JIKK | JURNAL ILMIAH KAJIAN KEIMIGRASIAN
P-ISSN 2622-4828 E-ISSN : 2808-7135
POLITEKNIK IMIGRASI POLTEKIM

A K-MEANS ALGORITHM FOR THE IMMIGRATION RESIDENTIAL PERMIT

https://10.0.205.137/jikk.v6i2.443
Submitted: 03-08-2023 Reviewed: 02-08-2023 Published: 12-08-2023

Priati Assiroj
Politeknik Imigrasi
priati.assiroj@poltekim.ac.id

I.A Pribadhi
Politeknik Imigrasi
anung56@gmail.com

Mohammad Rizal S
Politeknik Imigrasi
roniwen99@gmail.com

Abstract (In English). *Information technology utilization is increasing rapidly and massive, it falls within the realm of Immigration. Information technology has become one of the solutions to improve the performance index of Indonesian Immigration. Its utilization is not only at the immigration border but also at the Immigration Office, for example at the Immigration Office of Kediri. Data mining technology is utilized in finding and making a pattern of permit residence in its jurisdiction. We found the optimal cluster is 3. It is based on the Davies Bouldin index (DBI) that we got 0.184. There is a similarity between Cluster 1 and Cluster 3, but we found that the difference between Cluster 1 and Cluster 3 is in Cluster 1 foreigners stay in Kediri city and Kediri regency, as for Cluster 3 foreigners stay in Jombang regency, Nganjuk, etc.*

Keywords: immigration; residential permit; data mining; clustering; k-means

Abstract (In Bahasa). *Pemanfaatan teknologi informasi menjadi semakin masif termasuk dalam dunia keimigrasian. Teknologi merupakan salah satu solusi diantara solusi lainnya untuk menunjang kinerja keimigrasian khususnya di indonesia. Pemanfaatannya tidak hanya pada pos-pos pemeriksaan imigrasi di perbatasan akan tetapi juga pada kantor-kantor imigrasi di berbagai daerah, salah satunya adalah di Kantor Imigrasi Kelas II Non TPI Kediri. Teknologi data mining dimanfaatkan untuk mengetahui pola pengajuan izin tinggal di wilayah kerja kantor imigrasi kediri. Dengan teknologi ini didapatkan hasil pengklasteran yang paling optimal yaitu dengan 3 klaster, jumlah ini dapat diketahui dengan mengukur nilai indeks dari masing-masing klaster. Terlihat nilai indeks terbaik terdapat pada klaster 3, yaitu 0.184. Terdapat kemiripan antara klaster 1 dan klaster 3, akan tetatpi komponen pembedanya adalah orang asing pada klaster 1 mayoritas tinggal di Kota dan Kab. Kediri, sedangkan pada klaster 3 orang asing mayoritas tersebar di Kab. Jombang, Nganjuk dan lainnya.*

Keywords: imigrasi; izin tiggal; data mining; pengklasteran; k-means

1. INTRODUCTIONS

The migration of people has a significant effect on demographic policy in a nation. The demographics in a nation are influenced by natality and mortality (Mantra, 2000). The migration of people in a nation is called internal migration then the migration of people across the border of the nation is called international migration. As time goes on, that is not only the increase of internal migration but

also the international migration. Socio-economic, cultural, and political stability are the factors of international migration.

Indonesia, as a developed state and nation with 17K or more islands, is the biggest archipelago in the world, and one of the biggest problems is population, including people development and population dispersal. Nowadays, Indonesia has become the 4th nation with the biggest population in June 2021 (Kemendagri, 2022). Therefore, migration in Indonesia, local or international migration, is a common thing. Indonesia has become the destination of migration, especially international migration because it has natural beauty and is rich in natural resources. Table 1 below shows the immigration residence permit issued by the Kediri Immigration Office.

Table 1. Immigration residence permit issued in Kediri.

Year	Number
2019	920
2020	525
2021	415

The Indonesia Directorate General of Immigration is the government agency responsible for public service due to passport issuing, and foreigner monitoring due to its residential permits. The government must utilize modern technology to work more efficiently and effectively in public services.

The optimal public services can be realized through the adaptation and adoption of technology and innovations (Holle, 2011). Government must manage good governance in transparency to serve publicly (Lockwood, 2010). The good governance concept aligns with technology utilization, where digital public services and internet-based services are in it (Republik Indonesia, 2018). Therefore, the implementation of this policy is very slow due to the readiness that must be upscaled (Katharina et al., 2020).

The Indonesia Directorate General of Immigration has made many innovations in its services. These innovations include creativity from a high level, middle, and low level, to increase the satisfaction of people. With these innovations, hopefully, corruption, collusion, and nepotism will be decreased (Hapzah et al., 2020).

Nowadays, we have 127 immigration offices, and 13 immigration detention centers, which shows that many regions do not have an immigration office yet, and due to this condition, an immigration office often must serve many regions (more than one) in which immigration officers have also to work optimum. For example, Kediri Immigration Office Class II has 4 regions to serve, Kediri itself, the regency and the city, Jombang regency, and Nganjuk regency. This becomes a challenge for immigration officers in passport issuing, residential permits, and foreigner monitoring in these regions.

As a governance agency that serves Indonesian people and foreigners related to issuing residential permits and other immigration facilities, it also monitors and enforces immigration violations. The immigration office must implement regulations governing the entry of foreigners into the Indonesian territory. Foreigners that permitted to stay in Indonesia have a responsibility to

provide benefits, maintain public safety and order, comply with Indonesian law, and respect human rights.

In the Kediri immigration office, we see the need for data processing regarding the residential permit issues. With data mining technology implementation, we will discover the hidden knowledge. This technology works properly in predicting and classifying for next policies. Recently, much data from governance agencies has not been utilized effectively. This data mining technique is implemented to analyze the characteristics of residential permit applicants and supports the decision-making process. A stack of data can be utilized to discern useful information faster (Priati, 2018)(Assiroj, 2017)(Lia Hananto et al., 2021)(Assiroj, 2016). This research utilizes the clustering technic of data mining and implementing the k-means algorithm because of its ability to handle complex data (Assiroj & Fauzi, 2018). This clustering process discovers information in detail that helps and supports the policy of residential permits. Officials will easily understand the characteristics of applicants.

2. METHOD

This research uses the CRISP-DM method. This method is utilized because it is faster to process and analyze, more efficient, more accurate, and easily understood than other methods. Figure 1 below shows the process of the CRISP-DM method.

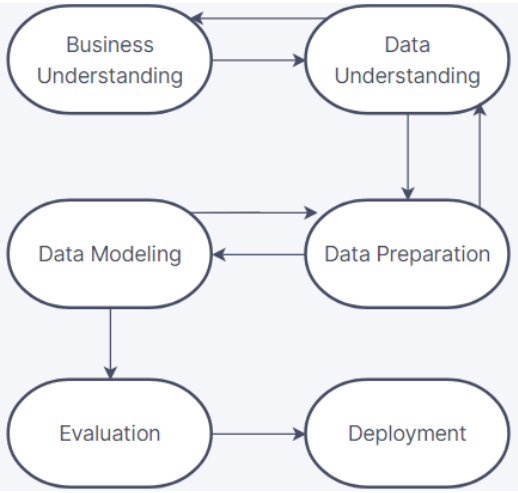

Figure 1. CRISP-DM process

2.1. BUSINESS UNDERSTANDING

Business understanding is a process to ensure that the goal of the research is achieved and delineate the boundaries, and problems formulation (Ginantra et al., 2021).

2.2. DATA UNDERSTANDING

Data understanding is a process of identifying problems in data and summarization. It needs expertise and precision in understanding data related to summarization in the business understanding step (Sari et al., 2020).

2.3. DATA PREPARATION

This step will solve data problems, by creating derived variables, selecting algorithms, and data cleansing (removing outliers and unused variables) (Astuti, 2019).

2.4. DATA MODELING

This step involves the creation of predictive or descriptive models with machine learning and statistical techniques, and the implementation of selected algorithms (Arhami, 2020).

2.5. EVALUATION

In this step, we interpret the result of the conducted experiment, evaluate the result, and make it align with the decided goals (Utari & Sofiani, 2021).

2.6. DEPLOYMENT

In this final step, we provide a report so everyone who reads this research results will understand (Tjaya et al., 2021).

3. RESEARCH

3.1. BUSINESS UNDERSTANDING PROCESS

A. Kediri Immigration Office analyzes residential permits to understand the pattern and as a support for policy and decision-making.
B. The clustering result hopefully gives information that supports residence permit policy and foreign monitoring and creates a foreign vulnerability map.
C. The first step is to obtain the data from the sub-section of immigration residential permits in the Kediri immigration office and we obtain 2096 foreigner residential permit data.

3.2. DATA UNDERSTANDING PROCESS

A. We make a description of our data as shown in the table below.

Table 2. Data type

No.	Attributes	Data type
1	Type of residential permit	Nominal
2	Registration number	Nominal
3	Completion date	Nominal
4	Effective date	Nominal
5	Name	Nominal
6	Place of birth	Nominal
7	Nationality	Nominal

8	Date of birth	Nominal
9	Gender	Nominal
10	No File	Nominal
11	No Paspor	Nominal
12	Expiry date	Nominal
13	Sponsorship	Nominal
14	Address	Nominal

B. Then we explore the data and convert the file format from .pdf to .xlsx with relevant attributes.
C. We verify data quality to identify null or missing values in every column. This process validates and ensures that data is proper for processing.

3.3. DATA PREPARATION PROCESS

A. Select the proper attributes such as type of residential permit, nationality, gender, sponsorship, and address.
B. Ensure the quality of data. Remove some noise and missing values, and initiate nominal data into numerical data. This process generates suitable data for the data mining process. The suitable data will produce a stable pattern. Nationality attribute is categorized as continental origin such as Asia, Europe, America, Africa, and Oceania. The sponsorship attribute is categorized as corporate and individual, address attribute is categorized as the area of work in the Kediri immigration office such as Kediri City and Regency, Jombang and Nganjuk Regency, etc.
   Data is then transformed into a numerical form that is suitable for k-means clustering. The type of residential permit is transformed into 1, 2, and 3 for temporary, visit, and permanent resident permit. Nationality is transformed into 1-6, 1 is for foreigners from Asia, 2 is for foreigners from Europe, 3 is for foreigners from America, 4 is for foreigners from Australia, 5 is for foreigners from Africa, and 6 is for foreigners from Oceania. Gender is transformed into 1 for male and 2 for female. Sponsorship is transformed into 1 for corporate and 2 for individual sponsors. The address is transformed into 1-5, 1 is for Kediri regency, 2 is for Kediri city, 3 is for Jombang regency, 4 is for Nganjuk regency, and 5 is for outside Kediri region.

3.4. MODELING PROCESS

This research utilizes k-means clustering algorithms in Rapid Miner. The experiment process is conducted to get the proper culture numbers. This process starts with 2 clusters and until 5 clusters. In the last process, when we make 5 clusters, we found that Cluster 0 consists of 336 items, Cluster 1 consists of 54 items, Cluster 2 consists of 715 items, Cluster 3 consists of 756 items, Cluster 4 consists of 60 items, and Cluster 5 consists of 175 items, as shown in the following figure.

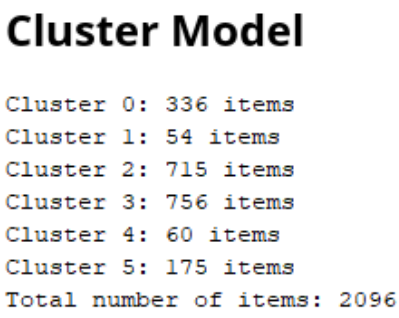

Figure 2. Cluster model result

3.5. EVALUATION PROCESS

The evaluation is carried out thoroughly to ensure that the modeling outcomes align with the predetermined objectives in the business understanding phase. This research utilizes the Davies-Bouldin Index (DBI) evaluation feature available in RapidMiner. DBI is used to measure the performance of clustering results and is presented in the form of a value. A smaller or closer-to-zero value of the index is interpreted as a good clustering process. Conversely, if the index value moves further from zero, it indicates that the clustering process is poor. The evaluation results can be seen in the following table.

Table 3. DBI values

k	Davies Bouldin Index
2	0,201
3	0,184
4	0,232
5	0,251
6	0,248

From Table 3 above, it can be observed that the value of k=3 represents the lowest value, indicating that the most optimal number of clusters in this process is 3.

3.5. DEPLOYMENT PROCESS

In this final phase of the entire CRISP-DM process, clustering produces characteristic patterns of residence permits at the Kediri immigration office. This data can serve as a reference to understand the characteristics of residence permit applications and issuances, as well as to enhance the quality of immigration services. Additionally, it can support decision-making for more effective oversight of foreigners.

4. CONCLUSION

The implementation of data mining on residence permit data at the Kediri immigration office can be performed effectively. The most optimal number of clusters is 3 with an index value of 0.814. This index value serves as a benchmark

for the optimality of the clustering process. The cluster membership is as follows: 1st cluster contains 1331 items, the 2nd cluster contains 128 items, and the 3rd cluster contains 637 items.

5. RESULT

Cluster 1 is dominated by countries from the Asian continent with a percentage of 92%, residing in Kediri Regency with a percentage of 62%, males with a percentage of 76%, and using Temporary Residence Permits with a percentage of 76%. Cluster 2 is dominated by countries from the American continent with a percentage of 44%, residing in Kediri Regency with a percentage of 45%, males with a percentage of 70%, and using Visitor's Visas with a percentage of 55%. Cluster 3 is dominated by countries from the Asian continent with a percentage of 92%, residing in Jombang Regency with a percentage of 68%, male with a percentage of 81% using Temporary Residence Permits with a percentage of 55%. The difference between Clusters 1 and 3 is that in Cluster 1, foreigners are spread across Kediri Regency and Kediri City, while in Cluster 3, foreigners are spread across Jombang Regency, Nganjuk Regency, and others.

REFERENCES

Arhami, M. (2020). *Data mining : algoritma dan implementasi*. Penerbit Andi.

Assiroj, P. (2016). Kajian Perbandingan Teknik Klasifikasi AlgoritmaC4.5, Naive Bayes, dan CART untk Prediksi Kelulusan Mahasiswa (Studi Kasus: STMIK ROSMA Karawang). *Media Informatika*, *15*(2), 1–17. https://doi.org/10.5281/zenodo.1184054

Assiroj, P. (2017). Penerapan Data Mining Pada Data Transaksi Superstore Untuk Mengetahui Kemungkinan Pelanggan Membeli Product Category Dan Product Container Secara Bersamaan Dengan Teknik Asosiasi Menggunakan Algoritma Apriori. *Techno Xplore : Jurnal Ilmu Komputer Dan Teknologi Informasi*. https://doi.org/10.36805/technoxplore.v1i2.106

Assiroj, P., & Fauzi, A. (2018). *Comparing CART and C5.0 Algorithm Performance of Human Development Index*. https://doi.org/10.1109/IAC.2018.8780439

Astuti, D. (2019). Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering. *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, *1*(2), 60–72. https://doi.org/10.20895/inista.v1i2.71

Ginantra, N. L. W. S. R., Nur Arifah, F., Hadi Wijaya, A., Septarini, R. S., Ahmad, N., Ardiana, D. P. Y., Effendy, F., Iskandar, A., Hazriani, Sari, I. Y., Gustiana, Z., Prianto, C., & Negara, E. S. (2021). *Data Mining dan Penerapan Algoritma*. Yayasan Kita Menulis.

Hapzah, S. N., Rianto, B., & Tamrin, M. H. (2020). Tata kelola inovasi pelayanan publik: dimensi kelembagaan, aktor kebijakan dan masyarakat. *Publisia: Jurnal Ilmu Administrasi Publik*, *5*(2), 153–164. https://doi.org/10.26905/pjiap.v5i2.4327

Holle, E. S. (2011). Pelayanan Publik Melalui Electronic Government: Upaya Meminimalisir Praktek Maladministrasi Dalam Meningkatan Public Service. *Sasi*, *17*(3), 21. https://doi.org/10.47268/sasi.v17i3.362

Katharina, R., Dewanti, S. C., Dharmaningtias, D. S., Sanur L, D., & Jaweng, R. N. E. (2020). *Pelayanan publik dan pemerintahan digital Indonesia* (1st ed.). Yayasan Obor Indonesia.

Kemendagri. (2022). *Rilis Data Kemendagri: Jumlah Penduduk Jabar Terbanyak, Kaltara Tersedikit*. https://dukcapil.kemendagri.go.id/phln/read/rilis-data-kemendagri-jumlah-penduduk-jabar-terbanyak-kaltara-tersedikit

Lia Hananto, A., Assiroj, P., Priyatna, B., Nurhayati, Fauzi, A., Yuniar Rahman, A., & Shofiah Hilabi, S. (2021). Analysis of Drug Data Mining with Clustering Technique Using K-Means Algorithm. *Journal of Physics: Conference Series*, *1908*(1). https://doi.org/10.1088/1742-6596/1908/1/012024

Lockwood, M. (2010). Good governance for terrestrial protected areas: A framework, principles, and performance outcomes. *Journal of Environmental Management*, *91*(3), 754–766. https://doi.org/10.1016/j.jenvman.2009.10.005

Mantra, I. B. (2000). *Demografi Umum*. Pustaka Pelajar.

Priati, F. A. (2018). Data Mining Dengan Teknik Clustering Menggunakan Algoritma K-Means Pada Data Transaksi Superstore. *Seminar Nasional Informatika Dan Aplikasinya*, *September 2017*, 15–19. https://doi.org/10.5281/zenodo.1186505

Republik Indonesia. (2018). Peraturan Presiden Republik Indonesia Nomor 95 Tahun 2018 Tentang Sistem Pemerintahan Berbasis Elektronik. *Menteri Hukum Dan Hak Asasi Manusia Republik Indonesia*, 110.

Sari, Y. P., Primajaya, A., & Irawan, A. S. Y. (2020). Implementasi Algoritma K-Means untuk Clustering Penyebaran Tuberkulosis di Kabupaten Karawang. *INOVTEK Polbeng - Seri Informatika*, *5*(2), 229. https://doi.org/10.35314/isi.v5i2.1457

Tjaya, P. P., Rino, R., & Wijaya, H. (2021). Implementasi Metode Clustering K-Means Untuk Rekomendasi Pengadaan Stock Lampu Pada Pt Global Lighting Indonesia. *Algor*, *3*(1), 50–59. https://doi.org/10.31253/algor.v3i1.628

Utari, D. R., & Sofiani, F. (2021). Klasterisasi Lulusan Pendidikan Tinggi Vokasi Berbasis Hasil Studi Pelacakan Menggunakan Algoritme K-Means. *Jurnal Serasi*, *19*(2), 75. https://doi.org/10.36080/js.v19i2.1543