is very essential. It gives the great impact as the universities discovered the knowledge from their resources and becomes an organizational intellectual asset. Estimation on the number of students who will apply for the appeal on the next semester is a great prospective of preserving the knowledge. It becomes intellectual capital assets of its colleges or universities. The university offers various programs in a field of *Engineering, Applied Sciences, Business Studies, Pharmacy and Information Technology*. The main focus of the study is mainly for the department of IT (Information Technology) that offers the following specialization which includes: *Software Engineering, Database, Internet & Security, Networking and Information Systems*. The students can only file an appeal for their enrolled courses in the previous semester on the first day of teaching of the next semester. Usually the period of announcing where the student can file an appeal is three working days from the date of announcements of results. The types of appeals are: appeal against the result, makeup and re-sit.

III. LITERATURE REVIEW

According to the study of Sharma, R. et. al., “K-means is a simple clustering techniques used to analyze the datasets that aims to find the best division of an entities into k groups (called clusters), so that total distance between the group’s members and corresponding centroid, irrespective of the group is minimized. It results into a partitioning of the data space into Voronoi Cells. The main focus of the study is to analyse the rice data set of a crop fields records from Industrial Statistics of India using Weka Tool. In addition, K-means techniques, clustered the data with the result of finding the rate of production of rice crop in consecutive years and reason why there harvesting of rice comes in a high or low of production.” Sharma, R. et. al.,(2012). In addition, another study entitled “*A Review of Density-Based Clustering in Spatial Data*”, that spatial mining is a branch of data mining that refers to the location or geo-referenced data. The study refers to the density based clustering that discusses the area of any given data. It represents the data records that mapped into grid structure and group it according to the grid based on density. It uses CluStream framework that deals in online and offline phase for clustering. Online phase deals with the use of an algorithm that records the summary of information about the data records and performs clustering on synopsis information on offline phase. Shrivastava, et. al. (2012). Similarly, according to the study of Jain, et. al.,(2010), in their study entitled as ‘*K- Means Clustering Using Weka Interface*’, that “K-means clustering techniques using WEKA 3.7 provides conversion of datasets into standard format of spreadsheets. The sample data is converted into csv format, where data sets of fields are initially been removed and the children fields are converted into groups categorization. The result is customer segments where the distribution of males and female in each cluster is presented. Each data presented will depends on the dataset assigned in WEKA Open source software that supports various clustering algorithms execution for data mining research process. Jain, et. al.,(2010). Consequently, the study of Sharma,et. al.,(2012) entitled ‘*Comparison the various clustering algorithms of Weka Tools,*’ discusses that “Cluster analysis or clustering is the task of assigning a set of objects into groups (called clusters) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. The main focus of this study is to provide detailed introduction of weka clustering algorithms that classify the data various types. In order to perform the clustering the promise data repository was used in projecting the data for analysis. In addition, with regards to the behaviour of the data, the k-means clustering algorithm is a simplest algorithm that uses weka as most suitable tool for data mining applications. Seemingly, according to the study of Chitraa,et.al.(2012), entitled ‘*An Enhanced Clustering Technique for Web Usage Mining*’, discusses the use of navigational browsing on the web and personalized recommendations of web pages according to the current interest of the users. It used to determine the web patterns through the use of web mining. It is a technique that automatically retrieve, extract and evaluate information for knowledge discovery from web documents and services. Clustering analysis is used in this study partitioning a set of data objects into a number of object clusters, where each data object shares the high similarity with the other objects within the same cluster but is quite dissimilar to objects in other clusters”.

IV. APPLICATION OF K-MEANS ALGORITHM IN STUDENT’S APPEAL DOMAIN

Hence, to apply the clustering technique, the common is K-Means Algorithm. Along with in the study of Akila, (2012), “K-Means algorithm is used to cluster the relevant web pages in order to increase the relevance rate of search results and reduce the computational time of the user. The objective is to partitions the input dataset into k clusters. Each cluster is represented by an adaptively changing centroid (also called cluster centre), starting from some initial values named seed-points. K-Means computes the distances between the inputs (also called input data points) and centroids, and assigns inputs to the nearest centroid”. Appeal Committee has a total of 5 (five) members acting as an Assessment Committee (AC) on each of the specialization namely: *Information System, Database, Software Engineering, Internet & Security and Networking* and has 3 levels: *Diploma, Higher Diploma and Bachelor Level*. Projection from large dataset of student’s appeal in the number of appeal cases to whom the students are expected to apply is a great potential and helpful to the college or university for the decision making. If the students are eligible for re-sit or make-up exam, it is very easy for the Appeal Committee to count the number of ratio of receiving and monitoring the appeal that soon they will prepare and analyse. The main objective of clustering is to determine the basic grouping of data in any given sets. It pointed on the study of Seddawy, et.al.(2012), that “clustering involves identifying a finite set of categories or segments to describe the data according to a certain metric. The clusters can be mutually exclusive, hierarchical or overlapping”. In addition, clustering optimizes a qualification metric during its learning process and considers all the features of datasets that are treated equally, which means each feature of datasets participated in clustering process equivalently (Hamid,P. 2015). To show, how the clustering will occur, the application of K-Means algorithm is used. K-Means algorithm mainly used to partition the clusters based on their means. Apparently, the author named Seddawy, et.al.(2012), define the clustering as “a technique that enables to find specific discriminative factors or attributes for the studied data. Each member of a cluster should be very similar to other members in its cluster and very dissimilar to other clusters. When a new data is introduced, it is classified into the most similar clusters. Techniques for creating clusters include partitioning methods as in k-means algorithm, and hierarchical methods as decision trees, and density-based methods”. Consequently, “K-Means Algorithm is a simplest technique to illustrate the clustering problem. It builds to classify or grouping the objects based in “k” in a group. K is positive integer, and grouping is done by mining the sum of squares of distance between data and corresponding cluster known as centroid. The groups of centroid refer as the average point in the multidimensional space.” (Seddawy, et.al.,2012). According to the study of Seddawy et.al.(2012), below are the commonly four steps in K-Means:

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centers.
2. Assign each object to the group that has the closest centered.
3. When all objects have been assigned, recalculate the positions of the K centered.
4. Repeat Steps 2 and 3 until the centered no longer change.

Fig. 1 – Steps in K-Means Algorithm

It produces a separation of the objects into groups from which the metric to be minimized can be calculated as shown below:

$$\sum_{k=1}^k \sum_{j=1}^{n_j} \|x_j - c_j\|^2$$

k: number of clusters
 nj: number of points in jth cluster
 xij: ith point in jth cluster

Fig. 2 - Formula metric in Separating of Objects - Seddawy et.al.(2012)

In addition, below is the procedures code for K-means Algorithm:

```

Input:
D = {t1, t2 ... tn} // Set of Elements
A // Adjacency matrix showing distance between elements.
K // Number of desired clusters.

Output:
K // Set of clusters.

K-Means Algorithm:
    Assign initial values for means m1, m1..... mk;
    Repeat
        Assign each item ti to the cluster which has the closest
        mean;
        Calculate new mean for each cluster;
    Until convergence criteria is met;
    
```

Fig.3 - Procedures Code for K-Means Algorithm - Seddawy et.al.(2012)

V. RESULTS AND OUTCOME

The main objective is to analyze the requirements needed in clustering technique applicable for grouping the student’s appeal in terms of re-sit, make-up and appeal against exam result. Clustered data from a datasets of student information, student transcript and curriculum is a knowledge discovery that provides the automatic projection in the number of students enrolled in the previous semester candidates to take re-sit appeal, make-up appeal and appeal against exam result. Apparently, to show the clustered data, the researchers used the K-means algorithm as a way to cluster each dataset from data enrolled last semester (previous semester). Below are the fig.4, fig. 5 and fig. 6 as shown below that serve as inputted values in order that clustered data will group accordingly as addressing to re-sit, make-up and appeal against exam result.

stud_fullname	stud_gen	stud_dob	stud_course	stud_dept	stud_advisor	stud_spec	Level	stud_probation	stud_email	stud_level
XXXXXX	Male	12/6/1999	IS	Information Technology	XXXXXX	IS	Diploma First Year	Active	XXXXXX@stu.hct.edu.om	1
XXXXXX	Male	12/6/1999	IS	Information Technology	XXXXXX	IS	Diploma First Year	Active	XXXXXX@stu.hct.edu.om	1
XXXXXX	Male	12/6/1999	IS	Information Technology	XXXXXX	IS	Diploma First Year	Active	XXXXXX@stu.hct.edu.om	2
XXXXXX	Male	12/6/1999	IS	Information Technology	XXXXXX	IS	Diploma First Year	Active	XXXXXX@stu.hct.edu.om	2
XXXXXX	Male	12/6/1999	IS	Information Technology	XXXXXX	IS	Diploma First Year	Active	XXXXXX@stu.hct.edu.om	3
XXXXXX	Male	12/6/1999	IS	Information Technology	XXXXXX	IS	Diploma First Year	Active	XXXXXX@stu.hct.edu.om	3
XXXXXX	Male	12/6/1999	IS	Information Technology	XXXXXX	IS	Diploma First Year	Active	XXXXXX@stu.hct.edu.om	4
XXXXXX	Male	12/6/1999	IS	Information Technology	XXXXXX	IS	Diploma First Year	Active	XXXXXX@stu.hct.edu.om	4
XXXXXX	Male	12/6/1999	IS	Information Technology	XXXXXX	IS	Diploma First Year	Active	XXXXXX@stu.hct.edu.om	4

Fig.4 – Sample Student Information

stud_id	rec_ay	rec_sem	rec_course_code	rec_course_title	rec_credit	rec_perce	rec_grade	rec_grade_pts	rec_note	rec_result
XXXXXX	2014-2015	1	XXXX	XXXX	3	83	B+	3.3		P
XXXXXX	2014-2015	1	XXXX	XXXX	0	69	C	2		P
XXXXXX	2014-2015	1	XXXX	XXXX	3	92	A	4		P
XXXXXX	2014-2015	1	XXXX	XXXX	3	77	B	3		P
XXXXXX	2014-2015	1	XXXX	XXXX	3	74	B-	2.7		P
XXXXXX	2014-2015	2	XXXX	XXXX	3	95	A	4		P

Fig.5 – Sample Student Transcript

Specialization	Level	Semester	Course	Description	Type	Credit Hrs
IT	Diploma First Year	1	XXX	XXX	CR	3
IT	Diploma First Year	1	XXX	XXX	Foundation Requirement	0
IT	Diploma First Year	1	XXX	XXX	CR	3
IT	Baccalaureate Year	1	XXX	XXX	MR	3
IT	Diploma Second Year	1	XXX	XXX	DR	3
IS	Advanced Diploma Year	2	XXX	XXX	CR	3

Fig.6 – Sample Curriculum

In addition, the study followed the 4 steps on clustering the data in the Student’s Appeal System and these are:

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centers.
2. Assign each object to the group that has the closest centered.
3. When all objects have been assigned, recalculate the positions of the K centered.
4. Repeat Steps 2 and 3 until the centered no longer change. This produces a separation of the objects into groups from metrics. Seddawy et.al.(2012)

To illustrate the K-means algorithm on the datasets for re-sit, make-up and appeal against result. The k points placed in order to represent initial group centers as shown on fig. 7.

```

$this->load->model('registrar_model');
$specs = $this->registrar_model->get_specializations();
$studs = array();
foreach($specs as $s)
{
    $stud[$s->specialization]['Diploma First Year'] = $this->registrar_model->get_students($s->specialization,'Diploma First Year');
    $stud[$s->specialization]['Diploma Second Year'] = $this->registrar_model->get_students($s->specialization,'Diploma Second Year');
    $stud[$s->specialization]['Advanced Diploma Year'] = $this->registrar_model->get_students($s->specialization,'Advanced Diploma Year');
    $stud[$s->specialization]['Baccalaureate Year'] = $this->registrar_model->get_students($s->specialization,'Baccalaureate Year');
}

```

Fig.7 – Sample Code for Grouping of Data

Furthermore, to be able to assign the k points on each object to the group that has closest to the clustered data, the study used the condition in terms of year level, filter the student records, number of failed subjects, curriculum subjects count, failed, absent and passed status as shown on fig.8.

```

foreach($studs as $kspec => $vspec){
    foreach($vspec as $klevel => $vlevel){
        $cur_subjs = $this->registrar_model->get_cur_subj($kspec,$klevel);
        $cur_cnt = count($cur_subjs);
        $cnt[$kspec][$klevel]=0;
        //$cnt_attended[$kspec][$klevel]=0;
        foreach($vlevel as $s){
            $failed = 0;
            $passed = 0;
            $absent = 0;
            $norec = 0;
            $attended = 0;
            $subjs = $this->registrar_model->get_subjs($s->stud_id);
            foreach($cur_subjs as $cs){
                foreach($subjs as $sb){
                    if($cs->course_code == $sb->rec_course_code && $cs->course_title == $sb->rec_course_title)
                    {
                        if($sb->rec_result=='Failed'){
                            $failed++;
                        } else if($sb->rec_result=='ABS'){
                            $absent++;
                        } else if($sb->rec_result=='P'){
                            $passed++;
                        } else {
                            $norec++;
                        }
                    }
                }
            }
        }
    }
}

```

Fig.8 – Sample Code for assigning each object to the group

Following the steps of Seddawy et.al.(2012), “when all objects have been assigned, recalculate the positions of the K centered and repeat it until the centered no longer change. This produces a separation of the objects into groups from metrics”. The fig. 9 illustrates the nodes that satisfied the condition, and check the student if it is qualified to take the re-

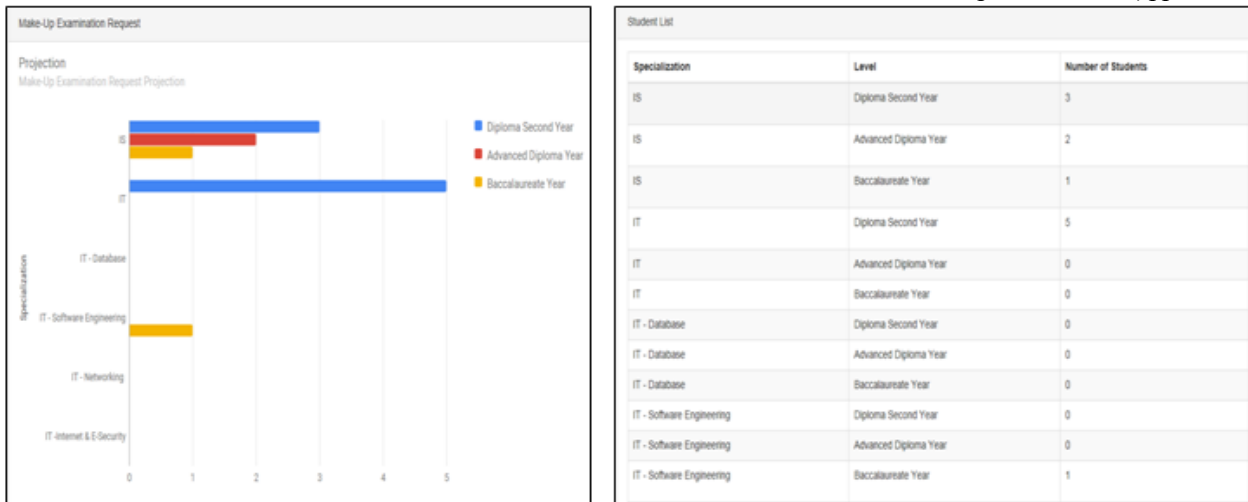


Fig.13 – Screen Output for Make-Up Appeal

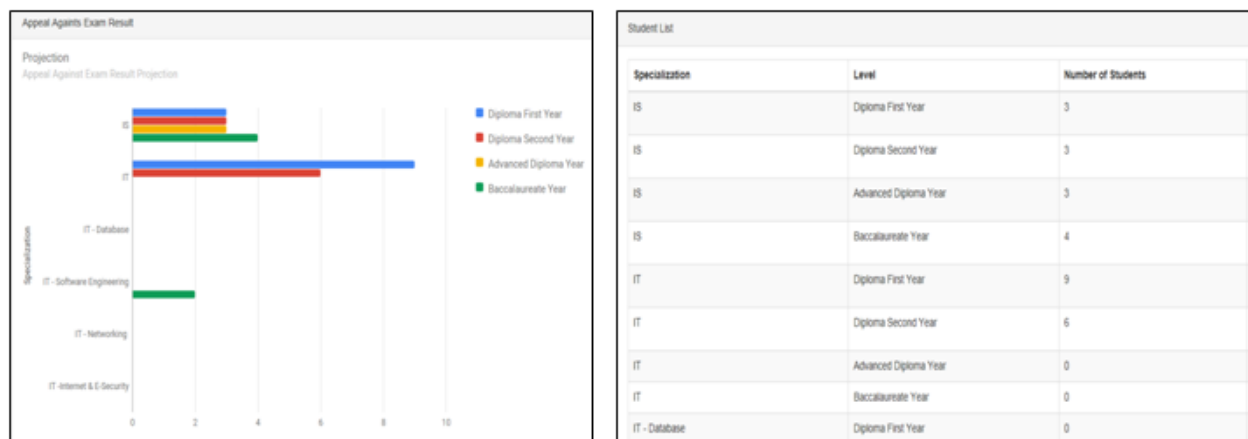


Fig.14 – Screen Output for Appeal against Exam Result

VI. CONCLUSIONS

The study used the clustering technique specifically in K-Means Algorithm perspective where centroids are divided or clustered it to the designated groups of type of appeals. In the Student Appeal System, it shows that clustering is an effective way of clustering the appeals in terms of re-sits appeal, make-up appeal, and appeal against exam result by filtering the case of proof or grounds of students. In this technique, the system can easily track the projected appeal for each of the student's enrolled last semester by grouping the student's data into re-sit, makeup and against exam result. In addition, this study helps a lot in projecting the number who are eligible in applying student's appeal easily, shorten the time in processing the appeal, increase the level of readiness in preparing the exam questionnaires ahead of time and projects the number of sections and subject to be open for the next academic semester.

REFERENCES

- [1] S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
- [2] Ahmed Bahgat El Seddawy, T. S. (2012). Enhanced k-mean algorithm to improve decision support system under uncertain situations. *International Journal of Modern Engineering Research(IJMER)* , 4094-4101
- [3] Akila, C. J. (2012). Enhanced bonding based web page information retrieval using clustering algorithm. *Ijreas* .
- [4] Chun-Wei sai, K.-W. H.-C.-S. (2009). A fast tree-based search algorithm for cluster search engine. *International Conference On Systems, Man, And Cybernetics* , 8.
- [5] Hamid Parvin, B. M.-B. (2015). A clustering ensemble framework based on selection of fuzzy weighted clusters in a locally adaptive clustering algorithm. *Journal Pattern Analysis And Applications* , 87-112.
- [6] Jain I.S, Aalam, M. A., Doja, M.N.(2010). K-means clustering using weka interface. *Proceedings Of The 4th National Conference*. New Delhi: Indiacom - 2010 Computing For Nation Development.
- [7] Narendra Sharma, A. B. (2012). Comparison the various clustering algorithms of weka tools. *International Journal of Emerging Technology and Advanced Engineering* .
- [8] Pragati Shrivastava, H. G. (2012). A review of density-based clustering in spatial data. *International Journal of Advanced Computer Research* .
- [9] V. Chitraa, A. S. (2012). An enhanced clustering technique for web usage mining. *International Journal of Engineering Research & Technology(IJERT)*