AJST
journal.amca2012.org

AMCA JOURNAL OF SCIENCE AND TECHNOLOGY
e-ISSN 2774-8324
Vol 3, No 2, August 2023, 5-7
DOI 10.51773/ajst.v3i2/274

Myedusolve personality clustering analysis using k-means machine learning algorithm

Sapto Hadi Riono[1*], Rizdania[1], Dimas Rega Hadiatullah[1], and Muhammad Rizky[1]

[1]Ilmu Komputer, Universitas PGRI Wiranegara, Indonesia
[*]Corresponding author: saptoenator@gmail.com

KEYWORDS
clustering
k-means
machine learning
myedusolve
personality

ABSTRACT Personality is a characteristic of a person in carrying out and responding to something and adapting to their environment. This personality difference can affect everyone in running a program or something that is being done. MyEdusolve is a technology company engaged in education, with a vision to realize the transformation of the Indonesian workforce through digital literacy, has been followed by more than 40,000 participants with several selected programs. The suitability of personality with the program being followed is an important factor in the success of participants to take part in the program held. This research is focused on analyzing and creating personality clustering on programs held by MyEduSolve. Clustering is done using the K-Means machine learning algorithm. Through this analysis, it is hoped that companies can carry out activities or increase branding programs according to personality types and the results of analysis on data so that participants are helped to achieve and prepare careers according to the programs held by the company.

1. INTRODUCTION

Personality is a characteristic or trait that reflects a person's actions in doing something. Each person has a different and diverse personality so the way a person responds to problems in his daily life tends to vary. As is the case when studying, some personality types tend to get bored quickly with organized and scheduled learning methods, and some are the opposite. From here it can be seen that personality is a very important factor and thing for someone to adjust to their environment (Dinata et al, 2020), to provide good encouragement and motivation in doing something. The typical learning of a person generally tends to be very diverse, so learning methods and components are very important in the success of a person to learn and understand something. In the digital era, there are now many online learning platforms with very applicable and interactive learning systems and methods, one of which is the digital education company *MyEduSolve*. *MyEduSolve* is a technology company engaged in education, to realize the transformation of the Indonesian workforce through digital literacy, by improving the quality of the Indonesian workforce through international certification and online learning.

MyEduSolve has more than 180 online classes taught by more than 50 professional instructors with approximately 40,000 participants who have joined the company's program. With so many registered participants, the personalities of each participant also vary; clustering or grouping of the personalities of each participant in the data must be done to provide exciting learning facilities or concepts and facilities in each program to comfort and help participants in carrying out learning on the platform. One method of clustering data can use the k-means machine learning algorithm. K-means is a non-hierarchical data clustering method that groups data in one or more groups so that groups with the same characteristics will be in one cluster, while if they have different characteristics they will be in different clusters: (Dinata et al, 2020). The clustering can be known by the amount of data in the cluster center based on the calculation of its distance from the cluster center.

In this study, the personality types to be analyzed are explorer and career-ready personality types. An explorer is the attitude or behaviour (personality) of an adventurous person who is creative, always curious, enthusiastic, and has an interest in specific fields or interests broadly and deeply. Individuals who have more interest in certain things than other individuals (Suminta , 2022). tend to take risks, think fast, change their piker patterns with little regret or guessing, and they tend to be competent to be social. Career readiness is a level and development of competence in a person that allows a person to recognize and be able to overcome any problems related to work or career. These two personality types are very suitable for individuals who will start a career, where individuals will tend to learn and will explore the potential of their interests and talents and learn new things that are by the career or goal to be addressed. By clustering the 2 personalities, it is hoped that it can provide insight or information from the cluster correlation with other column variables in the data so that companies can make policies by the business being run.

2. METHOD

The research conducted was completed through research stages which were divided into 4 stages, namely: 1. Data pre-processing, 2. Clustering, 3. Data analysis, 4. Conclusion. The dataset used is a company dataset obtained through a survey of participants participating in a program held by the company, which consists of age, gender, program choice, place of residence and questions about self-development.

3. RESULTS AND DISCUSSION

3.1 Results

The tools used are Python programming language using *Google Collab* for pre-processing and clustering and *Microsoft Excel* for data analysis.

Data pre-processing is a stage before clustering and analysis which is used to clean, delete, or change data sources in the form of non-alphabetic character data or unused words (Muttaqin , 2016). In this study, data pre-processing was carried out with several stages, namely

3.1.1 Remove missing values

Missing value is the loss of specific attribute values in the data, and missing values can be caused by errors when data collection is carried out, errors when entering data or the lack of respondents' ability to provide answers to questions (Arifianto et al, 2022).

3.1.2 Delete duplicated

Remove duplicate is used to remove the same value in each data.

3.1.3 Change the data type

In applying the K-means algorithm, one of the data types that can be used is integer, so some of the age column data types must be converted to integer so that clustering and analysis can be carried out.

3.1.4 Scale adjustment

At this stage, the scale of questions from the data obtained will be changed by adjusting the type of personality to be created, which will be processed by using the number on the question.

3.1.5 Feature selection

Feature Selection is a data modelling process to select influential features, where at this stage the division of features is carried out for clustering 2 explorer and career-ready personalities (Riono, 2022). Modelling is done by dividing the features in the data to be used where question numbers 5-31 for explorer and 32-end for career ready.

3.1.6 Data normalization

In data normalization, the process carried out in this study is to equalize the feature scale range according to the two personality types where -4 will be worth = 0, and -1 will be worth = 1.

The next stage is clustering, clustering is a method used to divide a collection of data into several groups or groups according to predetermined similarities (Metisen & Sari, 2015). The cluster data will be divided into 2 personality aspects, namely explorer and career-ready.

The next stage is analysis in Figure **??**, analysis is the process of decomposing a whole into components that can recognize the signs of the components, know the relationship with each other and function in one whole into the smallest unit (Septiani et al, 2020). The analysis carried out is to find a correlation between 2 explorer personalities and career readiness with age, place of residence, program of choice and gender.

After the data is complete, the next stage is to provide conclusions and suggestions, where the information results from data analysis are used to provide advice and input to company policies in running the business.

3.2 Discussion

3.2.1 Personality types by age

The number of those who are not career-ready and do not have an explorer personality is still greater than those who are career-ready and have an explorer personality. It can be seen that the most significant number is in the age range of 16-25 years.

3.2.2 Personality types by gender

The number of participants who are still not career-ready and do not have an explorer personality is still more than those who do, both for males and female gender. have an explorer personality is still more than those who have, both for male and female gender. The number of female participants who participated in the program created by *MyEdu-Solve* was more than the number of male participants, with a total of 2413: 3780

3.2.3 Personality types based on place of residence

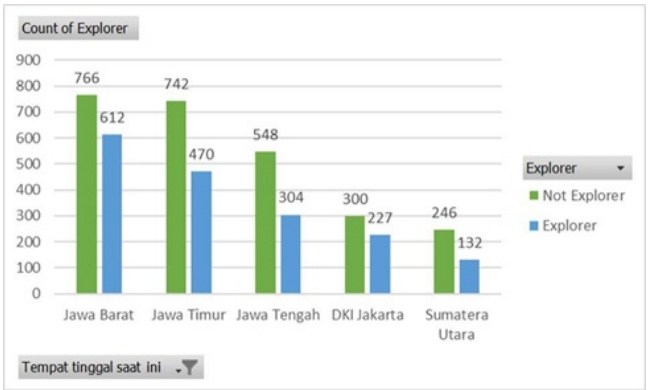

FIGURE 1. Personality explorer and program options

The number of participants who have a non-explorer personality and are not career-ready is more than the explorer personality and career-ready, and it can be seen that the top 3 positions are large provinces on the island of Java (West Java, East Java, and Central Java).

3.2.4 Personality types based on elective programs

In the personality type based on the program of choice, it can be seen that participants who are not career-ready with a typical not explorer tend to be high in the Microsoft

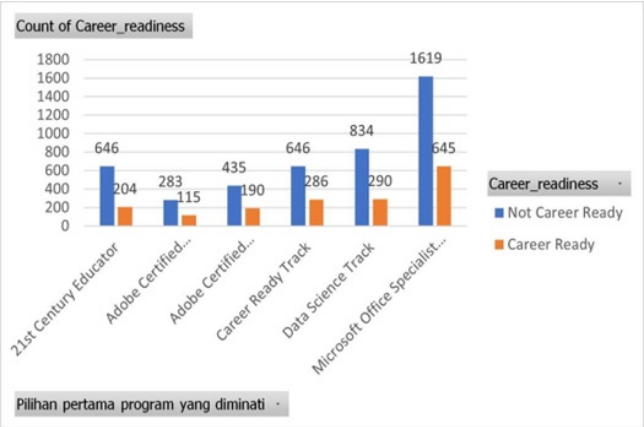

FIGURE 2. Personality explorer and program options

Office specialist program, as well as career-ready participants with a typical explorer have high scores in the same program, namely Microsoft office specialist.

3.2.5 Participants by age and gender

The lowest number of participants was found in East Nusa Tenggara, Maluku, West Papua, Papua, and North Kalimantan. This might happen because education platforms like *MyEduSolve* exist in those areas. Alternatively, maybe the name MyEduSolve is still not widely known by people from these regions. The age of users from the area varies from teenagers to over 26 years old. The age of *MyEduSolve* program participants is mostly 20 to 22 years old. These ages are generally students who are in their final year. It could be that these participants are students who are preparing to enter the career world.

4. CONCLUSION

Based on the results of research and analysis conducted, the number of those who are not career-ready and do not have explorer personality is still more than explorer personality, followed by participants with female gender with the dominant age of participants between 16-25 years. Java Island with the 3 largest provinces of West Java, Central Java and East Java are in the number of those who are not ready for a career and do not have an explorer personality, while the Indonesian part of the country is not yet ready for a career. Explorer personality, while the eastern and central parts of Indonesia are partly the regions with the smallest number of participants in the program. The Microsoft Office Specialist Expert program is the most popular program for both personality types.

References

Arifianto, A. S., Safitri, K. D., Agustianto, K., & Wiryawan, I. G. (2022). Pengaruh Prediksi Missing Value Pada Klasifikasi Decision Tree C4.5. *Jurnal Teknologi Informasi Dan Ilmu Komputer* (JTIIK), 9(4). https://doi.org/10.25126/jtiik.202294778

Dinata, R. K., Safwandi, S., Hasdyna, N., & Azizah, N. (2020). Analisis K-Means Clustering pada Data Sepeda Motor. INFORMAL: *Informatics Journal*, 5(1), 10–17. https://doi.org/10.19184/isj.v5i1.17071

Metisen, B. M., & Sari, H. L. (2015). Analisis Clusteringmenggunakan Metode K-Means Dalam Pengelompokkan Penjualan Produk Pada Swalayan Fadhila. *Jurnal Media Infotama*, 11(2). https://doi.org/10.37676/jmi.v11i2.258

Muliono, R., & Sembiring, Z. (2019). Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen. CESS (*Journal of Computer Engineering, System and Science*), 4(2), 272–279. https://doi.org/10.24114/cess.v4i2.13620

Muttaqin, F. A. (2016). Implementasi Text Mining Pada Aplikasi Pengawasan Penggunaan Internet Anak [Diploma, Universitas Komputer Indonesia]. http://elib.unikom.ac.id

Riono, S. H. (2022). Optimization of Naïve Bayes Based on Genetic Algorithm for Performance Evaluation of Lecturers of PGRI Wiranegara University", Program Studi Ilmu Komputer Universitas PGRI Wiranegara. SAINTEKBU, 14(02), Article 02. https://doi.org/10.32764/saintekbu.v14i02.3515

Sari, H. & Shabri. (2016). Hubungan Tipe Kepribadian Dengan Motivasi Belajar Pada Mahasiswa Kurikulum Berbasis Kompetensi Fakultas Keperawatan Universitas Syiah Kuala. *Idea Nursing Journal*, 7(2). https://doi.org/10.52199/inj.v7i2.6449

Septiani, Y., Aribbe, E., & Diansyah, R. (2020). Analisis Kualitas Layanan Sistem Informasi Akademik Universitas Abdurrab Terhadap Kepuasan Pengguna Menggunakan Metode Sevqual. *Jurnal Teknologi Dan Open Source*, 3(1), 131–143. https://doi.org/10.36378/jtos.v3i1.560

Suminta, R. R. (2017). Hubungan Antara Tipe Kepribadian Dengan Orentasi Religiusitas. *Fikrah*, 4(2), 214. https://doi.org/10.21043/fikrah.v4i2.1884