

# Clustering Data On Underage Marriage Using The Clustering Method

Sonadi Perangin angin <sup>1</sup>, Evitha <sup>2</sup>, Marto Sihombing <sup>3</sup>

<sup>1,2,3</sup>Sistem Informasi, STMIK Kaputama, Binjai, Sumatera Utara  
[sonadipa003@gmail.com](mailto:sonadipa003@gmail.com)<sup>1\*</sup>, [evitha12014@gmail.com](mailto:evitha12014@gmail.com)<sup>2</sup>, [martosihombing45@gmail.com](mailto:martosihombing45@gmail.com)<sup>3</sup>

## Abstract

In Law no. 1 of 1974, article 7 paragraph (1) states that marriage is only permitted if the man has reached the age of 19 and the woman has reached the age of 16. Nationally, early marriage to the age of under 16 is 26.95%. In fact, based on the findings of Bappenas in 2008, it was stated that 34.5% of the 2,049,000 marriages in 2008 until now were child marriages which were increasing rapidly (Rifiani, 2011: 126). The influence of foreign culture is also one of the causes of the large number of underage marriages, foreign cultures which are very famous for freedom of dating, are the views of today's youth to have relations outside of legal marriage. Not only culture, information technology in the 4.0 era has greatly influenced the occurrence of underage marriages, adult video sites that are easily accessible via the internet. For this reason, the K-means clustering method is used as the right solution for the problem of underage marriage data by grouping the data based on age, gender, and occupation to get definite data, so that data grouping using the application *matlab* and *rapid miner* can produce output from data mining that can be used in making decisions in the future

**Keywords:** *underage marriage, clustering, matlab*

## 1. Background problem

Nationally, early marriage to the age of under 16 is 26.95%. In fact, based on the findings of Bappenas in 2008, it was stated that 34.5% of the 2,049,000 marriages in 2008 until now were child marriages which were increasing rapidly (Rifiani, 2011: 126). The influence of foreign culture is also one of the causes of the large number of underage marriages, foreign cultures which are very famous for freedom of dating, are the views of today's youth to have relations outside of legal marriage. Not only culture, information technology in the 4.0 era has greatly influenced the occurrence of underage marriages, adult video sites that are easily accessible via the internet.

The role of parents is very important to educate children not to have relations that are out of bounds which results in underage marriages. Meanwhile, making a marriage certificate is not easy because the age requirement does not meet the criteria for marriage. One way to find out the data on underage marriages is by grouping the data of the prospective bride and groom based on age, gender, and occupation. With these problems, the clustering method can be used as the right solution for the problem of underage marriage data by grouping the data based on age, gender, and occupation to get definite data, so that data grouping using the application *matlab* and *rapid miner* can produce output from data mining that can be used in making decisions in the future.

## 2. Theoretical Basic

According to (Isnaini & Sari, 2019) Underage marriage is a marriage carried out by a couple or one of the partners is still categorized as a teenager under the age of 19 years. The practice of early marriage is most common in Africa and Southeast Asia.

Data Mining is often also called Knowledge Discovery in Database, is an activity that includes the collection, use of historical data to find regularities, patterns or relationships in large data sets. According to Alfa Saleh, (2015: 208), explains that: "Data Mining is a process or activity for collecting large data and then extracting this data into information that can later be used" (Mining et al., 2022).

### 2.1 Data Mining Techniques

According to Siska Haryati, Aji Sudarsono, Eko Suryana (2015) There are several data mining techniques that are often mentioned in the literature. However, there are 3 popular data mining techniques, namely:

1. Association Rule Mining

Association Rule mining is a mining technique to find associative between attribute combinations. An example of an associative rule from purchasing analysis in a supermarket is to arrange the placement of goods or design a marketing strategy by using discount coupons for certain combinations of goods.

## 2. Clustering

In contrast to association rule mining and classification where data classes have been predetermined, clustering can be used to label unknown data classes. Because of that clustering is often classified as an unsupervised learning method. The principle of clustering is to maximize the similarity between clusters. Clustering can be done on data that has several attributes that are mapped as a multidimensional space.

## 3. Classification

In classification, there is a target categorical variable. For example, the classification of income can be separated into three categories, namely high income, moderate income, and low income.

## 2.2 Clustering

Clustering is a data analysis method, which is often included as a data mining method, the purpose of which is to group data with the same characteristics. According to Hermawati (2012, p: 123) "Cluster analysis (clustering) is finding objects in one group that are the same (have a relationship) with others and are (points have a relationship) with objects in other groups". The main purpose of the clustering method is to group a number of data/objects into clusters (groups) so that each cluster contains data that is as similar as possible. The clustering method seeks to place similar (closely spaced) objects in one group and make the distance between groups as far as possible. This means that objects in one group are very similar to each other and different from objects in other groups.

## 3. Research Methods

The research method is a description of the steps so that the research can be carried out in a structured manner, so a framework is prepared starting from the beginning until the final results are achieved, while the steps are, describing the problem, studying the theory, conducting data collection, conducting data analysis, testing and implementation, and evaluation

**Table 1:** Data Yang Akan diolah

Age	Gender	Occupation
17	L	wiraswasta
16	P	Ibu Rumah Tangga
18	L	wiraswasta
17	P	Ibu Rumah Tangga
17	L	wiraswasta
16	P	Ibu Rumah Tangga
16	L	wiraswasta
16	P	Ibu Rumah Tangga
17	L	wiraswasta
15	P	Ibu Rumah Tangga
16	L	wiraswasta
15	P	Ibu Rumah Tangga
17	L	wiraswasta
16	P	Ibu Rumah Tangga
18	L	wiraswasta
16	P	Ibu Rumah Tangga
16	L	wiraswasta
15	P	Ibu Rumah Tangga
17	L	Wiraswasta
16	P	Ibu Rumah Tangga

The above data can be processed using the method clustering algorithm k-means, then data of nominal and non-nominal data types such as Age, Gender, and Occupation must first be initialized in the form of numbers. The grouping of data on underage marriages can be expressed in a number of variables independent namely Age (X), Gender (Y), Occupation (Z).

To determine group from one object, the first thing to do is measure the distance Euclidean between two points or objects or X, Y and Z.

Docluster into 3 groups (K=3) and determine the center point centroid. As for the calculation process clustering as follows:

Iteration I

Centroid 1 = (3,1,1) taken randomly from data 1

Centroid 2 = (2,2,2) taken randomly from data 2

Centroid 3 = (4,1,1) taken randomly from data 3

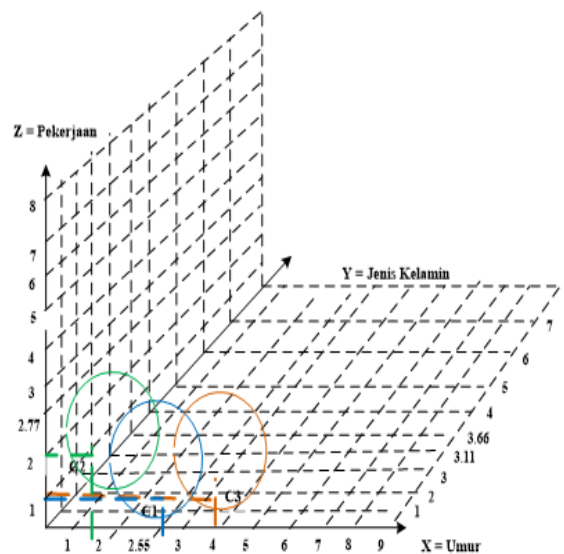


Figure 1: Graph of underage marriage data

## 4. Results and Discussion

As for the implementation results that have been carried out on the program or coding so that it can be seen whether it has fulfilled the objectives to be achieved. Implementation that will be carried out in application programming MATLAB by using the method Clustering so that the system can carry out the process of grouping data and also results cluster with centroid obtained and subsequent testing using the RapidMiner application.

Information :

Of the 1000 data on underage marriages, 3 were obtained cluster Where,

- cluster 1 totals 204 data ,
- cluster 2 amounted to 235 data, and
- cluster 3 totaling 79 data.

The following is a description of the center cluster on the graph:

1. 3.38 1.45 1.45

It can be seen that at cluster 3 Group data on underage marriages at group age is 17 is by Gender Male and Occupation is self-employed.

2. 5.48 1.53 1.53

It can be seen that at cluster 3 Group data on underage marriages at group age is 19 is with Gender Male and Occupation is self-employed.

3. 7.50 1.02 1.02

It can be seen that at cluster 3 Group data on underage marriages at group age is 21 is with Gender Male and Occupation is self-employed.

Comparative Analysis of Results

Based on the results of the process Clustering use MATLAB and RapidMiner, then it can be seen in the following information:

1. For comparative data results

Matlab :

- cluster 1 totals 217 data ,
- cluster 2 totaled 204 data, and
- cluster 3 totaling 79 data.

Rapidminer:

- cluster 1 total 91 data ,
- cluster 2 amounted to 225 data, and
- cluster 3 totaling 184 data.

**For Graph and Plot results**

- By using Matlab, results from the chart centroid it can be found that the criteria obtained are ages 19, 17, and 21 for the criteria of sex being more dominantly male, and work is self-employed.
- By using Rapidminer, the results of the plot are obtained for the age criteria of 17, 19, and 21 for the criteria of a male predominant sex and for work is self-employed.
- With Rapidminer testing cluster 4, that a comparison of the results of the data is obtained cluster 1 (67), cluster 2 (137), cluster 3 (79) and cluster 4 (217), with the most dominant criteria being the criteria with cluster 2.
- With Rapidminer testing cluster 5, that a comparison of the results of the data is obtained cluster 1 (90 data), cluster 2 (118 data), cluster 3 (91 data), cluster 4 (166 data), and cluster 5 (35 data).

Based on the results of the comparison above, it can be concluded that each of the applications has different results, be it rapiminer or matlab, therefore the results obtained will be a comparison and also a reference for writers and also produce information according to needs. necessary and will be implemented in the future

## 5. Conclusion

1. By using the methodclustering it will get results in the form of new information using 500 data that has been processed and also tested by the system..
2. By implementing usingmatlab andRapidminerthen it has succeeded in obtaining the results of grouping underage marriage data with the results of each of each application.
3. With the design that has been made, it becomes a designinterface and a system that can cluster underage marriages and solve the problems in this study.

## References

- [1] Ardiansyah, D., & Walim, W. (2018). Algoritma c4.5 untuk klasifikasi calon peserta lomba cerdas cermat siswa smp dengan menggunakan aplikasi rapid miner. *Jurnal Inkofar*, 1(2), 5–12..
- [2] Resdati, O., Hasanah, R., & Sosiologi, J. (2021). KENAKALAN REMAJA SEBAGAI SALAH SATU BENTUK PATOLOGI SOSIAL (PENYAKIT MASYARAKAT). In *JCI Jurnal Cakrawala Ilmiah* (Vol. 1, Issue 3). <http://bajangjournal.com/index.php/JCI>.
- [3] Cahyono, B. (2016). Penggunaan Software Matrix Laboratory (Matlab) Dalam Pembelajaran Aljabar Linier. *Phenomenon : Jurnal Pendidikan MIPA*, 3(1), 45–62. <https://doi.org/10.21580/phen.2013.3.1.174>.
- [4] Darmi, Y., & Setiawan, A. (2016). Penerapan Metode Clustering K-Means Dalam. *Y. Darmi, A. Setiawan*, 12(2), 148–157.
- [5] Haryati, S., Sudarsono, A., & Suryana, E. (2015). Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama*, 11(2), 130–138.