

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/356672427>

Implementation K-Means Clustering Method in Job Recommendation System

Conference Paper · October 2021

DOI: 10.1109/ICEEIE52663.2021.9616654

CITATIONS
22

READS
1,267

4 authors, including:

 **Betty Dewi Puspasari**
Institut Teknologi Nasional Malang
18 PUBLICATIONS 122 CITATIONS
[SEE PROFILE](#)

 **Andy Pramono**
State University of Malang
32 PUBLICATIONS 143 CITATIONS
[SEE PROFILE](#)

 **Aang Kisnu Darmawan**
Universitas Islam Madura
62 PUBLICATIONS 421 CITATIONS
[SEE PROFILE](#)

Implementation K-Means Clustering Method in Job Recommendation System

1st Betty Dewi Puspasari
Informatics Engineering Departement
STT Atlas Nusantara Malang
Malang, Indonesia
betty@sttar.ac.id

2nd Lany Lukita Damayanti
Informatics Engineering Departement
STT Atlas Nusantara Malang
Malang, Indonesia
lany.lukita1@gmail.com

3rd Andy Pramono
Art and Design Departement
Univeritas Negeri Malang
Malang, Indonesia
andy.pramono.fs@um.ac.id

4th Aang Kisnu Darmawan
Department of Information System
Universitas Islam Madura
Pamekasan, Indonesia
ak.darmawan@gmail.com

Abstract—Work is important for everyone to earn income. With the large number of new graduates each year, finding job vacancies is a problem for students who have just completed their studies in higher education because they still do not have work experience so they are required to look for jobs that really match their criteria. Applications made can recommend specific job vacancies for undergraduates from universities (undergraduates) with the K-Means Clustering method. Applications in the form of websites that become third parties for companies and applicants. This application is one of the means that can provide solutions to companies and applicants in finding workers or jobs using a recommendation system. The problem to be studied is how to apply the K-Means Clustering method to the job vacancy recommendation system. The recommendation system in this application will calculate the level of match of the applicant's main skills, salary, location, and other skills with the needs of the company. The stages of making a recommendation system are making system designs and designs which include context diagrams, DFD, ERD and interface design. built with PHP, Java, jQuery, JavaScript, HTML, and CSS. Program testing is done by black box testing method. The weight given shows that the K-Means Clustering method can be applied to the job vacancy recommendation system and can display job recommendations according to the applicant's personal data. Questionnaire testing is given to applicants, companies, and admins with elements of testing based on user satisfaction, user convenience and system quality, resulting in the conclusion that the system can run well by getting a percentage of 87.6%.

Keywords—Jobs, Job Vacancies, Recommendations, K-Means Clustering, Website

I. INTRODUCTION

Work is important for everyone to earn income. With the large number of new graduates each year, finding job vacancies is a problem for students who have just completed their studies in higher education because they still do not have work experience so they are required to look for jobs that really match their criteria.

According to Narendra Nathan that the identification of the parameters of interest is achieved by a priori specification of the distribution of wage-vacancies, as required by the applicants [3].

The divided the labor flow into different age groups and gender groups. For the young male population group, business cycle, government expenditure for education and reward of formal education are found to be the significant determinant factors and for the females of the same age group, business cycle and reward from formal education are the significant determinants factors. [1]–[3].

Along with widespread Internet access, job vacancies can not only be seen through print media, but can also be accessed online [4]. Until now, there have been many websites and applications providing job vacancies. There are many backgrounds of applicants, while companies also have different specifications in job vacancies. Therefore, an application is needed to help applicants find job vacancies that match their abilities and fields of interest and that can recommend specific job vacancies for undergraduates from universities.

Previous research related to applications such as the Puspasari research, where the use of applications can increase efficiency and data information on special processes that are displayed [5]. In other research related to applications such as in Pramono's research, application design with concepts and interfaces can provide alternative information solutions for users [6].

Data Clustering is one of the Data Mining methods that is unsupervised [7]. There are two types of data clustering that are often used in the data grouping process, namely hierarchical (hierarchical) data clustering and non-hierarchical (non-hierarchical) data clustering. K-Means is a non-hierarchical data clustering method that attempts to partition existing data into one or more clusters/groups [8]–[13].

This method partitions data into clusters so that data with the same characteristics are grouped into the same cluster and data with different characteristics are grouped into other groups [14]. The purpose of this data clustering is to minimize the objective function set in the clustering process, which generally tries to minimize variations within a cluster and maximize variations between clusters. The benefits of Clustering are as Object Identification (Recognition) for example in the field of Image Processing, Computer Vision or robot vision. In addition, it is a Decision Support System

and Data Mining such as market segmentation, area mapping, marketing management etc [15].

II. RESEARCH METHOD

A. Research Location

The location that became the place to carry out research on job vacancy recommendation systems for undergraduates was carried out at the Atlas Nusantara Engineering College (STTAR) Malang, JL. Pacitan no. 14, Arjosari Malang City. Data is taken from STTAR students and alumni in the form of data to complete the profiles of applicants, companies, and admins. STTAR students and alumni also register and access the website as applicants, companies, and admins.

B. Data Collection

Data collection is done by means of surveys and interviews that will be given to students and alumni of STTAR which includes several factors as follows like Main skill, Other skills, salary reference and Location

Job vacancy data collection is taken from social media. The next data collection is by observing studying the problems that have to do with the object under study, namely information about the system.

C. Functional and Non-Functional Needs

The recommendation system application has the following functional requirements:

- 1) The division of users into 3 groups in this recommendation system, namely: Admin, Applicants, Company
- 2) Administrator is in charge of managing job vacancies data. Admin can also manage applicant data and company data
- 3) Applicants and employers can apply online
- 4) Applicants can search for job vacancies by job type, city, and salary
- 5) Companies can add job vacancies
- 6) Employers can manage personal job vacancies

The job recommendation system has the following non-functional requirements:

- Web-based applications are placed on a web server
- Users can access the application via a web browser
- An internet connection is required to access the application

D. K-Mean Method

Data clustering using the K-Means method is generally carried out with the following basic algorithm:

- Determine the number of clusters
- Allocate data into clusters randomly
- Calculate the centroid / average of the data in each cluster
- Allocate each data to the nearest centroid/average
- Return to step 3, if there is still data that moves clusters or if the centroid value changes, some are above the specified threshold value or if the change in the value of

the objective function used is above the specified threshold value.

Characteristics of K-means:

- means is very fast in the clustering process.
- K-means is very sensitive on random initial centroid generation.
- Allows a cluster to have no members.
- The results of clustering with K-means are unique (always changing, sometimes good, sometimes bad)

Several distance spaces have been implemented in calculating the distance between the data and the centroid, including L1 (Manhattan/ City Block) distance space, L2 (Euclidean) distance space, and Lp (Minkowski) distance space. The distance between two points x_1 and x_2 in Manhattan/City Block distance space is calculated using the formula (1) [3]

$$D_{L1}(x_2, x_1) = \|x_2 - x_1\|_1 = \sum_{j=1}^p |x_{2j} - x_{1j}| \quad (1)$$

As for the L2 (Euclidean) distance space, the distance between two points is calculated using the formula (2).

$$D_{L2}(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^p (x_{2j} - x_{1j})^2} \quad (2)$$

Where:

D_{12} = Euclidean square distance between objects to x_2 by x_1 .

P = number of cluster variables.

X_{2j} = the value or data of the 2nd object in the j variable.

X_{1j} = value or data from the 1st object in the j variable

E. System Design

Context diagram contains a process that shows how the system as a whole works. The main data from and into the job vacancy recommendation system for undergraduates. The application has three (3) users, namely Admin, Company and Applicants. Each user has different access rights. See in Fig 1.

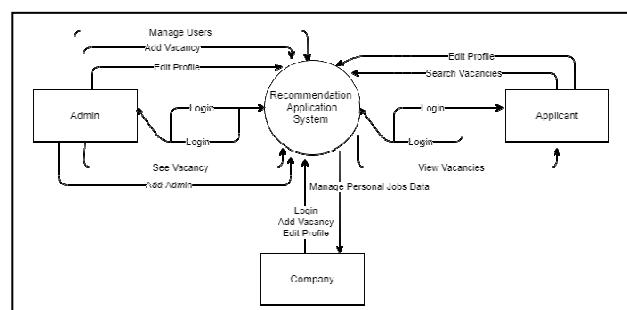


Fig. 1. Context Diagram

On the job vacancy recommendation system for undergraduates. There are 5 tables, namely admin, company, applicants, vacancies, and curriculum vitae. The underlined text is the primary key of each table. See in Fig. 2.

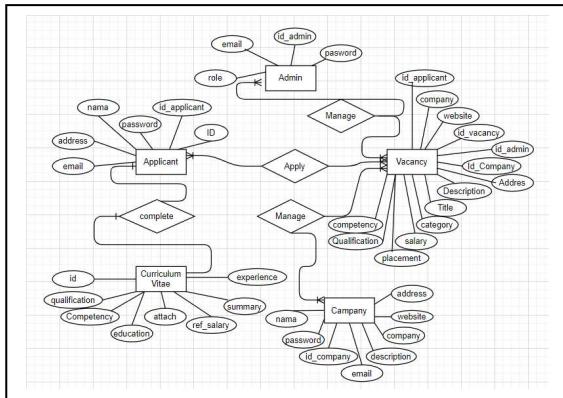


Fig. 2. Entity Relationship Diagram

III. DISCUSSION

A. Program Implementation

Program implementation is carried out to determine whether the system that has been built can run well and meet the specifications that have been determined.

B. Login

The login page is the first page that the user must access before being able to see the features in the application. Users can log in after registering their personal data, then the user logs in using an email and password. There are 2 types of login pages, the first is for admins and the second is for applicants and companies. Fig. 3 show the login screen for admins and Fig. 5 shows the login screen for applicants and companies

The login form contains fields for Email address (ozzi98@gmail.com) and Password (*****). It includes a 'Forgot Password' link and a 'Remember Me' checkbox. A blue 'Masuk' button is at the bottom.

Fig. 3. Login for Admin, Applicant and Company

C. Admin View

Admin can manage applicant data, which includes adding accounts, editing profiles, and deleting applicant accounts. The view of managing applicant data can be seen in Fig 4 is manage Application data and the display of the new applicant input form.



Fig. 4. Manage Applicant Data

Admin can manage company data, which includes adding accounts, editing profiles, and deleting company accounts. The view of corporate data management is the same with Fig 4. In Fig. 5 is the display of the new company input form.

(a) Tambah Lowongan (Add Job) form: Fields include Nama Pekerjaan (Job Title), Nama Perusahaan (Company Name), and Deskripsi Pekerjaan (Job Description). It also includes a file upload section for CV and a rich text editor for Description.

(b) Open/Close Vacancy screen: Shows a job listing for 'App Security Specialist/Engineer' at 'PT Angkasa Raya Sermata'. It includes buttons for 'Edit Lowongan' (Edit Job), 'Kelola Lowongan' (Manage Job), 'Tutup Sementara' (Temporary Close), 'Tutup Permanen' (Permanent Close), and 'Hapus' (Delete). It also shows 'Tingkat Pekerjaan' (Job Level) as 'Junior/Senior' and 'Kategori' (Category) as 'Teknik Informatika'.

Fig. 5. (a) Entry Job data, (b) Open/Close Vacancy

Views for managing admin data and adding admins almost the same with Fig.4, Fig 5. This feature can only be done by the main admin.

D. Company View

Fig 6 shows the company profile view. The company profile displayed includes the company address, website, industry category, and information about the company.

Companies can manage data on vacancies and incoming applications from their respective companies. Companies can also see notifications.

Companies can take action on incoming applications, namely accepting applications, rejecting applications, putting them on a shortlist, or delaying a review. The result of the action will appear in the notification.

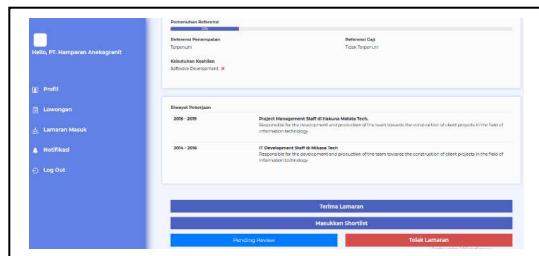


Fig. 6. Company Views

E. Applicant View

Applicant profiles contain reference features of skills, salaries, and locations that will be used to determine suitable vacancies for applicants. The appearance of the applicant's profile can be seen in Fig. 7.

The applicant dashboard displays the available vacancies, sorted from the vacancies that have the highest percentage of matches with the applicant's profile. The dashboard display of applicants.

The system will display recommended vacancies according to the applicant's main study program or expertise. An example of the applicant's main expertise is in the field of informatics engineering, so of the 9 available vacancies, only 6 vacancies are shown that require services in the field of informatics engineering.

Applicants can submit their applications if there are vacancies of interest. The list of applications that have been sent can be seen in the application menu, after being processed by the company the results will appear in the notification



Fig. 7. Applicant Views

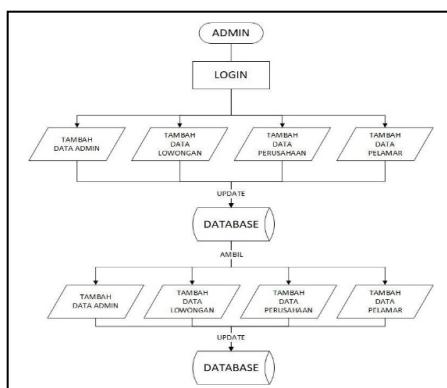
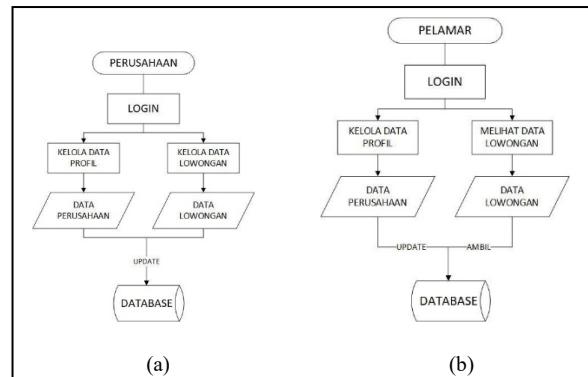


Fig. 8. Admin User Test

F. System Test

System testing is done by using the black-box method to test the system's functionality on three users, namely admin, applicant, and company. Fig. 8 is the test flow for the admin user. After logging in, the admin can perform all available functions, from adding data to editing data on the available functions.

Companies can access the website by logging in using an account that has been registered. Companies can manage their respective vacancies and can manage application data that enter the vacancies of each company. is a test flow for enterprise users. can be seen in fig. 9.



Testing on user applicants, applicants must login using an account that has been registered first. Applicants can see a list of all vacancies as well as recommended vacancies based on the profile that has been filled in, besides that applicants can also submit applications for vacancies of interest. Fig. 9 (b) is a test flow for user applicants.

G. K-Means Clustering Test Results

The application of K-Means Clustering is found in the recommendation function on the applicant search page. The results of the recommendations can appear after the applicant fills in the references on the profile. Applicant reference data is a variable that will determine the weight of the variables contained in the vacancy table. The system will display job vacancy recommendations according to the main skill because the main skill is given the highest weight, then job vacancies will appear according to the percentage of matches with the highest profile of the applicant.

Fig. 10. Reference Data on Applicant

In Fig. 10 is reference data on applicants that will be used as a determinant of the weight of job vacancies. After

determining the weight of the variables in the vacancies, the weights obtained are as shown in Table 1.

TABLE 1. QUALITY VACANCY

Vacancy	Key Skills	Location	Other Skill	Salary	Total
Mechanic Production	1	1	1	0	3
App Security Specialist Engineering	0	1	0	0	1
C++ Developer	0	1	0	0	1
Full stack Software Developer	0	0	0	1	1
Production Engineering Manager	1	0	1	0	2

Then the cluster center is taken randomly, namely production mechanic vacancies with a value of 3. The data distance from the first vacancy to the 5th vacancy is calculated using the Euclidean formula as follows:

$$\text{Centroid}_1 (C_1 / x_1) = 3$$

$$D_{1j} (x_2, x_1) = \sqrt{\sum_{j=1}^p (x_{2j} - x_{1j})^2}$$

$$D_1(3, 3) = \sqrt{(3-3)^2} = \sqrt{0} = 0$$

$$D_1(1, 3) = \sqrt{(3-1)^2} = \sqrt{4} = 2$$

$$D_1(1, 3) = \sqrt{(3-1)^2} = \sqrt{4} = 2$$

$$D_1(1, 3) = \sqrt{(3-1)^2} = \sqrt{4} = 2$$

$$D_1(2, 3) = \sqrt{(3-2)^2} = \sqrt{1} = 1$$

The results of the distance calculation are then sorted from the value closest to the center, the results are as Table 2.

TABLE 2. CALCULATION TEST RESULTS D_1

Vacancy	Total	D_1	Order D_1
Mechanic Production	3	0	5
App Security Specialist Engineering	1	2	1
C++ Developer	1	2	2
Full stack Software Developer	1	2	3
Production Engineering Manager	2	1	4

H. Blackbox testing Result

The website system questionnaire was also conducted using the black-box method. The items on the system that will be used in the black-box method include: User satisfaction, User convenience, System Quality

The questionnaire has 10 questions which are grouped into 3 groups of items, namely: user satisfaction, user convenience and system quality. Each user gets a different questionnaire.

TABEL 3 RESULT

Respondent	Question										Total
	1	2	3	4	5	6	7	8	9	10	
1	5	5	4	5	5	5	5	4	5	5	48
2	4	4	5	4	4	5	4	3	5	5	43
3	5	5	5	5	5	5	5	5	5	5	50
4	4	4	4	4	3	3	4	3	5	4	38
5	5	5	5	4	3	4	4	5	4	3	42
6	4	4	4	4	5	4	4	4	5	5	42
7	4	4	5	3	3	5	5	4	4	4	41
8	5	5	5	5	5	5	5	5	5	5	50

Respondent	9	4	4	4	4	4	4	4	4	4	40
	Question										Total
1	2	3	4	5	6	7	8	9	10		
10	4	3	3	4	4	3	4	5	4	4	38
11	4	5	5	4	5	5	5	5	5	5	48
12	4	4	4	4	4	4	4	4	4	4	40
13	5	5	4	4	4	3	5	4	3	5	42
14	5	4	4	4	4	4	5	5	5	5	45
15	5	5	5	5	5	5	5	5	5	5	50
Total Actual Skor											675
Total Ideal Skor											750
Nilai % actual											87,6%

Table 6 shows the results of the questionnaire as a whole which shows a value of 87.6%, these results are obtained from calculations using the index formula.

$$\text{Index \% formula} = \text{Actual score}/\text{ideal score} * 100.$$

Description = the ideal score is obtained from the results of the number of weights (5) * number of questions (10) * number of respondents (15) = 750

Respondents totaling 15 respondents are ideal enough to test the functionality of the whole system, then the results of the questionnaire which show the actual percent value of 87.6% of the total perfect score which is worth 100%, it can be concluded that the assessment for the whole system can work in accordance with the functions contained in each user role because the actual score obtained exceeds 60%.

IV. CONCLUSION

After designing the system, it is tested using the black box testing method. Tests are carried out starting from the login function on admin users, companies, and applicants where the user must first register the account to be used, then add data, edit data, and delete data. From the results of testing the recommendation function, it can be concluded that the system can show recommendations as expected. The K-Means Clustering method can group applicants and vacancies according to their main skills, salary, location, and other skills by giving weights to variables, then the system will display vacancies starting from the highest percentage of matches. Questionnaire testing is given to applicants and companies with elements of testing based on user satisfaction, user convenience and system quality, resulting in the conclusion that the system can run well by getting a percentage of 87.6%.

V. ACKNOWLEDGEMENT

Alhamdulillahi rabil 'alamin, the researcher expresses his highest gratitude to Allah subhanahu wa ta'ala for blessing, love, opportunity, health, and mercy to complete this research.

I am very grateful to have research team who always support, work together, data contribution for research, until we finish our research.

REFERENCE

- [1] R. Megasari, E. Piantari, and R. Nugraha, "Graduates Profile Mapping based on Job Vacancy Information Clustering," in 2020 6th International Conference on Science in Information Technology

(ICSTech), Oct. 2020, pp. 35–39. doi: 10.1109/ICSTech49800.2020.9392067.

- [2] S. H. B. Harmadi, “LABOR FLOW DETERMINANTS IN INDONESIA,” *J. Popul.*, p. 32.
- [3] W. Narendranathan and S. Nickell, “Modelling the process of job search,” *J. Econom.*, vol. 28, no. 1, pp. 29–49, Apr. 1985, doi: 10.1016/0304-4076(85)90065-X.
- [4] A. Pramono, S. Kom, B. D. Puspasari, and S. Kom, “First aid instructional media using Android platform,” in 2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT), Aug. 2017, pp. 1–5. doi: 10.1109/CAIPT.2017.8320712.
- [5] B. D. Puspasari, H. A. Ardani, P. Stephanus, and A. Pramono, “E-COUNSELING FOR HANDLING THE SELECTION MAJORS PROBLEM FOR SENIOR HIGH SCHOOL STUDENTS IN WEB-BASED USING BACKWARD CHAINING METHOD,” *Asian J. Behav. Sci.*, vol. 1, no. 1, Art. no. 1, Oct. 2019.
- [6] A. Pramono, P. Pujiyanto, B. D. Puspasari, and N. S. Dhanti, “Character Thematic Education Game ‘AK@R’ of Society Themes for Children with Malang-Indonesian Visualize,” *Int. J. Instr.*, vol. 14, no. 2, Art. no. 2, Apr. 2021, doi: 10.29333/iji.2021.14211a.
- [7] A. Pramono and B. D. Puspasari, “APLIKASI VISUALISASI DATA SEKOLAH DI KABUPATEN PASURUAN JAWA TIMUR BERBASIS ANDROID,” *MATICS*, no. 0, Art. no. 0, Oct. 2013, doi: 10.18860/mat.v0i0.2421.
- [8] I. Mohamad and D. Usman, “Standardization and Its Effects on K-Means Clustering Algorithm,” *Res. J. Appl. Sci. Eng. Technol.*, vol. 6, pp. 3299–3303, Sep. 2013, doi: 10.19026/rjaset.6.3638.
- [9] M. Yedla, S. R. Pathakota, and T. M. Srinivasa, “Enhancing K-means Clustering Algorithm with Improved Initial Centre,” *Int. J. Comput. Sci. Inf. Technol.*, pp. 121–125.
- [10] K. Krishna and M. Narasimha Murty, “Genetic K-means algorithm,” *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 29, no. 3, pp. 433–439, Jun. 1999, doi: 10.1109/3477.764879.
- [11] J. Yadav and M. Sharma, “A Review of K-mean Algorithm.” 2013.
- [12] A. Likas, N. Vlassis, and J. J. Verbeek, “The global k-means clustering algorithm,” *Pattern Recognit.*, vol. 36, no. 2, pp. 451–461, Feb. 2003, doi: 10.1016/S0031-3203(02)00060-2.
- [13] T. I. Saputra and R. Arianty, “IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING PADA ANALISIS SENTIMEN KELUHAN PENGGUNA INDOSAT,” *J. Ilm. Inform. Komput.*, vol. 24, no. 3, Art. no. 3, Feb. 2020, doi: 10.35760/ik.2019.v24i3.2361.
- [14] B. D. Puspasari, “Aplikasi Pengenalan Huruf Alfabet Bagi Anak Usia Dini Menggunakan Metode Pengolahan Citra Berbasis Data Suara,” *J. ELTEK*, vol. 11, no. 1, Art. no. 1, Oct. 2017.
- [15] Y. Agusta, “K-Means – Penerapan, Permasalahan dan Metode Terkait,” vol. 3, p. 14, 2007.