International Journal of Pure and Applied Mathematics
Volume 118 No. 24 2018
ISSN: 1314-3395 (on-line version)
url: http://www.acadpubl.eu/hub/
Special Issue

http://www.acadpubl.eu/hub/

Consultancy staffing portal system based on clustring algorithm

Mohit Jain[1], Prof. Shola Usha Rani [2]
[1]Department of SITE, [2]Asst. Professor(Sr),
Department of SCSE,
VIT University, Chennai Tamil Nadu, India
mohit.jain2016@vitstudnt..in,
sholausha.rani@vit.ac.in

May 23, 2018

Abstract

In many organizations, employee data must be maintained and used for many purposes. Here, in this document, we're going to use that data to calculate an employee's performance. The data of this employee can be converted into useful information using data mining techniques such as K-means and the decision tree. K-means is used to find the rank of the employee means that the employee can be included in their criteria. The decision tree is used to find that an employee's review means that the employee needs to improve or that they meet expectations. This algorithm, when used, can identify the best employee that can be considered for the evaluation or eligible candidates for the promotion. Therefore, these algorithms, such as K-mean and the decision tree, help find the best employees for any association and help us make a good decision in less time. There are several factors that should be considered and are limited to this algorithm, so human intervention is required to consider these factors. However, classification and evaluation are seen in many companies, and this algorithm will definitely identify potential candidates. Employee performance,

Grouping, Decision tree, K-means, Employee performance, Data extraction, Euclidean distance. Therefore, it is based on the portal system. It gives candidates the opportunity to work with many respected company organizations, which requires talented individuals to shape their current and future success. Recruitment of the hiring strategy in which the client company hires skilled labor in sufficient quantity to achieve the objectives of the organization. The organization finds the people with the best skills, experience and availability to help the organization become more profitable.

***Index Terms*:**Employee performance, Clustering, Decision tree, K-means, Employee performance, Data mining, Euclidean distance.

1 INTRODUCTION

The industrial sector has a collection of employee data. This information can become useful information to make a good decision. In all industrial companies, there is a lot of employee data, such as position, salary and attendance. These data can be analyzed with quality using some mining technique. The employee's performance shows a very important role to get a good job in any company. Performance plays a role based on the employee's work experience. The very important element that is used to evaluate the performance of an employee in a company is their position and experience. The other elements, such as salary, projects, knowledge and skills, also play an important role in the employment opportunity for an employee. The employee should not evaluate only his experience and position he achieved. Therefore, a grouping of employees is based on the performance of the employee. The research has been based on the k-mean grouping and the decision tree algorithm to categorize employees in different clusters. Here, k-mean is used to find the employee's cluster, and this can help us find the rank of the employee, whether it is under criteria or not. Then, the decision tree is used here for the purpose of finding the employee's review. The decision tree helps us find the result of the predefined data media, and gives us feedback or employee review.

2 LITERATURE SURVEY

The article describes the Silhouette method in the K-mean grouping. The silhouette refers to a method of clarification and justification within groups of data [1, 2]. This methodology provides a graphic representation of how all the values are found within the cluster. The K-mean algorithm divides the number of observations in K clusters [3], where each observation can belong to the cluster with the closest average [4]. To use K, silhouette measurements are used. The cluster can be classified or classified to obtain improved student performance [5,6]. This algorithm can help to obtain a result of student performance in a classification [7]. The article deals with the algorithm of the decision tree [8]. Algorithm classifies map data into a predefined group of class data that can be classified according to their gain entropy [8, 9]. Then, the highest weighted class goes to the root and then only to a lower class than the one classified later in another class [10, 11]. When the classification reaches the last class, it is divided into a single value that is the employee's response [12]. The result of the employee is based on the performance of the employee.

3 PROPOSED WORK

A. Clustering technique (K-means)

Cluster research or clustering is the way of group a set of entities in which those entities in similar group (cluster) are more same (similar) to each other than other clusters (groups). Euclidean distance is also used to find the clusters of the K-means algorithm. To find Euclidean distance between observations, first take initial cluster centroid, cluster 1 and cluster 2 make them to centroid and calculate. Based on the calculation (Euclidean distance), every observation assigns to one cluster - calculating minimum distance.

Where ,XH: Observation value of variable height

H1: Centroid value of cluster 1 for variable height

XW: Observation value of variable weight

W1: Centroid value of cluster 1 for variable weight

B. Clustering technique (decision tree)

Decision tree is a way that is used as a hierarchy such as a map of choice and available issues, along with their results, ability

amount, and convenience. This is the root, which we are using to show this approach.

To create a decision tree, used these following rules:

- Choose one instance of practice example divide it into available instances so that the practice example is scattered in some small-small sets.
- After this method is applied for all instance of practice example. The knob of entire practice example refers to equal hierarchy, and no left instance will be done for another divination then break diving the hierarchy. Decision tree is further classified into two parts or nodes as follows:
- Leaf node - it shows the appraisal of the instance.
- Decision node - it defines many analysis for single instance appraisal distinct branch with one decision tree for every available outcome of the analysis.

C. Design of the system

Fig 1. Design of the System

4 SCREEN SHOTS OF WEB PORTAL

On this portal page number of clients post their requirement according to the skills. After we manage job on the basses of resume and manage placement basses on job requirement and resume.

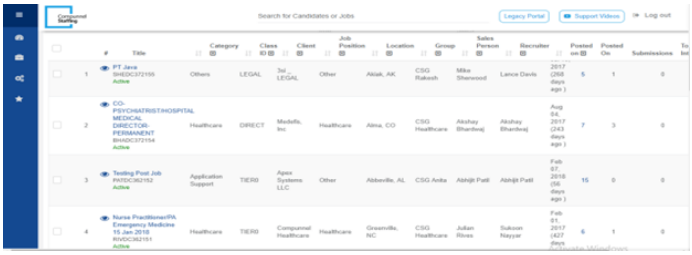

Fig 2. Job Post Page

A. Manage jobs

Recruiter uploads the candidate resume. The services work at the background of the resume and Third Party parsing services all over the data of the resume. Relational database is used for converting the data that is stored in the database.

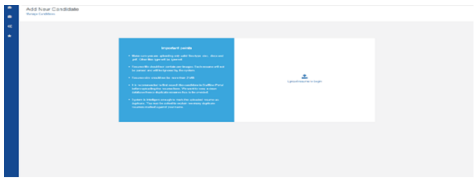

Fig 3. Add New Candidate

B. Search Resume

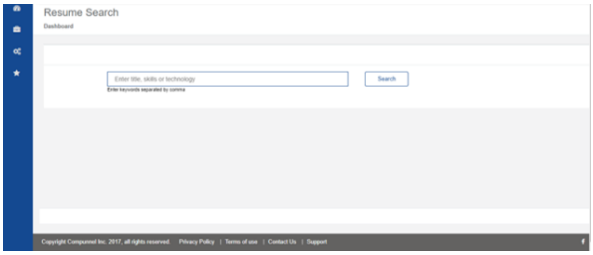

Fig 4. Search Resume

According to the skills we search the resume here. As based on the Quality and skills we cluster the resume. By the help ok K-means and decision tree algorithm it is very easy to find the best resume. For the Clustering purpose the Employee dataset is divided into two part. One is for the employee skills dataset in which the

clustering is done by the K-Means Algorithm. And the second one is on the Quality of employee in which for the best result we calculate the dataset by the decision tree. By dividing the dataset by the clustering purpose and by the help of decision tree (binary algorithm) the result will find the best employee and it will update the best employee resume on the top list. *C. K-means algorithm*

ID	Salary	Age	Experience	Skills
1	35000	26	5	Java
2	47000	25	3	PHP, ASP.net, Python, java
3	52000	28	3	Java, C, C++
4	28000	32	8	Angular, node.js, C++, Java
5	50000	23	2	Java

Table 1. Dataset of Employees

Resumes are searched in the portal on basis of their skills and requirements from the client. Each user details will be clustered and crawled according to the skills.

K-means:

Step1: Data are mapped from the database according to the requirement.

Step2: Detail of the users will be clustered on the basis of their skill using k-means algorithms.

Step3: Details of different user will be further clustered and then mapped for assigning the job.

K-means (Euclidean distance)

Apply the formula of Euclidean distance in Table 1 to find the value of the each data set. Then, we get values for dataset.

$$\sqrt{(35000-35000)^2+(26-26)^2+(5-5)^2}=0$$

$$\sqrt{(35000-47000)^2+(26-25)^2+(5-3)^2}=12000.00021$$

$$\sqrt{(35000-52000)^2+(26-28)^2+(5-3)^2}=17000.00024$$

$$\sqrt{(35000-28000)^2+(26-32)^2+(5-8)^2}=7000.003214$$

$$\sqrt{(35000-50000)^2+(26-23)^2+(5-2)^2}=15000.0006$$

$$\sqrt{(47000-35000)^2+(25-26)^2+(3-5)^2}=12000.00021$$

$$\sqrt{(47000-47000)^2+(25-25)^2+(3-3)^2}=0$$

$$\sqrt{(47000-52000)^2+(25-28)^2+(3-3)^2}=5000.0009$$

$$\sqrt{(47000-28000)^2+(25-32)^2+(3-8)^2}=19000.00195$$

$$\sqrt{(47000-50000)^2+(25-23)^2+(3-2)^2}=3000.000833$$

In Table 2, after getting the values of dataset, then assign each row to a cluster according to their values. Hence, in previous to find cluster through K-means using Euclidean distance technique, we first considered two centroids then calculate the cluster value of first two rows. Then, after this, we update the cluster again, and again, by taking the mean of the centroid. However, in the modified algorithm, there is no need to update centroid again, and again, fix the value of centroid then calculate the values of the cluster. It remains same, but it decreases the calculation time. If column one gets the highest values, then assigns that cluster value of that row is 1 otherwise gives it to cluster 2.

D. Decision tree

Table 4 contains the employees dataset for decision tree. The result of "proper" is 7 and result of "improper" is 3, there are two different values.

Table 2: Clustered data

Id	Cluster 1	Cluster 2	Assignment
1	0	12000.00021	2
2	12000.00021	0	1
3	17000.00024	5000.0009	1
4	7000.003214	19000.00195	2
5	15000.0006	3000.000833	1

Table 3: Assign cluster to dataset

Id	Salary	Age	Experience	Assignment
1	35000	26	5	2
2	47000	25	3	1
3	52000	28	3	1
4	28000	32	8	2
5	50000	23	2	1

Table 4: Dataset for decision tree

Work task	Skills	Inventiveness	Quality	Result
Weighty	Genuine	Excellent	Great	Proper
Weighty	Simple	Excellent	Great	Proper
Weighty	Genuine	Normal	Great	Proper
Weighty	Genuine	Normal	Great	Proper
Weighty	Genuine	Normal	Poor	Improper
Light	Genuine	Excellent	Poor	Improper
Light	Genuine	Excellent	Poor	Proper
Light	Simple	Excellent	Great	Proper
Light	Genuine	Excellent	Poor	Proper
Light	Simple	Normal	Poor	Improper

$$E(\text{work task}) = \frac{5}{10} * I(4,1) + \frac{5}{10} * I(3,2) = 0.8464 \text{ bits}$$

$$E(\text{quality}) = \frac{5}{10} * I(2,3) = 0.4885 \text{ bits}$$

$$E(\text{skills}) = \frac{7}{10} * I(5,2) + \frac{3}{10} * I(2,1) = 0.7810 \text{ bits}$$

$$E(\text{inventiveness}) = \frac{6}{10} * I(5,1) + \frac{4}{10} * I(2,2) = 0.87074 \text{ bits}$$

Hence, calculate info gain: Gain (work task) = 0.8812-0.8464 = 0.0348 bits Gain (quality) = 0.8812-0.7812 = 0.1 bits Gain (initiative) = 0.8812-0.87074 = 0.0104 bits The maximum gain information is quality so select quality as root. Then, split the tree into just minimum gain of quality and then do where every node is node

split. It shows in Fig. 1. Decision tree is also creating by if else command in any programing language.

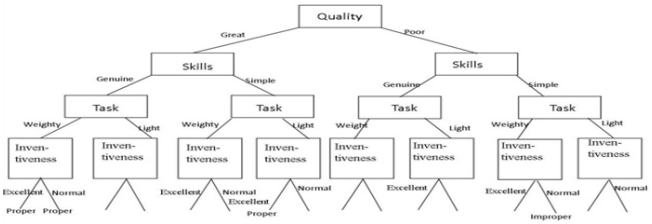

Fig 5. Decision Tree

IF (quality = "great" and skills = "genuine" and task = "weighty" and inventiveness = "excellent"), then result = "proper"

ELSE IF (quality = "great" and skills = "genuine" and task = "light" and inventiveness = "normal"), then result = "proper"

ELSE IF (quality = "great" and skills = "simple" and task = "weighty" and inventiveness = "normal"),

then result = "proper"

ELSE IF (quality = "poor" and skills = "genuine" and task = "light" and inventiveness = "excellent"),

then result = "improper"

ELSE IF (quality = "poor" and skills = "simple" and task = weighty and inventiveness = normal), then result = "improper"

Best Resume Updated

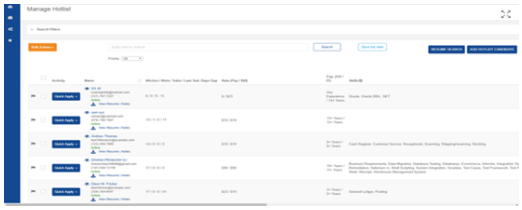

Fig 6. Resume Updated Page

To find the exact requirements given by the client and to find the matching candidates for the given jobs given by client the based resume is updated. Schedule an interview for the perfect candidate to the job and convert it to the Employee. To Place the matching candidate to the client.

5 RESULTS AND DISCUSSION

Users will be categorized in different form according to their Skills. Each user details will be clustered and according to the skills and the nearest skill using K-means Algorithm. After Clustering the best set of skill user will be selected and further assigned for the job. The Best skill performer will be mapped.

6 Conclusion

It offers user to enter the data through simple and interactive forms. This is very helpful for the client to enter the desired information through so much simplicity. The user is mainly more concerned about the validity of the data, whatever he is entering. There are checks on every stages of any new creation, data entry or update so that the user cannot enter the invalid data, which can create problems at later date. Data storage and retrieval will become faster and easier to maintain because data is stored in a systematic manner and in a single database. This may also help the manager to check the employee performance and those employees want extra attention for decreasing falling ratio for taking the strict action right time. Decision tree method is used on the previous year data of employee performance. This data mining approach helps the institutes, companies, and anywhere, where the need of employee and this ranking and review technique will help to the manager to find best employees in minimal time when he has so many number of employee data or big dataset. Hence, here, we describe about raking and reviews that help us to take a good decision in less time.

References

[1] Kumar SA, Vijayalakshmi MN. Mining of student academic evaluation records in higher education. In: Recent Advances in Computing and Software Systems (RACSS), 2012 International Conference on IEEE; 2012. p. 67-70.

[2] Geng X, Luo L. Multilabel ranking with inconsistent rankers. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2014. p. 3742-7.

[3] Bouhmala N. How good is the Euclidean distance metric for the clustering problem. In: Advanced Applied Informatics (IIAI-AAI), 2016 5th IIAI International Congress on IEEE; 2016. p. 312-5.

[4] Esteves RM, Hacker T, Rong C. Competitive k-means, a new accurate and distributed k-means algorithm for large datasets. In: Cloud Computing Technology and Science (Cloud Com), 2013 IEEE 5th International Conference on IEEE. Vol. 1; 2013. p. 17-24.

[5] Kumar KM, Reddy AR. A fast K-means clustering using prototypes for initial cluster center selection. In: Intelligent Systems and Control (ISCO), 2015 IEEE 9th International Conference on IEEE; 2015. p. 1-4.

[6] Potera CM, Mocanu ML. Evaluation of an optimized K- means algorithm based on real data. In: Computer Science and Information Systems (Fed CSIS), 2016 Federated Conference on IEEE; 2016. p. 831-5.

[7] Kotalwar R, Gandhi S, Chavan R. Data mining: Evaluating performance of employees using classification algorithm based on decision tree. Eng Sci Technol Int J 2014;4:29-35.

[8] Yang Y, Chen W. Taiga: Performance optimization of the C4. 5 decision tree construction algorithm. Tsinghua Sci Technol 2016;21(4):415-25.

[9] Guleria P, Thakur N, Sood M. Predicting student performance using decision tree classifiers and information gain. In: Parallel, Distributed and Grid Computing (PDGC), 2014 International Conference on IEEE; 2014. p. 126-9.

[10] Vaidya J, Shafiq B, Fan W, Mehmood D, Lorenzi D. A random decision tree framework for privacy-preserving data mining. IEEE Trans Dependable Secure Comput 2014;11(5):399-411.

[11] Lin C, Du X, Jiang X, Wang D. An efficient and effective performance estimation method for DSE. In: VLSI Design, Automation and Test (VLSI-DAT), 2016 International Symposium on IEEE; 2016. pp. 1-4.

[12] Chen Q, Gong Z. Data mining modelling of employee engagement for it enterprises based on decision tree algorithm. In: Information Management, Innovation Management and Industrial Engineering (ICIII), 2013 6th International Conference on IEEE. Vol. 2; 2013. p. 305-8.