International Journal of Computer Trends and Technology
ISSN: 2231-2803 / https://doi.org/10.14445/22312803/IJCTT-V67I4P107
Volume 67 Issue 4, 33-36, April 2019

Type of article

The Application of Data Mining in Sports and Extracurricular Activities

Tansen Patel[1], Uttam Kumar Sahu[2]

[1]*Asst. Professor Computer Science and Engineering,*
Shri Shankaracharya Institute of Professional Management and Technology, Raipur, India
[2]*Research ScholarComputer Science and Engineering,*
Shri Shankaracharya Institute of Professional Management and Technology, Raipur, India

Abstract *- Data mining has played a more important role in all fields and had an increasingly greater influence. By using association analysis of data mining and classification algorithm, this paper analyzes the result data of college sports and extracurricular activities held in the last four years data of csvtu - youth festival. It studies the correlation between different students, branches, semesters, events, and years. This research paper uses the k-means technique in data mining for mining our dataset, which comes under clustering techniques of data mining. WEKA software is used to implement the result part. The paper suggests the implementation of the k-means Clustering technique of data mining.*

Keywords *- Data Mining, WEKA, Clustering, k-means.*

I. INTRODUCTION

The data being generated and stored is growing exponentially due largely to the continuing advances in computer technology. This presents opportunities for those who can unlock the information embedded within this data and introduce new challenges. In this paper, we discuss how the modern field of data mining can be used to extract useful knowledge from the data. Those that can master this technology and its methods can derive great benefits and gain a competitive advantage [1].

Data mining has various uses. It uses pattern matching and statistical techniques. The data we contain with us is often vast and noisy, meaning that it is not precise, and the data structure is complex. This is where a purely statistical technique can never be successful, so data mining is its solution. The main areas used are missing value, heterogeneity, size of data, noisy data, static data, relevance, interestingness, algorithm efficiency, dynamic data, sparse data, and complexity of data. Data mining has become an efficient tool for analyzing large datasets [2]. The concept of data mining comes into existence when the data is to be verified and stored by using the data mining so that no error should be there in the data. There are various data mining techniques available to us through which we can mine data. Some of them are Classification analysis, Association rule learning, Anomaly or Outlier detection, Clustering analysis, Regression analysis, Decision trees, Combinations, Sequential patterns, etc.

This paper is organized as follows: Chapter 2 discusses different data mining techniques, Chapter 3 describes the methodology, and Chapter 4 the result and discussion. Chapter 5 concludes the paper.

II. DATA MINING TECHNIQUES

A. Cluster Analysis

Cluster analysis enables identifying a given user group according to common features within a database. Clustering can be defined specifically as identifying similar classes of objects. This technique can be used to find distinct groups or classes of objects. Still, it is expensive, so clustering can be used pre-processing for attribute subset selections and classifications. We can identify sparse and dense regions in object space and discover total distribution patterns and relations among the data attributes using different clustering techniques. For example, to make a group of customers based on purchasing patterns, distinctly distribute genes with similar functionality. There are different types of clustering methods. Some of them are Partitioning Methods, Hierarchical Agglomerative (divisive) methods, Density-based methods, Grid-based methods, and Model-based methods [3]. Cluster analysis enables a given user group according to common features within a database. These features can include age, geographic location, education level, etc. It's a data mining technique useful in marketing to segment the database.

B. Classification Analysis

There is just one answer: classification analysis, the data mining technique that enables recognizing the pattern inside the database. An effective solution to improve marketing strategy performance is to delete any extra information to create improved sub-archives.

C. Association Rule

Association rule is one of the most active data mining methods. This rule was originally proposed for market basket analysis. Its purpose is to find a connection between commodities trading rules in the different databases. These rules are characterized by the customer buying behavior patterns that can be used to guide businesses which are scientifically arranged stock inventory and shelf design. After much research on association rule mining and doing a lot of research, their work involves exploring the theory of association rule mining with original algorithm improvements and newly designed algorithms. Association rule mining and the number of association rules mining and other issues were introduced to improve the efficiency of the mining rule algorithm, adaptability, availability, and application promotion [4].

D. Decision Trees

The Decision tree method is one of the most important classification measures in data mining. A decision tree classifier the type of classifier is a flow chart-like tree structure where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node represents the class. A decision tree model is used to classify records to find a path from the root to the leaf by measuring the attribute test, and the attribute on the leaf is the classification result. The decision tree method is the main technology used for classification and prediction. [5].

III. METHODOLOGY

A. Clustering

Clustering is an important unsupervised learning problem; every other problem of this kind deals with finding a structure in a collection of unlabeled datasets. A loose definition of clustering may be organizing objects into groups whose members are similar in some way. A cluster is a collection of objects which are similar between them and are dissimilar to the objects belonging to another cluster. We are going to verify the tests by using the clustering technique. There are various techniques and algorithms in data mining to check the validity of data. Some of them are: Hierarchical Clustering, K means Clustering, Canopy clustering, Cobweb clustering, etc.

We used the K means cluster technique to mine the data. K means clustering is fast, easy, and simple, and it can make clusters accurately. This algorithm is most accurately used as a pre-processing approach to clustering techniques like K-means. This can decrease computational expenses by starting with initial clustering as it can ignore points not in use. This clustering is a type of unsupervised learning used when you have unlabeled data. The goal of the k-means algorithm is to find groups in the data, with the number of groups represented by the variable *K*.

We are taking sports and extracurricular events (2017-18) data from the Department of CSE SSIPMT Raipur. These data have 10 attributes, i.e., roll num, student name, gender, branch, semester, participation, event position, organizer, and session. These data have a branch of CSE 3rd, 5th, and 7th semester's student records of session 2017-18. All non-numerical data is set as a specific numerical value that makes applying a k-means clustering algorithm easier.

$$J(V) = \sum_{i-1}^{c} \sum_{j=1}^{ci} (||xi - vj||)^2$$

Where,

||xi – vj || is the Euclidean distance between xi and vj.

ci is the number of data painted i^{th} cluster,

c is the number of cluster centers.

B. Algorithmic steps for k-means clustering

Let $X = \{x_1, x_2 \ldots, x_n\}$ be the set of data point and $V = \{v_1, v_2, \ldots\ldots, v_c\}$ be the set of centres.

1. Select *'c'* cluster centers randomly.
2. Calculation of the distance between each data point and cluster centers.
3. Assign the data point to the cluster center whose distance from the cluster canter is the minimum of all the cluster centers.
4. Recalculate the new cluster canter using:

$$J(V) = \sum_{i-1}^{c} \sum_{j=1}^{ci} (||xi - vj||)^2$$

where

'c_i' represents the number of data points in *the* i^{th} cluster.

5) Recalculate the distance between each data point and newly obtained cluster centers.

6) If no data point was reassigned, stop; else, repeat from step 3.

IV. RESULT AND DISCUSSION

In this paper, we have used WEKA software implemented in java language. It is open-source software. It offers data mining algorithms for pre-processing, clustering, association rule, classification, and Machine learning. In WEKA, the whole dataset is in (.arff) file format, consisting of different attributes to indicate various things in the file. First of all, the excel sheet is converted into (.csv) format, and then this file is converted into (.arff) format. Figure 4.1 shows the sample view of the lessee dataset, which consists of major attributes and values.Fig4.1 is showing data set which was used to operate data mining.

We had taken 226 student records (fig 4.1) for the data mining and analysis in which 101 male and 125 female students were involved in CSE Department. The total number of students is 83 from the 3rd, 5th, and 7th Semester Department of CSE, participating

in at least one event. The total number of students was 40, the winner in various events, and 18 students were the second position in various events.

	A	D	F	G	H	I	J	K
1	S.N.	GENDER	SEM	PARTICIF	EVENT	POSITIOI	ORGANIZER	SESSION
2	1	MALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
3	2	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
4	3	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
5	4	MALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
6	5	FEMALE	7	YES	EVENT	PARTICIF	CSVTU-YOUTH FES	2017-18
7	6	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
8	7	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
9	8	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
10	9	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
11	10	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
12	11	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
13	12	FEMALE	7	YES	EVENT	PARTICIF	CSVTU-YOUTH FES	2017-18
14	13	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
15	14	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
16	15	FEMALE	7	YES	EVENT	PARTICIF	CSVTU-YOUTH FES	2017-18
17	16	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
18	17	FEMALE	7	YES	VOLLEYI	RUNNER	CSVTU-YOUTH FES	2017-18
19	18	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
20	19	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
21	20	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
22	21	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
23	22	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
24	23	FEMALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
25	24	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
26	25	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
27	26	MALE	7	NO			CSVTU-YOUTH FES	2017-18
28	27	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
29	28	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
30	29	MALE	7	NO	,	,	CSA CRICKET TOUR	2017-18
31	30	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
32	31	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
33	32	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
34	33	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
35	34	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
36	35	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
37	36	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
38	37	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
39	38	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
40	39	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
41	40	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
42	41	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
43	42	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
44	43	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
45	44	MALE	7	YES	CRICKET	WINNER	CSA CRICKET TOUR	2017-18
46	45	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
47	46	FEMALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18
48	47	MALE	7	NO	,	,	CSVTU-YOUTH FES	2017-18

Fig. 4.1 Sports and Extracurricular Data of CSE 2017-18

The above (fig4.1) shows the data set in which we applied the k-means clustering technique. Fig4.2 shows the k means data mining technique of weka software. Here we showed the result based on student participation in various events in 2017-18.

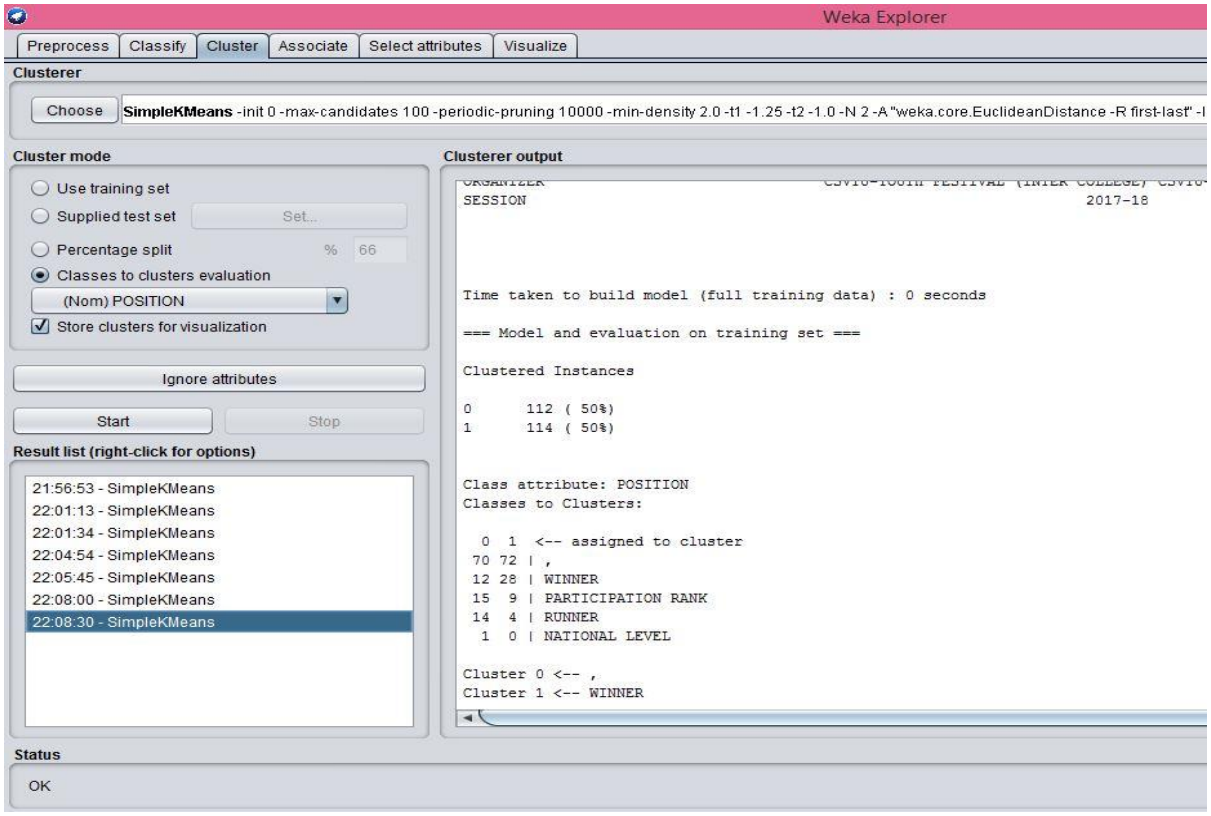

Fig. 4.2 Class to clusters evaluation – POSITION

V. CONCLUSION

Using the clustering technique in data mining, we used Simple K Means clustering, and we found the result as displayed above in the figures. We now conclude that the data inserted are valid, and the result is shown in the figure. So that we successfully applied the clustering technique in the database of Sports and extracurricular activities, and we found good results. This type of clustering is very fast, easy, and simple, and it can make clusters with precision. This algorithm is mostly used as a pre-processing approach to clustering techniques. We have used this approach on the sports and extracurricular event 2017-18 of CSE Department of SSIPMT Raipur and applied it to the details of the lessee present, which helped us obtain a group of information of objects which are alike.

REFERENCES

[1.] Gary M. Weiss, Ph.D., Department of Computer and Information Science, Fordham University.

[2.] Gary M. Weiss, Ph.D., Brian D. Davison, Ph.D.," DATA MINING," To appear in the Handbook of Technology Management, H. Bengali (Ed.), John Wiley and Sons, 2010.

[3.] M. Ramageri, Mrs. Bharati, "DATA MINING TECHNIQUES AND APPLICATIONS," Indian Journal of

[4.] Computer Science and Engineering Vol. 1 No. 4 301-305, ISSN: 0976-5166. 2010.

[5.] Liang Zhao, Deng - Feng Chen, Sheng - Jun Xu and Jun Lu, "The Research of Data Mining Classification Algorithm that Based on SJEP," International Journal of Database Theory and Application Vol.8, pp. 223-234, 2015.

[6.] Qing – Yun Dai, Chun-ping Zhang, and Hao Wu," Research of Decision Tree Classification Algorithm in Data Mining," International Journal of Database Theory and Application Vol.9, pp.1-8, No.5, 2016.