Jurnal Statistika Dan Komputasi (STATKOM), Vol. 4 No. 1 (June 2025)
ISSN : 2963-038X (print), ISSN : 2963-0398
DOI : https://doi.org/10.32665/statkom.v4i1.4651

Clustering Job Seekers in Bojonegoro Using K-Means and Fuzzy K-Means

Adam Al Avin Faisal Hutin[1]

[1]Statistics Study Program, Universitas Nahdlatul Ulama Sunan Giri
E-mail: adamalavin@gmail.com[1]

Diajukan 29 Mei 2025 ***Diperbaiki*** 18 Juni 2025 ***Diterima*** 25 Juni 2025

Abstract

Background: *Job seekers are part of the labor force who are unemployed and actively looking for work. One of the efforts to address the rising number of job seekers is by expanding job openings or employment opportunities. Employment is an essential need for individuals to meet various aspects of life, ranging from basic needs to education and housing.*

Objective: *This paper aims to analyze the frequency distribution of job seeker attributes in Bojonegoro, compare the K-Means and Fuzzy K-Means methods in clustering sub-districts, determine the best clustering method, and describe frequency distribution for each formed cluster.*

Methods: *The methods used are K-Means and Fuzzy K-Means, both known for their ease of implementation and effectiveness in clustering large datasets by minimizing the average distance between data points within each cluster.*

Results: *The majority of job seekers in Bojonegoro in 2022 are aged 15–24, unmarried, and senior high school graduates, with males comprising 59.6% of the total. The clustering analysis, with an optimal k equal to 5, reveals five balanced groups with distinct variations in age, gender, and marital status, suggesting a range of employment needs among subgroups.*

Conclusion: *The findings indicate that most job seekers in Bojonegoro are young, male, unmarried, and secondary school graduates. The clustering process identified five relatively even groups, with K-Means slightly outperforming Fuzzy K-Means in cluster cohesion.*

Keywords : *Job Seekers, Clustering, K-Means, Fuzzy K-Means, Workforce Age.*

Abstrak

Latar Belakang: Pencari kerja adalah bagian dari angkatan kerja yang sedang menganggur dan secara aktif mencari pekerjaan. Salah satu upaya untuk mengatasi meningkatnya jumlah pencari kerja adalah dengan memperluas lowongan pekerjaan atau kesempatan kerja. Lapangan pekerjaan merupakan kebutuhan penting bagi individu untuk memenuhi berbagai aspek kehidupan, mulai dari kebutuhan pokok hingga pendidikan dan perumahan.

Tujuan: Penelitian ini bertujuan untuk menganalisis distribusi frekuensi atribut pencari kerja di Bojonegoro, membandingkan metode K-Means dan Fuzzy K-Means dalam melakukan pengelompokan kecamatan, menentukan metode pengelompokan yang paling baik, dan mendeskripsikan distribusi frekuensi pada setiap cluster yang terbentuk.

Metode: Metode yang digunakan adalah K-Means dan Fuzzy K-Means, keduanya dikenal karena kemudahan implementasi dan keefektifannya dalam mengelompokkan dataset yang besar dengan meminimalkan jarak rata-rata antar titik data dalam setiap cluster.

Hasil: Mayoritas pencari kerja di Bojonegoro pada tahun 2022 berusia 15-24 tahun, belum menikah, dan lulusan sekolah menengah atas, dengan laki-laki mencakup 59,6% dari total pencari kerja. Analisis pengelompokan, dengan k = 5 yang optimal, menunjukkan lima kelompok yang seimbang dengan variasi yang berbeda dalam hal usia, jenis kelamin, dan status pernikahan, yang menunjukkan berbagai kebutuhan pekerjaan di antara subkelompok.

Kesimpulan: Temuan ini menunjukkan bahwa sebagian besar pencari kerja di Bojonegoro berusia muda, laki-laki, belum menikah, dan lulusan sekolah menengah. Proses pengelompokan mengidentifikasi lima kelompok yang relatif sama, dengan K-Means sedikit mengungguli Fuzzy K-Means dalam hal kohesi klaster.

Kata kunci: Pencari kerja, Klasterisasi, K-Means, Fuzzy K-Means, Usia Tenaga Kerja.

INTRODUCTION

Human needs continue to evolve in line with societal and technological changes. It is no longer only to fulfill biological needs, but nowadays many people actively seek employment opportunities amidst limited job availability (Desviona & Rahmawati, 2022). The population has various needs to survive and must earn an income through working. Job vacancy information can be obtained through various media such as social media, newspapers, job portals, and others (Angela & Honni, 2018). Thus, technological advancement in the global era plays a significant role in employment dynamics.

Current technological development affects many sectors, including employers who use technology to disseminate job vacancy information to job seekers. Given these challenges, this study aims to analyze the data of job seekers in Bojonegoro Regency in 2022 using clustering methods to group the data based on age, gender, education level, and district of residence.

Given the increasing number of job seekers in Indonesia and the challenges faced by companies in selecting suitable candidates, there is a growing need for an efficient and technology-driven job vacancy information system (Yuliani et al., 2020). Job seekers are part of the workforce who are unemployed and actively looking for jobs. One way to address the increasing number of job seekers is to expand job vacancies or employment opportunities (Ningsih & Abdullah, 2021). People work not only to meet basic needs but also to fulfill broader social and economic aspirations such as clothing, housing, and education.

In seeking employment, there is a mandatory requirement to have a "Yellow Card" (AK1), which can be obtained at the local Industry and Labor Office. The AK1 card serves as proof that an individual is actively seeking employmen (Sukirman et al., 2020). Indonesia's large population and diverse human resources create complex unemployment issues.

Human resources are an essential element for job seekers. Improving the quality of human resources can be done through education and job training as regulated by Law Number 13 of 2003 concerning Manpower. In this millennial era, technological progress is rapid, so education must equip graduates with both soft and hard skills to compete in the job market (Alkordi et al., 2020). Indonesia is a country with a large population where many people compete for jobs. According to the Central Statistics Agency (BPS), in 2022, there were 937,176 unemployed people, a decrease of approximately 67.76% compared to 2.74 million people in the previous year (Simbolon, 2023).

To illustrate regional responses, in East Java, especially Surabaya, the government launched a web-based application called ASSIK (Arek Suroboyo Siap Kerjo) to facilitate job searches without requiring physical documents (Salman, 2022). At the local level, in Bojonegoro Regency, the unemployment rate in 2022 was recorded at 2,063 based on AK1 data (Kuncoro, 2022). However, these data have not been fully consolidated. The main factor of unemployment is the dominance of small and medium enterprises. The COVID-19 pandemic also caused an increase in unemployment in 2020 (Arnovia, 2022). The Bojonegoro government through DISPERINAKER disseminates job vacancy information via digital media, radio, announcement boards, official websites, job fairs, and optimizes BLK, BKK in vocational schools, and private job training institutions.

Given these contextual challenges at both national and regional levels, this study aims to analyze the frequency distribution of job seeker attributes in Bojonegoro, compare the K-Means and Fuzzy K-Means methods in clustering sub-districts, determine the best clustering method, and describe frequency distribution for each formed cluster.

Clustering is a statistical method used to determine the grouping of data by separating relevant and irrelevant attributes (Jain, 2020)(Xu & Tian, 2019). There are many methods used in clustering, such as K-Means and Fuzzy K-Means. The K-Means algorithm is a simple and widely used clustering technique due to its computational efficiency and ease of implementation. However, this method only works with numerical data, requiring categorical variables to be converted into numerical form prior to analysis (Almeida et al., 2021)(Arora & Malik, 2020). K-Means is a simple, fast, and accurate clustering algorithm (Sallaby & Suryana, 2018). K-Means is the best method in clustering education big data than K-Medoids and Random clustering (Nurdiansyah et al., 2023). This finding aligns with research on clustering sub-districts in Jember Regency based on paddy and secondary crop production (Khan et al., 2023). The K-Means method is also the main method applied in clustering sub-districts for disaster mitigation in Bojonegoro (Nurdiansyah et al., 2024).

K-Means belongs to partitioning clustering, where each data point is assigned to a specific cluster and may change membership through iterative updates. This method partitions the dataset into k distinct, non-overlapping clusters, where k is a user-defined positive integer (Jain, 2020)(Arora & Malik, 2020). In contrast, the Fuzzy K-Means method (also known as Fuzzy C-Means) determines cluster centers that represent the average position of all points weighted by their degrees of membership. Each data point can belong to multiple clusters with varying membership values, and cluster centers are updated iteratively to minimize the objective function (Bezdek et al., 2019).

Fuzzy K-Means is an extension of K-Means that is more flexible and robust in handling overlapping data characteristics. It provides better results when data distributions are not well-separated or exhibit ambiguity in class boundaries (Wang et al., 2019). However, the method has limitations in clearly identifying whether a data point belongs to multiple clusters or is simply an outlier, due to its probabilistic membership structure (Abdel-Aty et al., 2021).

This study explains the clustering of sub-districts in Bojonegoro Regency based on job seeker attributes using K-Means and Fuzzy K-Means clustering methods. The novelty of this study lies in the application of job seeker attribute data for regional-level clustering analysis, particularly in Bojonegoro Regency. The application of both methods was performed using the open-source R software, a free statistical software easily obtained from its official website (Nurdiansyah & Sulistiawan, 2023).

The benefits of this research are expected to be felt by various parties. For academics, the results can increase knowledge and insight, especially regarding the clustering of sub-districts in Bojonegoro based on job seeker attributes. Additionally, this study can serve as a reference for future research on similar topics. For the Bojonegoro Regency Government, especially the Industry and Labor Office, the results can provide input on the number of job seekers in the area and serve as a basis for formulating more targeted employment policies. Meanwhile, for the community, this research provides useful information about the number of job seekers in Bojonegoro, which can raise awareness about employment conditions in the local environment.

METHOD

Research Design

This study employs a quantitative research design using a comparative approach between clustering methods. The statistical methods used are K-Means and Fuzzy K-Means clustering algorithms to analyze secondary dataset on job seekers.

Population and Sample

The population of this research includes all job seekers registered in Bojonegoro Regency in 2022. The sample consists of a secondary dataset of job seekers registered in Bojonegoro Regency in 2022. The data were obtained from the Industry and Labor Office (DISPERINAKER) of Bojonegoro Regency and are considered representative of the job seeker distribution in the area.

Sampling Techniques

No primary sampling was conducted due to the nature of secondary data used in this study. The entire dataset obtained from DISPERINAKER was analyzed without applying any sampling techniques.

Research Subjects

The research subjects are records in the secondary dataset of job seekers in Bojonegoro Regency during 2022. The variables studied are shown in Table 1, which include age, gender, education level, marital status, and district of residence.

Table 1. Definition of Research Variables.

Variable	Measurement Scale	Value
Age	Ordinal	1 = 15 – 24 years (Entry-Level Job Seekers) 2 = 25 – 34 years (Young Professionals) 3 = 35 – 44 years (Mid-Career Job Seekers) 4 = 45 – 54 years (Senior Job Seekers) 5 = 55 years and above (Retiree Job Seekers)
Gender	Nominal	1 = Man 2 = Woman
Marital Status	Nominal	1 = Not Married 2 = Married
Education	Ordinal	0 = Not Graduated from Elementary School 1 = Elementary School 2 = Junior High School / Islamic Junior High School 3 = Senior High School / Vocational High School / Islamic Senior High School 4 = Diploma / Bachelor 5 = Postgraduate
District	Nominal	1 = Balen, 2 = Baureno, 3 = Bojonegoro, 4 = Bubulan, 5 = Dander, 6 = Gayam, 7 = Gondang, 8 = Kalitidu, 9 = Kanor, 10 = Kapas, 11 = Kasiman, 12 = Kedewan, 13 = Kedungadem, 14 = Kepohbaru, 15 = Malo, 16 = Margomulyo, 17 = Ngambon, 18 = Ngasem, 19 = Ngraho, 20 = Padangan, 21 = Purwosari, 22 = Sekar, 23 = Sugihwaras, 24 = Sukosewu, 25 = Sumberrejo, 26 = Tambakrejo, 27 = Temayang, and 28 = Trucuk.

Data Analysis Techniques

The data analysis process includes several steps:

1) Importing the data into R software in CSV format.
2) Presenting the frequency distribution to find out the information of each variable in the data (Kartini et.al, 2024).
3) Checking for missing values and removing incomplete entries to ensure data quality.
4) Encoding categorical variables (e.g., age, gender, marital status, education, district) into numeric format.
5) Determining the optimal number of clusters using the Elbow method.
6) Applying K-Means clustering using the kmeans() function and Fuzzy K-Means using the fkmeans() function in R.
7) Evaluating clustering performance with Sum of Squared Errors (SSE).
8) Visualizing SSE values across different numbers of clusters to identify the optimal cluster number using the Elbow criterion (Fitriyah et al., 2023).
9) Assigning data to clusters based on the best method.
10) Calculating the frequency distribution

for each cluster.

11) Interpreting the clustering results.

RESULTS AND DISCUSSION

Job Seeker Dataset Overview

This study uses job seeker dataset from Bojonegoro Regency in 2022, consisting of 260 entries with five variables: age, gender, marital status, education, and district. Before conducting further analysis using the K-Means and Fuzzy K-Means methods, a summary distribution was first carried out to obtain a general overview of each variable under study.

To better understand the composition of job seekers in Bojonegoro Regency, the dataset was analyzed across several demographic variables: age group, gender, marital status, education level, and district of residence. These variables provide valuable context for interpreting the clustering results. Rather than presenting separate frequency tables for each variable, a consolidated summary is offered in Table 2, allowing for a more efficient overview while maintaining analytical depth.

In Figure 1, the data reveal that the majority of job seekers are young (15–24 years), comprising over 70% of the total, and are likely first-time entrants into the workforce. This is consistent with the finding that nearly 90% are unmarried, suggesting limited financial or familial responsibilities, which may influence job preferences and mobility. In terms of gender, male job seekers account for almost 60%, indicating a modest gender disparity in labor force engagement. While not extreme, this pattern may reflect traditional gender roles or varying access to job information and networks. Educational attainment is notably concentrated at the secondary level, with over three-quarters of job seekers having completed Senior or Vocational High School. Only a small portion (7.3%) hold a Diploma, Bachelor's, or Postgraduate degree. This implies a workforce with mid-level qualifications, which aligns with the types of job opportunities commonly available in the region—such as administrative, retail, or technician-level roles.

These findings provide a demographic snapshot that is crucial for both clustering analysis and policy formulation. The dominance of young, unmarried, and secondary school-educated individuals suggests the urgent need for youth-focused employment programs, vocational training, and early-career job placement services. Additionally, the gender composition and educational gaps could guide the design of inclusive job matching platforms and targeted skill development programs. Understanding this demographic baseline also supports more nuanced interpretation of cluster formation later in the study. For instance, the clustering may reflect not only statistical groupings but also real-world segmentation in labor readiness, access, and opportunity.

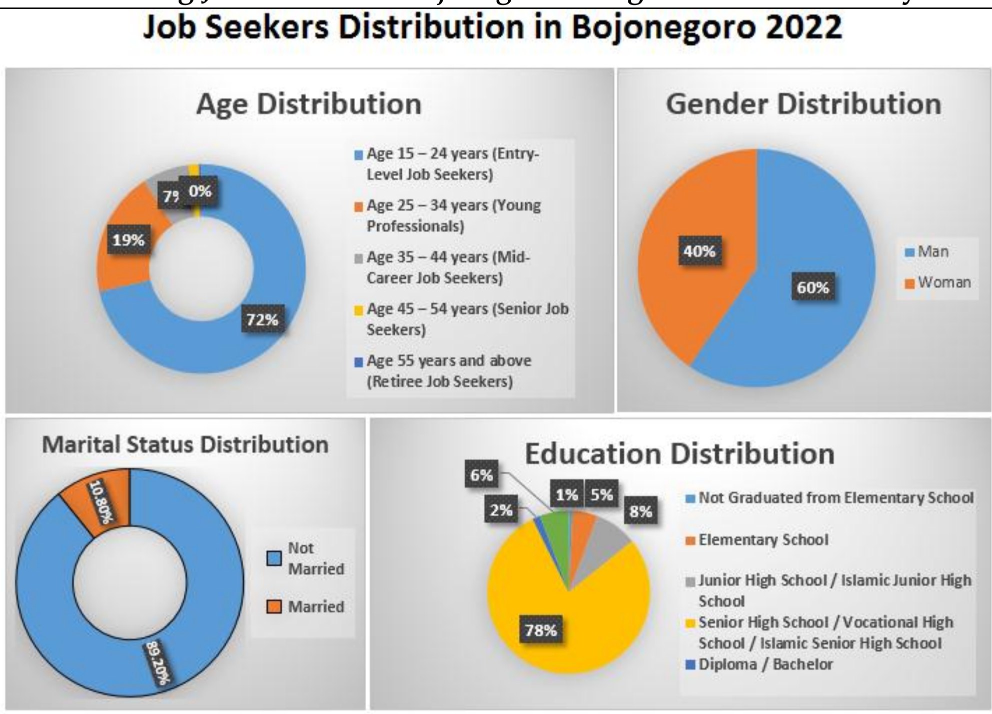

Figure 1. Visualization of the Distribution of Job Seekers by Age, Gender, Education, and Marital Status in Bojonegoro Regency in 2022

In order to understand the geographical distribution of job seekers, their distribution by sub-district of residence was analyzed. This is important to identify areas with high levels of job seekers as well as to assist in making more targeted employment policies at the local level.

Table 2. Results of Frequency Distribution of Job Seekers by Subdistrict.

District	Frequency	Percent
Balen	13	5.0
Baureno	4	1.5
Bojonegoro	13	5.0
Bubulan	1	.4
Dander	21	8.1
Gayam	8	3.1
Gondang	1	.4
Kalitidu	9	3.5
Kanor	4	1.5
Kapas	8	3.1
Kasiman	22	8.5
Kedewan	10	3.8
Kedungadem	34	13.1
Kepohbaru	4	1.5
Malo	9	3.5
Margomulyo	5	1.9
Ngambon	1	.4
Ngasem	4	1.5
Ngraho	20	7.7
Padangan	24	9.2
Purwosari	6	2.3
Sekar	3	1.2
Sugihwaras	8	3.1
Sukosewu	7	2.7
Sumberrejo	6	2.3
Tambakrejo	10	3.8
Temayang	3	1.2
Trucuk	2	.8
Total	260	100.0

The distribution results show that Kedungadem sub-district recorded the highest number of job seekers, with 34 people (13.1% of the total data), followed by Padangan (24 people, 9.2%), Kasiman (22 people, 8.5%), and Dander (21 people, 8.1%). These sub-districts are the main concentration points for job seekers in Kabupaten Bojonegoro. In contrast, some sub-districts such as Bubulan, Gondang, and Ngambon only recorded 1 job seeker (0.4%), indicating a low number of job seekers or possibly limited access to data from these areas.

This distribution shows the unequal geographical distribution of job seekers, which could be caused by factors such as population density, access to job vacancy information, or the dominance of the informal sector in certain areas. This information can be used by local governments to design job training programs that are more focused on sub-districts with high numbers of job seekers, as well as expanding the reach of employment information services in areas with low but potentially increasing numbers of job seekers.

Clustering Method Comparison Process

Before the clustering process, data preprocessing was carried out, including handling missing values, and encoding categorical variables (e.g., age, gender, marital status, and district) to prepare the dataset for clustering algorithms. The clustering method comparison was conducted by comparing the best average within distance values produced by each method. Based on the evaluation results of the K-Means and Fuzzy K-Means clustering, the comparison of both methods is presented in Table 3 below:

Table 3. Results of Clustering Method Comparison Process

	Sum Squared Error	
k	K-Means	Fuzzy K-Mean
2	4205.064	4270.104
3	1628.439	1636.271
4	997.194	1002.481
5	712.204	719.671
6	544.664	556.901
7	479.189	494.481
8	409.036	424.163
9	381.086	357.706
10	322.524	337.158

Table 3 presents the results of a clustering method comparison using the Sum of Squared Errors (SSE) across different values of k (number of clusters), for both K-Means and Fuzzy K-Means algorithms. As the number of clusters increases from 2 to 10, the SSE values for both methods consistently decrease, indicating better fit with higher k. However, based on the elbow method, the optimal number of clusters is determined by identifying the point at which the rate of SSE reduction begins to level off. This "elbow point" represents a balance between model complexity and performance.

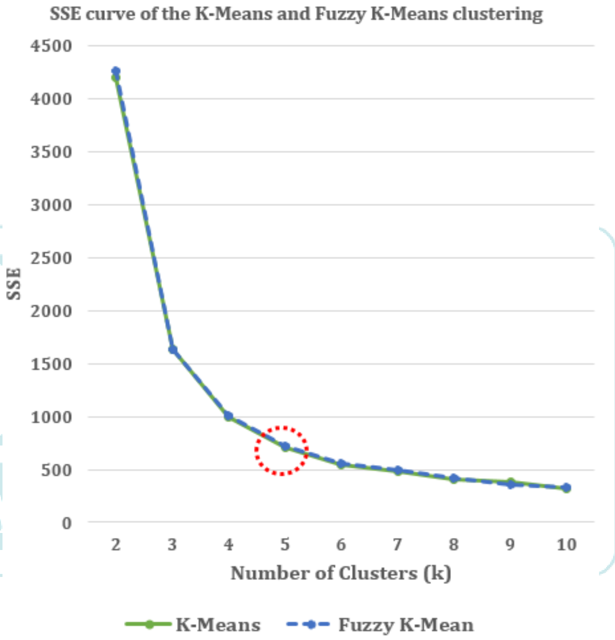

Figure 1. Plot of SSE curve of the K-Means and Fuzzy K-Means clustering.

In this case, the first noticeable Elbow occurs at k = 5, where the decrease in SSE begins to slow down significantly. At this point, the K-Means method yields an SSE of 712.204, and the Fuzzy K-Means method results in 719.671. These values reflect a substantial reduction from previous k-values, especially compared to k = 2 and k = 3. Beyond k = 5, the SSE continues to decrease but at a much slower rate, indicating diminishing returns in clustering performance. Thus, k = 5 is identified as the most appropriate number of clusters for both methods based on the Elbow method applied to the first major drop in the SSE curve. This suggests that segmenting the data into five clusters provides a suitable balance between accuracy and simplicity for further analysis or decision-making.

Frequency Distribution for Each Formed Cluster

The clusters formed can be used as a reference in splitting the data to determine the distribution of each cluster so as to identify the characteristics of each cluster formed.

Table 4. Results of Count of Data for Each Formed Cluster

Cluster	Count of Data
1	44
2	36
3	60
4	62
5	58

Table 4 presents the distribution of data across each cluster formed through the clustering process. Out of a total of 260 data points, Cluster 1 contains 44 data points, Cluster 2 has 36, Cluster 3 includes 60, Cluster 4 has 62, and Cluster 5 consists of 58 data points. This result indicates that the data are relatively evenly distributed among the clusters, although there are slight differences in their sizes. Cluster 2 has the fewest members, while Cluster 4 contains the most. The relatively balanced distribution suggests that the clustering process successfully grouped the data into proportionally sized clusters. Each cluster can be further analyzed to identify its distinct characteristics, which may provide deeper insights depending on the objective of the study such as population segmentation, behavioral grouping, or classification based on similarities.

Table 5. Results of Age of Job Seekers for Each Formed Cluster

	Age	Frequency	Percent
Cluster 1	15 – 24 years	38	86.4
	25 – 34 years	5	11.4
	45 – 54 years	1	2.3
	Total	44	100.0
Cluster 2	15 – 24 years	21	58.3
	25 – 34 years	6	16.7
	35 – 44 years	8	22.2
	45 – 54 years	1	2.8
	Total	36	100.0
Cluster 3	15 – 24 years	40	66.7
	25 – 34 years	15	25.0
	35 – 44 years	5	8.3
	Total	60	100.0
Cluster 4	15 – 24 years	36	58.1
	25 – 34 years	20	32.3
	35 – 44 years	5	8.1
	55 years and above	1	1.6
	Total	62	100.0
Cluster 5	15 – 24 years	51	87.9
	25 – 34 years	4	6.9
	35 – 44 years	1	1.7
	45 – 54 years	2	3.4
	Total	58	100.0

Table 5 presents the age distribution of job seekers across each of the five clusters, categorized into distinct career stages: Entry-Level Job Seekers (15–24 years), Young Professionals (25–34 years), Mid-Career Job Seekers (35–44 years), Senior Job Seekers (45–54 years), and Retiree Job Seekers (55 years and above). The data clearly shows that Entry-Level Job Seekers form the majority in all clusters, indicating that most job seekers are in the early stages of their careers. Cluster 1 is predominantly composed of Entry-Level Job Seekers, accounting for 86.4% of its members, followed by a small proportion of Young Professionals (11.4%) and a few Senior Job Seekers (2.3%). Similarly, Cluster 5 has a very high concentration of Entry-Level Job Seekers at 87.9%, with minimal representation from other age groups. Cluster 3 also follows this pattern, with 66.7% Entry-Level Job Seekers and 25% Young Professionals.

Clusters 2 and 4, however, show greater diversity in age composition. In Cluster 2, while Entry-Level Job Seekers still make up the majority (58.3%), there is a more notable presence of Mid-Career Job Seekers (22.2%)

and Young Professionals (16.7%), along with a small portion of Senior Job Seekers (2.8%). Cluster 4 includes 58.1% Entry-Level Job Seekers and 32.3% Young Professionals, with additional representation from Mid-Career (8.1%) and Retiree Job Seekers (1.6%). In summary, while most clusters are dominated by Entry-Level Job Seekers, Clusters 2 and 4 reflect a more varied demographic, potentially representing more experienced or career-transitioning individuals. These patterns can inform the design of targeted employment programs or interventions tailored to the specific needs of each job seeker group.

Table 6. Results of Gender of Job Seekers for Each Formed Cluster

Gender		Frequency	Percent
Cluster 1	Man	26	59.1
	Woman	18	40.9
	Total	44	100.0
Cluster 2	Man	20	55.6
	Woman	16	44.4
	Total	36	100.0
Cluster 3	Man	36	60.0
	Woman	24	40.0
	Total	60	100.0
Cluster 4	Man	43	69.4
	Woman	19	30.6
	Total	62	100.0
Cluster 5	Man	30	51.7
	Woman	28	48.3
	Total	58	100.0

Table 6 presents the gender distribution of job seekers across the five clusters formed during the clustering process. The results indicate that male job seekers consistently outnumber female job seekers in all clusters, although the degree of gender imbalance varies. In Cluster 1, 59.1% of the job seekers are men, while 40.9% are women. A similar pattern is seen in Cluster 2, where men make up 55.6% and women 44.4%. Cluster 3 also shows a moderate gender gap, with 60% men and 40% women. The most significant gender disparity is found in Cluster 4, where men constitute 69.4% of the cluster, and women represent only 30.6%. On the other hand, Cluster 5 displays the most balanced gender composition, with men making up 51.7% and women 48.3%, suggesting near gender parity.

Overall, while all clusters have a higher proportion of male job seekers, Cluster 5 stands out for its relatively equal gender representation, which may reflect a more diverse or inclusive group. These findings can be useful for developing gender-sensitive employment programs or analyzing potential gender-based barriers in job-seeking behavior.

Table 7. Results of Marital Status of Job Seekers for Each Formed Cluster

Marital Status		Frequency	Percent
Cluster 1	Not Married	42	95.5
	Married	2	4.5
	Total	44	100.0
Cluster 2	Not Married	32	88.9
	Married	4	11.1
	Total	36	100.0
Cluster 3	Not Married	55	91.7
	Married	5	8.3
	Total	60	100.0
Cluster 4	Not Married	48	77.4
	Married	14	22.6
	Total	62	100.0
Cluster 5	Not Married	55	94.8
	Married	3	5.2
	Total	58	100.0

Table 7 displays the distribution of job seekers' marital status across the five clusters. The data reveals that unmarried individuals dominate all clusters, indicating that the majority of job seekers are single. This trend is particularly strong in Cluster 1, where 95.5% of the members are not married, and only 4.5% are married. Similarly, Cluster 5 has 94.8% unmarried individuals, while Cluster 3 and Cluster 2 also show high proportions of unmarried job seekers at 91.7% and 88.9% respectively. Cluster 4 is the most diverse in terms of marital status, with 77.4% not married and 22.6% married, representing the highest percentage of married individuals among all clusters. This may suggest that Cluster 4 includes a relatively older or more experienced group of job seekers, possibly aligned with the age distribution seen in previous tables.

Overall, the data suggests that most job seekers in the sample are unmarried, which could reflect their age demographics—predominantly younger individuals likely at the early stages of their careers. Understanding this pattern is important for tailoring employment support services, as marital status may influence job preferences, availability, and mobility.

Table 8. Results of Education of Job Seekers for Each Formed Cluster

	Education	Frequency	Percent
Cluster 1	Elementary School	1	2.3
	Junior High School / Islamic Junior High School	1	2.3
	Senior High School / Vocational High School / Islamic Senior High School	38	86.4
	Diploma / Bachelor	2	4.5
	Postgraduate	2	4.5
	Total	44	100.0
Cluster 2	Elementary School	4	11.1
	Junior High School / Islamic Junior High School	6	16.7
	Senior High School / Vocational High School / Islamic Senior High School	24	66.7
	Postgraduate	2	5.6
	Total	36	100.0
Cluster 3	Elementary School	2	3.3
	Junior High School / Islamic Junior High School	3	5.0
	Senior High School / Vocational High School / Islamic Senior High School	47	78.3
	Postgraduate	8	13.3
	Total	60	100.0
Cluster 4	Not Graduated from Elementary School	2	3.2
	Elementary School	2	3.2
	Junior High School / Islamic Junior High School	9	14.5
	Senior High School / Vocational High School / Islamic Senior High School	47	75.8
	Diploma / Bachelor	1	1.6
	Postgraduate	1	1.6
	Total	62	100.0
Cluster 5	Elementary School	4	6.9
	Junior High School / Islamic Junior High School	3	5.2
	Senior High School / Vocational High School / Islamic Senior High School	48	82.8
	Diploma / Bachelor	1	1.7
	Postgraduate	2	3.4
	Total	58	100.0

Table 8 illustrates the educational background of job seekers across the five clusters. The data shows that the majority of job seekers in all clusters have attained Senior High School, Vocational High School, or Islamic Senior High School as their highest level of education. This suggests that most job seekers are at an intermediate educational level, suitable for entry-level or semi-skilled job positions. In Cluster 1, a significant 86.4% of individuals have completed senior or vocational high school education, while a small percentage hold a Diploma/Bachelor's degree (4.5%) or Postgraduate degree (4.5%). There are also a few with only elementary (2.3%) or junior high school education (2.3%). Similarly, Cluster 3 reflects a high concentration (78.3%) of senior/vocational high school graduates, along with 13.3% having completed postgraduate education—indicating the presence of some highly educated individuals.

Cluster 2 shows slightly lower educational attainment overall, with 66.7% holding senior high school-level education, while 11.1% only completed elementary school, and 16.7% junior high school. Postgraduate degree holders make up 5.6% of this cluster, suggesting limited higher education representation. Cluster 4 is the most educationally diverse. While 75.8% have completed senior/vocational high school, there are also job seekers who did not graduate from elementary school (3.2%), or have only elementary (3.2%) and junior high school education (14.5%). Only 1.6% of this cluster have a Diploma/Bachelor's degree, and 1.6% have a Postgraduate degree.

The data for Cluster 5 appears to be incomplete, but from the available information, 6.9% have completed elementary school and 5.2% junior high school. The full picture would be clearer with the rest of the cluster's data, especially the proportion of senior high school and higher education graduates. In summary, across all clusters, senior and vocational high school graduates dominate the job-seeking population, indicating a strong presence of mid-level educational attainment. However, Cluster 4 shows the widest range in educational background, from those with no elementary education to postgraduate degrees. These insights can help tailor employment services and training programs based on the specific educational needs of each cluster.

Table 9. Results of Education of Job Seekers for Each Formed Cluster

District		Frequency	Percent
Cluster 1	Gondang	1	2.3
	Kalitidu	9	20.5
	Kanor	4	9.1
	Kapas	8	18.2
	Kasiman	22	50.0
	Total	44	100.0
Cluster 2	Sugihwaras	8	22.2
	Sukosewu	7	19.4
	Sumberrejo	6	16.7
	Tambakrejo	10	27.8
	Temayang	3	8.3
	Trucuk	2	5.6
	Total	36	100.0
Cluster 3	Balen	13	21.7
	Baureno	4	6.7
	Bojonegoro	13	21.7
	Bubulan	1	1.7
	Dander	21	35.0
	Gayam	8	13.3
	Total	60	100.0
Cluster 4	Kedewan	10	16.1
	Kedungadem	34	54.8
	Kepohbaru	4	6.5
	Malo	9	14.5
	Margomulyo	5	8.1
	Total	62	100.0
Cluster 5	Ngambon	1	1.7
	Ngasem	4	6.9

District	Frequency	Percent
Ngraho	20	34.5
Padangan	24	41.4
Purwosari	6	10.3
Sekar	3	5.2
Total	58	100.0

Table 9 shows the geographical distribution of job seekers across the five clusters based on their district of origin. The data reveals that each cluster is associated with specific districts, indicating potential regional patterns in job-seeking behavior or socio-economic characteristics. In Cluster 1, job seekers are concentrated primarily in Kasiman (50.0%), followed by Kalitidu (20.5%) and Kapas (18.2%). This suggests that Cluster 1 may represent individuals from semi-urban or rural areas with similar demographic or employment characteristics.

Cluster 2 is more evenly spread across several districts, with Tambakrejo (27.8%), Sugihwaras (22.2%), and Sukosewu (19.4%) being the top contributors. This diverse spread might indicate a group with varied backgrounds or challenges in job accessibility. In Cluster 3, the highest concentration is in Dander (35.0%), followed by equal contributions from Balen and Bojonegoro (both 21.7%). This cluster likely represents job seekers from districts with moderate economic activity or close proximity to the central Bojonegoro area. Cluster 4 is dominated by Kedungadem, which accounts for more than half of the cluster (54.8%), suggesting a strong regional pattern. Other contributing districts include Kedewan (16.1%) and Malo (14.5%), possibly reflecting similar educational or occupational trends.

Lastly, Cluster 5 features significant representation from Padangan (41.4%) and Ngraho (34.5%), together making up over 75% of the cluster. This points to a highly localized group of job seekers, likely shaped by regional economic factors or specific labor market conditions. In summary, the clustering results show clear geographical segmentation among job seekers, with each cluster being strongly associated with certain districts. These insights can help local governments or employment services to develop targeted programs tailored to the unique needs and challenges of job seekers in specific regions.

CONCLUSION

Conclusion

This study analyzed the characteristics of job seekers in Bojonegoro Regency in 2022 using frequency distribution and clustering methods (K-Means and Fuzzy K-Means). The majority of job seekers are young individuals aged 15–24 years, mostly unmarried, and predominantly male. Most job seekers have completed senior or vocational high school education, indicating a concentration in secondary education level. The clustering analysis identified five distinct groups with relatively balanced sizes, mostly dominated by entry-level job seekers but with some clusters showing more age and educational diversity. K-Means outperformed Fuzzy K-Means in terms of lower Sum of Squared Errors (SSE), indicating better compactness of clusters and clearer internal cohesion among grouped data. These findings highlight the importance of tailoring employment programs to address the needs of young, early-career job seekers, while also considering gender and regional differences. To further validate the clustering results, future studies could incorporate longitudinal or follow-up job-seeking data to assess cluster stability and relevance over time.

Suggestions

Based on the results, it is recommended that local government and employment agencies develop targeted job training and placement programs that focus on entry-level young job seekers with secondary education backgrounds. Efforts should also be made to create gender-sensitive employment policies to reduce the gender gap observed in the job seeker population. Additionally, district-specific interventions could be implemented to address the varied job-seeking dynamics across different regions, especially in districts with higher concentrations of job seekers. Further research is encouraged to examine the post-clustering outcomes and employment trajectories of different groups, particularly older or more educated job seekers, to inform inclusive and sustainable workforce development strategies.

BIBLIOGRAPHY

Abdel-Aty, A., Mostafa, M. G., & El-Sherif, M. (2021). Comparative study of hard and fuzzy clustering algorithms in spatial data analysis. International *Journal of Advanced Computer Science and Applications*, 12(1), 145–152. https://doi.org/10.14569/IJACSA.2021.0120118

Alkodri, A. A., Sujono, S., Isnanto, B., & Maxrizal, M. (2020). Peningkatan Bidang Pendidikan Dan Pelatihan Keterampilan Bagi Pencari Kerja Dengan Basic Graphic Design. *Jurnal Abdimastek (Pengabdian Masyarakat Berbasis Teknologi)*, 1(2). https://doi.org/10.32736/abdimastek.v1i2.1026

Almeida, J. A., de Azevedo, F. M., & Neto, P. A. (2021). Performance analysis of K-Means clustering with categorical data transformation techniques. *Expert Systems with Applications*, 173, 114669. https://doi.org/10.1016/j.eswa.2021.114669

Angela, A., & Honni, H. (2018). Prototipe Aplikasi Mobile Android YourDreamJS untuk Pelamar dan Pencari Kerja. *JBASE - Journal of Business and Audit Information Systems*, 1(1), 19–29. https://doi.org/10.30813/.v1i1.1141

Arnofia, L. (2022). fasilitas pencari kerja yang dilakukan disperinaker Bojonegoro. blokbojonegoro.com. https://blokbojonegoro.com/2022/06/02/fasilitasi-pencari-kerja-ini-yang-dilakukan-disperinaker-bojonegoro/?m=0

Arora, S., & Malik, S. (2020). Clustering techniques for data analysis: A review. *International Journal of Computer Applications*, 176(30), 15–19. https://doi.org/10.5120/ijca2020920672

Bezdek, J. C., Keller, J. M., Krishnapuram, R., & Pal, N. R. (2019). *Fuzzy models and algorithms for pattern recognition and image processing*. Springer. https://link.springer.com/book/10.1007/b106267

Fitriyah, H., Safitri, E. M., Muna, N., Khasanah, M., Aprilia, D. A., & Nurdiansyah, D. (2023). Implementasi Algoritma Clustering Dengan Modifikasi Metode Elbow Untuk Mendukung Strategi Pemerataan Bantuan Sosial Di Kabupaten Bojonegoro. *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 4(3), 1598-1607. https://doi.org/10.46306/lb.v4i3.453

Jain, A. K. (2020). Data clustering: 50 years beyond K-Means. *Pattern Recognition Letters*, 138, 489–499. https://doi.org/10.1016/j.patrec.2020.01.008

Kartini, A.Y., Nurdiansyah, D., & Cahyani, N. (2024). *Metode Statistika*. Madza Media.

Khan, A. S. S., Fatekurohman, M., & Dewi, Y. S. (2023). Perbandingan Algoritma K-Medoids Dan K-Means Dalam Pengelompokan Kecamatan Berdasarkan Produksi Padi Dan Palawija Di Jember. *Jurnal Statistika Dan Komputasi*, 2(2), 67–75. https://doi.org/10.32665/statkom.v2i2.

2301

Kuncoro, J. (2022). Tahun 2022 tercatat 2.063 warga Bojonegoro menganggur. Suarabanyuurip.com. https://suarabanyuurip.com/2023/01/27/tahun-2022-tercatat-2063-warga-bojonegoro-menganggur/

Ningsih, W., & Abdullah, F. (2021). Analisis Perbedaan Pencari Kerja dan Lowongan Kerja Sebelum dan Pada Saat Pandemi Covid-19 di Kota Malang. *Journal of Regional Economics Indonesia*, 2(1), 42–56. https://doi.org/10.26905/jrei.v2i1.6181

Nurdiansyah, D., Hayati, E., Purnamasari, I., Hidayanti, A. A., & Rahayu, Y. F. (2024). Implementation of Clustering and Association for Early Warning of Disasters in Bojonegoro Regency. *ComTech: Computer, Mathematics and Engineering Applications*, 15(2), 119–127. https://doi.org/10.21512/comtech.v15i2.11933

Nurdiansyah, D., Saidah, S., & Cahyani, N. (2023). Data Mining Study For Grouping Elementary Schools In Bojonegoro Regency Based On Capacity And Educational Facilities. *Barekeng: Jurnal Ilmu Matematika Dan Terapan*, 17(2), 1081-1092. https://doi.org/10.30598/barekeng vol17iss2pp1081-1092

Nurdiansyah, D., & Sulistiawan, A. (2023). *Dasar Pemrograman Komputer Dengan Open Source Software R (Untuk Bidang Sains dan Teknologi)*. CV. AA. RIZKY.

Rochman, E. M. S., & Rachmad, A. (2020). Clustering Tourist Destinations Based on Number of Visitors Using the K-Mean Method. *Proceedings of the 1st International Multidisciplinary Conference on Education, Technology, and Engineering (IMCETE 2019)*, 312–314. https://doi.org/10.2991/assehr.k.200303.075

Sallaby, A. F., & Suryana, E. (2018). Penerapan Data Mining untuk Menentukan Jumlah Pencari Kerja Terdaftar Berdasarkan Umur dan Pendidikan Menggunakan K-Means Clustering (Studi Kasus di Dinas Tenaga Kerja Dan Transmigrasi Provinsi Bengkulu). *Journal of Technopreneurship and Information System (JTIS)*, 1(1), 35–38. https://doi.org/10.36085/jtis.v1i2.28

Salman, G. & Kurniati, P. (2022). Pencari Kerja di Surabaya Kini Lebih Mudah Akses Lowongan Pekerjaan lewat Aplikasi ASSIK. Kompas.com.

Simbolon, T. G. (2023). ADA 937.176 Pencari Kerja Di Indonesia Pada 2022. DataIndonesia.id.

Sukirman, S., Rosmiati, R., Alamsyah, N., & Kamal, K. (2020). Perancangan Sistem Informasi Kartu Kuning (Ak.1) Untuk Pencari Kerja Kantor Dinas Ketenagakerjaan Kota Makassar. *ILTEK : Jurnal Teknologi*, 14(2), 2067–2072. https://doi.org/10.47398/iltek.v14i2.418

Wang, Y., Li, J., & Gao, H. (2019). A hybrid fuzzy K-means clustering algorithm for data with outliers. *Expert Systems with Applications*, 124, 291–300. https://doi.org/10.1016/j.eswa.2019.01.016

Xu, C., & Tian, Y. (2019). A comprehensive survey of clustering algorithms. *Annals of Data Science*, 6(2), 221–252. https://doi.org/10.1007/s40745-018-0149-1

Yuliani, R., Setiawan, A., & Prasetyo, A. (2020). Analysis and design of job information systems for employment services. *Jurnal Teknologi dan Sistem Komputer*, 8(4), 237–244. https://doi.org/10.14710/jtsiskom.8.4.2020.237-244