Evaluation of Supporting Work Quality Using K-Means Algorithm

Dwi Ardiada[1*], Pasek Agus Ariawan[2], and Made Sudarma[3]
[1,2]Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University
[3]Department of Electrical and Computer Engineering, Udayana University
*Email: dwiardiada99@gmail.com

***Abstract*—One of the factors that to improve the performance productivity of an organization or agency is Human Resources. During this time many government agencies that do not have employees with adequate competence, this is evidenced by the low productivity of employees and the difficulty of measuring employee performance in the scope of government agencies. This research discusses the application of K-means Clustering conducted at Udayana University, especially on annual performance data for contract workers. this study aims to classify cluster clusters that are determined to facilitate the evaluate quality of work of contract workers. This research uses data used as many as 1613 data. And done preprocessing get 544 data. In preprocessing data, K-means Clustering method is performed. In K-means Clustering determined the number of K as much as 5 Cluster. To determine Data Cluster used Ecludian Distance calculation. From the results of K-means Clustering Applying it takes 10 iterations. From 5 clusters conducted on 544 data there are clusters 0 as much as 38, Cluster 1 as much as 473, Cluster 2 as much as 130, Cluster 3 as much as 26 and from cluster 4 as many as 3. From the results of K-means Clustering implementation is used as a supporter of Quality Evaluation Work.**

***Index Terms*— Data Mining, Evaluation, K-means Clustering, Performance, Work Quality.**

I. INTRODUCTION

The quality of human resources (HR) is one factor that to improve the productivity of an organization's performance or agency. Therefore, it is necessary Human Resources who have high competence because the expertise or competence will be able to support the achievement of employee performance. During this time many government agencies that do not have employees with adequate competence, this is evidenced by low employee productivity and the difficulty of measuring employee performance in the scope of government agencies. During this assessment performance of employees at Udayana University has not been implemented optimally, especially in assessing the performance of contract workers. During this assessment of honorary employees only determined from the results of his work, there is no clear assessment criteria. The chart for employee performance appraisal as shown in Figure 1

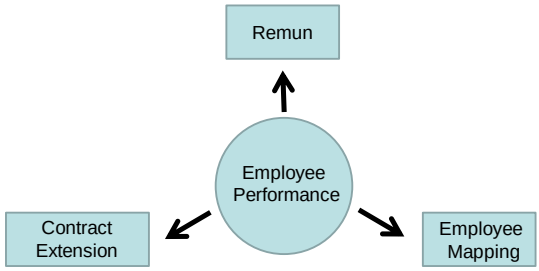

Fig. 1. Assessment of Contract Workers

Therefore, in this study will be developed performance appraisal of contractors based on competence, which is able to accommodate the performance of contract employees. To determine the priority or weight value of each competency factor is required assessment by using K-Means Clustering. The description of the evaluation of performance appraisal as in Figure 2.

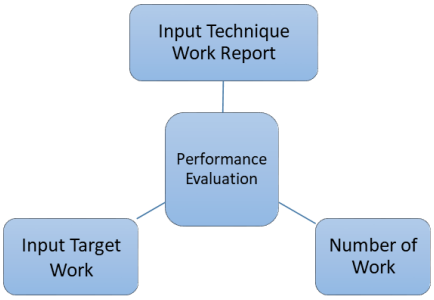

Fig. 2. Evaluation of Contract Workers

The performance appraisal to be proposed will be compared with manual assessment in terms of scale, criteria and benefits so that the research produced can be used by Udayana University in assessing the performance performance of contract employees. With proper performance appraisal, employee performance productivity can be assessed and rewarded according to its business. Thus the incentives received by contract employees are in line with their performance score [1]. The main objective of this research is to

classify the data of contract employee which then result from the grouping will be used as supporting performance evaluation of contract employee.

II. LITERATUR REVIEW

A. Data Mining

Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from large databases [2]. The term data mining has the essence as a discipline whose sole purpose is to discover, dig, or mine the knowledge of the data or information we have. Data mining, also known as Knowledge Discovery in Database (KDD) [3]. KDD is an activity that includes the collection, use of data, historically to find regularity, patterns or relationships in large data sets [9].

1) Training Methods

Broadly speaking the training methods used in data mining techniques are divided into two approaches, namely [4]:

a) Unsupervised learning,

this method is applied without any training and without any teacher. The teacher here is the label of the data

b) Supervised learning,

that is learning method with training and trainer. In this approach, to find the decision function, separator function or regression function, multiple examples of data that have output or labels are used during the training process.

2) Grouping Data Mining

There are several techniques that data mining has on the task that can be done, namely [5]:

a) Description

Researchers usually try to find ways to describe hidden patterns and trends in the data.

b) Estimates

Estimates are similar to classifications, except destination variables that are more numerical in direction than on the category

c) Prediction

Predictions bear a resemblance to estimation and classification. Only, the prediction of the results shows something that has not happened (may happen in the future).

d) Classification

In the classification of variables, objectives are categorical. For example, we will classify income in three classes, which are high income, medium income, and low income.

e) Clustering

Clustering is more toward grouping records, observations, or cases in a class that have similarities.

f) Association

Identify the relationship between the events that occur at one time.

B. Data Mining Phase

As a series of processes, data mining can be divided into several stages of the process illustrated in Figure 3.

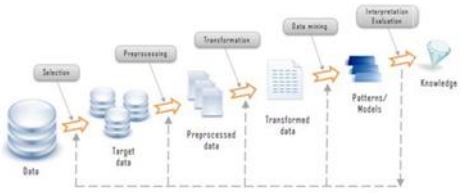

Fig. 3. Data Mining Phase

directly or by means of knowledge base. The stages of data mining are as follows [6]:

1) Data Selection dan Cleaning

Data Selection and Cleaning: Data cleaning is a process of removing noise, eliminating duplicate data and eliminating inconsistent data or irrelevant data.

2) Data Transformation

In this process the data is converted or merged into the appropriate format for processing in data mining.

3) Data Mining

The mining process is a major process when methods are applied to find valuable and hidden knowledge of data.

4) Process Mining

It is a major process when methods are applied to discover valuable and hidden knowledge of data. Some methods that can be used based on data mining grouping can be seen in Figure 4.

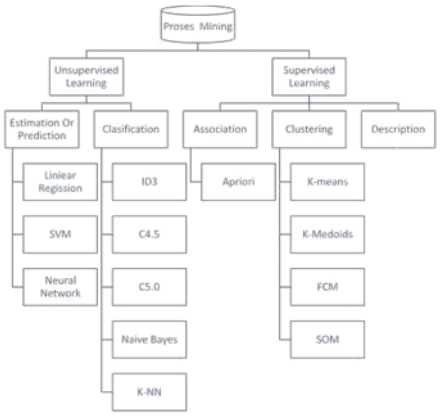

Fig. 4. Process Mining

5) Pattern Evaluation

Pattern evaluation is the stage to identify interesting patterns into the found knowledge base.

6) Knowledge Presentation

Visualization and presentation of knowledge about the methods used to acquire user acquired knowledge.

C. Clustering

Clustering is one method of data mining that is unsupervised which means that the character of each cluster is not specified before but based on the similarity of the attributes of a group or

cluster. Clusters divide data into groups or clusters based on a hazard of attributes between sets of data, similar attributes are presented as dots in multidimensional spaces. In data mining there are two types of Clustering methods used in data grouping, namely hierarchical clustering and non-hierarchical clustering [7].

D. K-Means Clustering Algorithm

One of the clustering algorithm in data mining is the K-Means Clustering algorithm to be able to produce groups that share similar attributes. K-Means is the most popular and widely used clustering method in many fields because it is simple, easy to implement, has the ability to cluster large data, able to handle data outliers. K-Means is a non-hierarchical data clustering method that attempts to partition existing data into one or more clusters. This method partitions the data into clusters / groups so that data that have the same characteristics are grouped into the same clusters and data that have different characteristics are grouped into other groups [8]. The K-Means algorithm is a relatively simple algorithm for classifying or grouping large numbers of objects with certain attributes into groups as much as K. K-Means one of the non-hierarchical clustered data methods that attempts to partition the existing data into in the form of one or more clusters or groups [9]. The K-Means grouping algorithm will result in a group of notes as much as k pieces. The K-Means algorithm was first conceived by J. MacQueen [7]. the steps of doing Clustering with the K-Means method are as follows [10].

- Select the number of clusters k.
- The initialization of this cluster center k can be done in various ways. But the most frequently done is by random. Cluster centers are preliminarily scored with random numbers
- Allocate all data / objects to the nearest cluster. The proximity of two objects is determined by the distance of the two objects. Likewise the proximity of a data to a particular cluster is determined the distance between the data with the cluster center. In this stage we need to calculate the distance of each data to each cluster center. The most distance between one data and one particular cluster will determine which data to enter in which cluster. To distance all data to each cluster center point can use Euclidean distance theory formulated in formula 1:
- Recalculate cluster center with current cluster membership. The cluster center is the average of all data / objects in a particular cluster. If desired it can also use the median of the cluster. So the mean (mean) is not the only size that can be used

E. WEKA (Waikato Environment for Knowledge analysis)

WEKA (Waikato Environment for Knowledge analysis) is a java pop-engine learning software written in java, developed at Waikato University in New Zealand. WEKA is free software available under the GNU General Public License. WEKA provides the use of clustering techniques using kmeans with the Simple Kmeans algorithm. The clustering technique and algorithm used in WEKA are called clus-terer [11]. Here's the look of WEKA app in Figure 5.

Fig. 5. WEKA App Interface

III. DISCUSSION AND RESULT

A. Data Source

Sources of data in research taken from the dataset on Udayana University is the performance data of existing contractors on each year with the number of 1613 data with Nip field, Employee Name, Status, NCP Performance, Performance Count Points, Support Count Points, Support Points, Total Points Performance, Points of Use Performance. following some data sets of contract labor performance on the table 1.

TABLE I. RESEARCH DATA

Nip	Employee Name	Status	NCP Performance	Performance Count Point	Support Count Points	Support Points	Total Points Performance	Points of Use Performance
196712312008021200 1	I Ketut Sudana	Contract	0.000	0.000	0.000	0.000	0.000	0.000
194512312000011200 1	I Made Rengkig	Contract	0.000	0.000	0.000	0.000	0.000	0.000
198703172010022200 1	Made Dian Pritasari Mahendra, SS., M.Hum	Contract	0.000	0.000	0.000	0.000	0.000	0.000
198911142014051200 1	A.A. Bagus Dharma Putra Dewangga	Contract	0.000	0.000	0.000	0.000	0.000	0.000
198801132012091200 1	I Made Dudik Suardika	Contract	0.000	0.000	0.000	0.000	0.000	0.000
198909152010012200 1	Si Luh Putu Arik Aryawathi, S.Pd	Contract	0.272	45.696	0.000	0.000	45.696	45.696
197001192012121200 1	I Nyoman Gunariadi	Contract	0.000	0.000	0.000	0.000	0.000	0.000
199109212014032200 1	Winda Saktiani Rahayu Ratna Ningsih, S.E.	Contract	0.854	143.427	0.000	0.000	143.427	143.427
198911302015041200 1	I Wayan Agus Surya Darma	Contract	0.000	0.000	0.000	0.000	0.000	0.000
195012312007011200 2	I Ketut Budiasta	Contract	0.000	0.000	0.000	0.000	0.000	0.000
199104192014102200 1	Ni Made Sukmasari Kartini	Contract	0.000	0.000	0.000	0.000	0.000	0.000
195512312016021200 1	I Nyoman Ngertiyasa, S.S	Contract	0.000	0.000	0.000	0.000	0.000	0.000
199312282013011200 1	I Kadek Agus Wiguna	Contract	0.000	0.000	0.000	0.000	0.000	0.000
198801012013031200 1	Putu Ari Arjana, S.T.	Contract	0.000	0.000	0.000	0.000	0.000	0.000
198410242017071200 2	dr Dewa Putu Agusta Cahya Udayana, S.Ked	Contract	0.000	0.000	0.000	0.000	0.000	0.000
195002242006031200 1	I Nyoman Sudana, B.A.	Contract	0.000	0.000	0.000	0.000	0.000	0.000
199304222017071200 1	Ns Gede Adi Ramananda, S.Kep	Contract	0.000	0.000	0.000	0.000	0.000	0.000
199805072017072200 1	Ni Kadek Ayu Suriani	Contract	0.000	0.000	0.000	0.000	0.000	0.000
195905092017082200 1	ERIANY MACHMUD	Contract	0.000	0.000	0.000	0.000	0.000	0.000

B. Data Transformation

At this stage, a process of data change that can be processed can be processed using K-Means Clustering algorithm. The variables selected in the performance data are NCP Performance (x1), Performance Count Points (x2), Support Count Points (x3), Support Points (x4), Total Performance Points (x5), Points Performance (x6). as in the table 2.

TABLE II. DATA PREPROCESSING

Nip	NCP Performance	Performance Count Point	Support Count Point	Support Point	Total Performance Point	Point Performance
198909152010012200 1	0.272	45.696	0.000	0.000	45.696	45.696
199109212014032200 1	0.854	143.427	0.000	0.000	143.427	143.427
199006012016092200 1	0.864	145.107	0.000	0.000	145.107	145.107
198808212011011200 1	0.748	125.631	10.058	8.133	133.764	133.764
197704062008111200 1	1.324	222.384	2.475	2.475	224.859	168.000
198507022013022200 1	2.760	463.716	2.750	2.750	466.466	168.000
198206102001102200 1	0.926	155.577	4.950	4.950	160.527	160.527
198109152004061200 1	0.734	123.352	0.000	0.000	123.352	123.352
197107092013021200 1	0.733	123.204	0.000	0.000	123.204	123.204
197302102003101200 1	0.961	161.493	0.000	0.000	161.493	161.493
198110082002091200 1	0.849	142.572	0.000	0.000	142.572	142.572
197406022008011200 1	3.398	570.792	0.000	0.000	570.792	168.000
199009042015041200 1	1.677	281.693	0.000	0.000	281.693	168.000
198503232013032200 1	0.825	138.554	0.000	0.000	138.554	138.554
197412142008011200 1	4.645	780.376	0.000	0.000	780.376	168.000
197007152015031200 1	0.876	147.097	0.000	0.000	147.097	147.097
198511042005041200 1	2.695	452.726	0.000	0.000	452.726	168.000
195112312000041200 1	5.406	908.208	0.000	0.000	908.208	168.000
198703172005082200 1	1.833	307.929	0.000	0.000	307.929	168.000

C. Data Processing

The next step is the process of data processing K-Means clustering algorithm. 1. Determine the k value of the cluster you want to create. The cluster you want to create is 5 clusters.

2. Determine the starting center point of each cluster. In this study the initial center point was determined by the random (random) method obtained from each cluster. following the initial center point of the cluster in the Table 3.

TABLE III. CLUSTER INITIAL CENTER POINT

Cluster	Nip	NCP Perfomance	Performance Count Point	Support Count Point	Support Point	Total Performance Point	Point Performance
Cluster 0	19760203200807113940	0.45	54.9	0	0	149.8	149.8
Cluster 1	19920131201605922110	0.91	82.36	6.25	6.25	164.73	164.73
Cluster 2	19880929200706111970	0.87	72.7	0	0	145.4	145.4
Cluster 3	19910504201605912130	0.74	51.63	0	0	103.26	103.26
Cluster 4	19891022201405122110	0.93	83.49	3.02	3.02	166.99	166.99

The next step is to allocate data to the cluster, then calculate the distance of each data to each cluster center by using Equation 1. A data will be a member of a cluster that has the smallest distance from its cluster center. After that calculate the new cluster center as a reference for the next iteration. because in the next iteration the position of the data does not change, then the iteration is stopped and the final result obtained is 5 cluster centers on the 16th iteration. with the final centeroid cluster in the table 4.

TABLE IV. Cluster Centroid Final

Attribute	Cluster					
	Full Data (544)	Cluster 0 (193)	Cluster 1 (27)	Cluster 2 (244)	Cluster 3 (64)	Cluster 4 (6)
x1	1.0114	1.2431	1.0859	0.7983	0.5061	7.716
x2	105.2674	138.7571	135.4559	69.7054	27.2402	1219.972
x3	2.5508	1.4773	18.3551	1.2914	1.4786	2.168
x4	2.5117	1.4773	17.8327	1.2833	1.4786	2.168
x5	210.4956	277.514	270.3895	139.4028	54.4797	2439.944
x6	141.4512	166.6453	166.5019	139.4028	54.4797	168

D. Results And Evaluation

From the application of K-means Clustering in this study obtained the result of cluster 5 groups with percentage of members in the table. From the results of the study on cluster 0.

TABLE V. RESULT OF CONTRACT PERFORMANCE CLUSTER DATA PERFORMANCE

Cluster	Instance
0	201 (37%)
1	37 (7%)
2	237 (44%)
3	64 (12%)
4	5(1%)

there are 201 members with Performance (Very Good), cluster 1 there are 37 members with Performance (Less), cluster 2 there are 237 with Performance (Good), Cluster 3 there are 64 members with Performance (Bad) and Cluster 4 there are 5 People With Performance diverse. With the data grouped the performance of contract workers facilitate management to evaluate the contract workers for future.

IV. CONCLUSION

Based on the research done, it can be concluded that K-Means algorithm can be used to group the data of contract labor based on 6 attribute of performance appraisal. From the results of clustering can support the information to perform performance evaluation of contract workers.

REFERENCES

[1] E. Nurmianto, N. Siswanto, and S. Sapuwan, "Perancangan Penilaian Kinerja Karyawan Dengan Metode Analytical Hierarchy Process (Studi Kasus di Sub Dinas Pengairan, Dinas Pekerjaan Umum, Kota Probol-inggo)," Jurnal Teknik Industri, vol. 8, no. 1, pp. 40–53, 2006.

[2] A. N. Riski Pristi Ananto, Yudha Purwanto, "Deteksi Jenis Serangan pada Distributed Denial of Service Berbasis Clustering dan Classification Menggunakan Algoritma Minkowski Weighted K-Means dan Decision Tree Detection," e-Proceeding of Engineering, vol. 68, no. 6, pp. 2015– 2017, 2017.

[3] I. P. A. Wijaya, I. Bagus, and A. Swamardika, "Data Mining for Clustering Revenue Plan Expense Area (APBD) by using K-Means Algorithm,", International Journal of Engineering and Emerging Technology. vol. 2, no. 1, pp. 87–93, 2017.

[4] M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," Eeccis, vol. 7, no. 1, pp. 59–64, 2013.

[5] A. F. Hadi, "Analisis Data Mining Untuk Menentukan Variabel Variabel Yang Mempengaruhi Kelayakan Kredit Kepemilikan Rumah Menggu-nakan Teknik Klasifikasi," UPI YPTK J. KomTekInfo, vol. 4, no. 1, pp. 108–116, 2017.

[6] B. Karim, S. R. Sentinuwo, and A. M. Sambul, "Penentuan Besaran Uang Kuliah Tunggal untuk Mahasiswa Baru di Universitas Sam Ratu-langi Menggunakan Data Mining," E-Journal Tek. Inform., vol. 11, no. 1, 2017.

[7] R. Setiawan, "Penerapan Data Mining Menggunakan Algoritma K-means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik LP3I Jakarta)," Jurnal Lentera Ict, vol. 3, no. 1, pp. 76–92, 2016.

[8] Y. Agusta, "K-Means Penerapan, Permasalahan dan Metode Terkait," Jurnal Sistem dan Informatika, vol. 3, no. Pebruari, pp. 47–60, 2007.

[9] M. Muhammad and L. Akhsani, "Kemampuan Komunikasi Matematis Dengan Metode K-Means Clustering Melalui Model Problem Based Learning," PYTHAGORAS, vol. 5, no. 2, pp. 120–130, 2016.

[10] D. Prima, T. Hapsari, and E. Widodo, "Pengelompokan Daerah Rawan Kriminalitas di Indonesia Menggunakan Analisis K-Means Clustering," Prosiding SI MaNIs (Seminar Nasional Integrasi Matematika dan Nilai Islami), vol. 1, no. 1, pp. 147–153, 2017.

[11] A. Wibowo, "Purwarupa sistem pakar indentifikasi jamur layak kon-sumsi berbasis web," Journal of Computer Engineering System and Science, vol. 2, no. 2, pp. 112–118, 2017.