JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 161 - 166 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.2045
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

IMPLEMENTATION OF DATA MINING BY USING K-MEANS TO CLASSIFY MARRIAGE AGE

Wiwin Handoko[1*], Auliana Nasution [2]
[1]Sistem Informasi, Sekolah Tinggi Manajemen Informatika dan Komputer Royal
[2] Informatika, Universitas Battuta
email: * win.van.handoko@gmail.com

Abstract: Marriage is a husband and wife relationship between a man and a woman to form a family. There are several conditions in marriage that must be fulfilled both religiously and legally in force in Indonesia. To carry out the marriage, the prospective bride and groom must register at the nearest Religious Affairs Office (KUA), KUA is an institution established by the government to handle marriage matters. At marriages, various age groups are often found registering at the KUA. This research was conducted using the Data Mining technique through the K-Means Clustering Model to determine the age grouping of marriage which aims to make it easier for the Office of Religious Affairs in educating the prospective bride and groom from a future perspective and an economic perspective in terms of having a child. The research dataset is data on prospective wedding brides at KUA Rawang Lama, Panca Arga in 2022 with a total of 102 samples, by forming 3 clusters, namely: the Ideal cluster of 76 prospective wedding brides (age 19-30 based on husband's age and age 18-25 based on age wife), a good cluster of 20 prospective marriage brides (age 28-44 based on husband's age and age 24-37 based on wife's age), and a risky cluster of 6 prospective marriage brides (age 49-72 based on husband's age and age 39-58 based on wife's age), and produces a Silhouette Score of 0.57.

Keywords: clustering; data mining; k-means; marriage

Abstrak: Pernikahan merupakan hubungan sebagai suami dan istri antara seorang laki-laki dan perempuan untuk membentuk sebuah keluarga. Terdapat beberapa syarat dalam pernikahan yang wajib dipenuhi baik secara agama maupun secara hukum yang berlaku di Indonesia. Untuk melakukan pernikahan, calon kedua mempelai harus mendaftarkan diri pada Kantor Urusan Agama (KUA) terdekat, KUA merupakan lembaga yang dibentuk oleh pemerintah untuk menangani masalah pernikahan. Pada pernikahan sering ditemukan berbagai kalangan umur yang mendaftarkan diri di KUA. Penelitian ini dilakukan dengan menggunakan teknik *Data Mining* melalui Model *K-Means Clustering* untuk menentukan pengelompokkan umur pernikahan yang bertujuan untuk mempermudah pihak KUA dalam mengedukasi calon mempelai pernikahan dalam sudut pandang masa depan dan sudut pandang ekonomi dalam hal memiliki seorang anak. Dataset penelitian ini adalah data calon mempelai pernikahan pada KUA Rawang Lama, Panca Arga pada tahun 2022 sebanyak 102 sampel, dengan membentuk 3 klaster yaitu : *cluster Ideal* sebanyak 76 calon mempelai pernikahan (usia 19-30 berdasarkan umur suami dan usia 18-25 berdasarkan umur istri), *cluster* baik sebanyak 20 calon mempelai pernikahan (usia 28-44 berdasarkan umur suami dan usia 24-37 berdasarkan umur istri), dan *cluster* beresiko sebanyak 6 calon mempelai pernikahan (usia 49-72 berdasarkan umur suami dan usia 39-58 berdasarkan umur istri), dan menghasilkan *Silhouette Score* 0.57.

Kata kunci: clustering; data mining; k-means; pernikahan

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 161 - 166 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.2045
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

Information technology is always developing rapidly. This development is the right opportunity to obtain more effective and efficient but diverse data. To process this data, a technique is needed so that the processing results or information obtained are appropriate. One technique that can be used is data mining[1].

The Office of Religious Affairs (KUA) is the frontline work unit of the Ministry of Religion which carries out governmental duties in the field of Islamic Religion, in the District area. It is said to be the foremost work unit because the Office of Religious Affairs (KUA) directly deals with the community. Because of that, it is only natural that the existence of the Office of Religious Affairs (KUA) is considered very urgent along with the existence of the Ministry of Religion[2].

The fact, there are still some people who do not understand the duties and functions of the Office of Religious Affairs (KUA). The result is not surprising, there is an impression that the duties and functions of the Office of Religious Affairs (KUA) are only limited to reading prayers and marrying off[3].

KUA Rawang Lama, Panca Arga is one of the religious affairs offices that handles various marriages where many age groups register their marriages both religiously and officially in the state. However, from the information available, the age of marriage is 19 years old who can legally marry in the country which has been regulated in the laws currently in force. Marriage under the age of 19 can only be married according to religion with various conditions and considerations that have been agreed upon. In 2022 at the KUA Rawang Lama, Panca Arga, there are 102 marriage registrars, the age groups who register marriages at the KUA are: ages 19-72 for men and ages 18-58 for women.

The issue of marriage related to age limits, namely underage or even below the minimum age for marriage is a complex discourse related to both legal and non-legal aspects. In this regard, the question of marriage collides diametrically with legal provisions that stipulate a minimum age for marriage[4]. This problem lies in how mentally prepared one has to live a household life with a partner. Not ready mentally and materially, it's better to postpone wedding plans in advance. Instead of living life as an unhappy husband and wife.

It is common knowledge that the ideal age for marriage in Indonesia is less than 20 years, especially for women[5]. With the data owned by the KUA Rawang Lama, Panca Arga, it fits the grouping of three categories, so the categories are ideal, good, and risky in determining the division of marriage ages.

Tahaga (Ketahanan Keluarga) is also known as the strength or resilience family. This is related to personal and family abilities to utilize their potential to face life's challenges, including the ability to restore family functions to their original state in facing challenges and crises[6]. This problem affects the need for housing to the cost of raising a child and reducing divorce cases, which often occur due to economic problems and readiness for a household.

METHOD

Data mining and machine learning techniques can be used to make predictions based on past data. Data mining is the process of finding useful patterns in large data sets. From other sources, data

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 161 - 166 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.2045
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

mining is the study of collecting, cleaning, processing, analyzing, and obtaining useful insights from data[7]. This data was obtained from KUA Rawang Lama, Panca Arga and has not gone through a cleaning process, so a lot of data is private and must be deleted.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 102 entries, 0 to 101
Data columns (total 11 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Tanggal Daftar        102 non-null    object
 1   NIK Suami             102 non-null    int64
 2   Nama Suami            102 non-null    object
 3   Tempat Lahir Suami    102 non-null    object
 4   Tanggal Lahir Suami   102 non-null    object
 5   Umur Suami            102 non-null    int64
 6   NIK Istri             102 non-null    int64
 7   Nama Istri            102 non-null    object
 8   Tempat Lahir Istri    102 non-null    object
 9   Tanggal Lahir Istri   102 non-null    object
 10  Umur_Istri            102 non-null    int64
dtypes: int64(4), object(7)
memory usage: 8.9+ KB
```

Image 1. Decription of Data KUA

There are several different approaches classified as information seeking techniques in KDD. There are quantitative approaches, such as proballistic and statistical approaches.

Several approaches make use of visualization techniques, classification approaches such as inductive logic, decision tree analysis and pattern finding. Other approaches include genetic algorithms, trend analysis, artificial neural networks, deviations and a mixed approach of two or more of the existing approaches[8].

The first step in the data mining process is data cleaning. the activities carried out in this process are checking inconsistent data, correcting errors in data and removing data duplication[9].

The dataset below has gone through a cleaning process, which deleted some unnecessary data such as tanggal daftar, NIK suami, tempat lahir suami, NIK istri, tempat lahir istri where this data is very private and may not be published.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 102 entries, 0 to 101
Data columns (total 6 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Nama Suami           102 non-null    object
 1   Tanggal Lahir Suami  102 non-null    object
 2   Umur_Suami           102 non-null    int64
 3   Nama Istri           102 non-null    object
 4   Tanggal Lahir Istri  102 non-null    object
 5   Umur_Istri           102 non-null    int64
dtypes: int64(2), object(4)
memory usage: 4.9+ KB
```

Image 2. Decription of Data Used

Clustering model analysis is a technique of multivariable analysis that is used to group objects (variables or data) that have similarities into one group so that they can produce information in testing the object and then present a hypothesis based on the relationships that occure[10].

Centroid is the data center point for calculating the vector mean as a centroid. In applying the K-means algorithm, the midpoint or centroid value is generated from the data obtained from each cluster[11]. Clustering steps contained in the K-Means algorithm[12].

```
# Menjalankan K-Means Clustering ke dataset
from sklearn.cluster import KMeans

km = KMeans(n_clusters=3, random_state=0)
km
```

```
KMeans(n_clusters=3, random_state=0)
```

Image 3. K-Means Clustering Method

The first step of the K-Means algorithm is to determine the number of clusters, in this study 3 clusters were determined[13]. Namely the ideal cluster (C0), the risky cluster (C1), and the good cluster (C2).

use the Euclidean Distance formula to calculate the distance of each input data to each centroid until the closest

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. IX No 2, Maret 2023, hlm. 161 - 166 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v9i2.2045
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

distance of each data to the centroid is found. Euclidean Distance equation:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x - y)^2} \quad (1)$$

Information :
d(x,y): the distance between the data at the position of the x and y points
x: position of first data point (cluster center)
y: second data point position (data from n)
n: the number of data attributes

RESULT AND DISCUSSION

The initial stage for analysis is importing the required modules and the dataset of the Office of Religious Affairs (KUA).

	Nama Suami	Tanggal Lahir Suami	Umur_Suami	Nama Istri	Tanggal Lahir Istri	Umur_Istri
0	ARIS MUNANDAR	2000-05-04	21	CYNDI RAHMADANI BR SINAGA	2001-12-04	20
1	HENDRA	26-11-1984	37	RATNA DEWI SIRAIT	04-03-1982	29
2	LIAM	1998-02-14	23	DELA PUSPITA	2002-05-31	19
3	WAHYU NURHADI	1996-07-30	25	NIA SAFITRI	1998-01-30	22
4	SUGIARIAN	1998-06-15	23	RAHAYU MINGSIH	1998-08-01	22

Image 4. Datasets

After building a model with cluster data, use the cluster data to perform tests. The centroid gain for the K-Means method is:

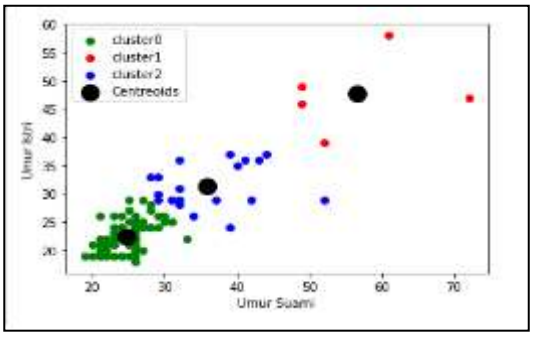

Image 5. Visualization Centroid

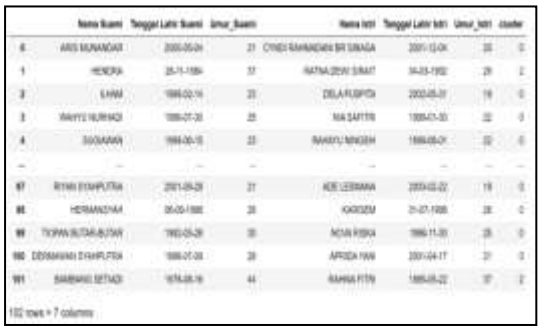

Image 6. View Cluster

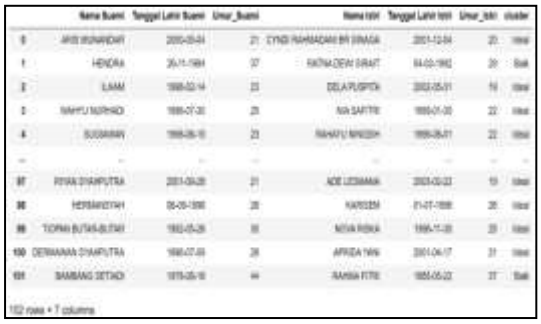

Image 7. Name of Cluster

Based on Figure 7, there are 3 clusters, namely: The Ideal Cluster of 76 prospective wedding brides (age 19-30 based on husband's age and age 18-25 based on age wife), A Good Cluster of 20 prospective marriage brides (age 28-44 based on husband's age and age 24-37 based on wife's age), and A Risky Cluster of 6 prospective marriage brides (age 49-72 based on husband's age and age 39-58 based on wife's age)

```
#Evaluasi Model silhouette_score (Nilainya Antara -1 hingga 1)
#Semakin mendekati 1, maka modelnya semakin bagus
from sklearn.metrics import silhouette_samples, silhouette_score

score = silhouette_score (df[['Umur_Suami','Umur_Istri']], km.labels_)

print('silhouette_score: %.2f' % score)
```

```
silhouette_score: 0.57
```

Image 8. Silhouette Score

Based on Figure 8, the results of Silhouette Score is 0.57 for the K-Means method. The closer to 1, the better the model.

CONCLUSION

Based on analysis by using the K-Means Algorithm, there are 3 clusters, namely: The Ideal Cluster of 76 prospective wedding brides (age 19-30 based on husband's age and age 18-25 based on age wife), A Good Cluster of 20 prospective marriage brides (age 28-44 based on husband's age and age 24-37 based on wife's age), and A Risky Cluster of 6 prospective marriage brides (age 49-72 based on husband's age and age 39-58 based on wife's age) then it produces a Silhouette Score of 0.57 for the K-Means method. The Silhouette Score model has a value between -1 to 1, the closer to 1, the better the model, which is included in the Good Classification category. Therefore, the K-Means method is a model that is categorized as good and is implemented to predict clusters based on previous historical experience to simplify the process of classifying the marriage age of the bride and groom.

BIBLIOGRAPHY

[1] I. Virgo, S. Defit, and Y. Yuhandri, "Klasterisasi Tingkat Kehadiran Dosen Menggunakan Algoritma K-Means Clustering," *J. Sistim Inf. dan Teknol.*, vol. 2, pp. 23–28, 2020, doi: 10.37034/jsisfotek.v2i1.17.

[2] H. Hijriani, "Implementasi Pelayanan Pencatatan Pernikahan di Kantor Urusan Agama (KUA) Kecamatan Sangasanga Kabupaten Kutai Kartanegara," *J. Adm. Negara*, vol. 3, no. 2, pp. 534–538, 2015.

[3] A. P. Nanda, "Analisa Dan Perancangan Sistem Informasi Pengolahan Data Pernikahan Pada Kantor Urusan Agama (Kua)," *J. J – Click*, vol. 5, no. 1, pp. 85–97, 2018, [Online]. Available: http://ejurnal.jayanusa.ac.id/index.php/J-Click/article/view/70

[4] R. Sulistyarini, "Rasio Legis Pengaturan Batas Minimal Usia Perkawinan Menurut Undang Undang Nomor 16 Tahun 2019 Tentang Perubahan Atas Undang Undang Nomor 1 Tahun 1974 Tentang Perkawinan," *Arena Huk.*, vol. 15, no. 1, pp. 135–159, 2022, doi: 10.21776/ub.arenahukum.2022.01501.7.

[5] A. Annistri, "Ini Usia Ideal untuk Menikah, Kamu Termasuk?," *cekaja.com*, 2020. https://www.cekaja.com/info/usia-ideal-untuk-menikah

[6] I. S. Atmaja, A. Irawan, Z. Arifin, I. Habudin, N. M. Zakaria, and S. Rusmanto, "Peranan Kantor Urusan Agama (KUA) Dalam Penguatan Ketahanan Keluarga di Kecamatan Tepus," *Nuansa Akad. J. Pembang. Masy.*, vol. 5, no. 2, pp. 75–88, 2020, doi: 10.47200/jnajpm.v5i2.575.

[7] A. Saifudin, "Metode Data Mining Untuk Seleksi Calon Mahasiswa," vol. 10, no. 1, pp. 25–36, 2018.

[8] M. L. Sibuea and A. Safta, "Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustring," *Jurteksi*, vol. 4, no. 1, pp. 85–92, 2017, doi: 10.33330/jurteksi.v4i1.28.

[9] H. Astuti, "Penerapan Data

Mining Menggunakan Metode K-Means Clustering Untuk Pengelompokkan Data Pelanggan (Studi Kasus : PT. Pinus Merah Abadi)," *J. Web Inform. Teknol.*, vol. 4, no. 1, p. 9, 2019.

[10] D. T. Kusuma and N. Agani, "Prototipe Komparasi Model Clustering Menggunakan Metode K-Means Dan FCM Untuk Menentukan Strategi Promosi : Study Kasus Sekolah Tinggi Teknik-PLN Jakarta," *J. TICOM*, vol. 3, no. 3, p. 93460, 2015, [Online]. Available: https://www.neliti.com/id/publications/93460/

[11] F. Sembiring, O. Octaviana, and S. Saepudin, "Implementasi Metode K-Means Dalam Pengklasteran Daerah Pungutan Liar Di Kabupaten Sukabumi (Studi Kasus : Dinas Kependudukan Dan Pencatatan Sipil)," *J. Tekno Insentif*, vol. 14, no. 1, pp. 40–47, 2020, doi: 10.36787/jti.v14i1.165.

[12] A. upi Fitriyadi, "Analisis Algoritma K-Means dan K-Medoids Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional," *Kilat*, vol. 10, no. 1, pp. 157–168, 2021, doi: 10.33322/kilat.v10i1.1174.

[13] Y. Hartati, S. Defit, and G. W. Nurcahyo, "Klasterisasi Bibit Terbaik Menggunakan Algoritma K-Means dalam Meningkatkan Penjualan," *J. Inform. Ekon. Bisnis*, vol. 3, pp. 4–10, 2021, doi: 10.37034/infeb.v3i1.56.