



## Passport Application Data Mining Implementation at The Muara Enim Non-TPI Class II Immigration Office

**Nurul Maharani  
Piranti<sup>1\*</sup>**

Politeknik Pengayoman Indonesia,  
Indonesia

**Fadhluhrahman  
Khalish<sup>2</sup>**

Politeknik Pengayoman Indonesia,  
Indonesia

**Okky Pratama  
Martadireja<sup>3</sup>**

Politeknik Pengayoman Indonesia,  
Indonesia

---

**\*Corresponding author:**

Nurul Maharani Piranti, Politeknik  
Pengayoman Indonesia, Indonesia.  
✉ [nurulpiranti20@gmail.com](mailto:nurulpiranti20@gmail.com)

---

**Article Info:**

**Article history:**

Received: February 14, 2026

Revised: February 24, 2026

Accepted: February 28, 2026

---

**Keywords:**

data mining; k-means clustering;  
passport applications; python;  
immigration office.

---

**Abstract**

**Background:** The Class II Non-TPI Muara Enim Immigration Office processes thousands of passport applications annually. However, existing data utilization remains limited to administrative recording and routine reporting, constraining its strategic value for service planning.

**Objective:** This study aims to identify meaningful patterns in passport application data and determine the optimal cluster structure using the K-Means algorithm to support evidence-based decision-making.

**Methods:** Therefore, this study was conducted to apply data mining techniques using the K-Means clustering method to analyze passport application patterns based on data from 2022 to 2024, with attributes such as gender, travel purpose, type of application, type of passport, and year of submission, implemented through Python with the support of Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn libraries.

**Results:** The results show that passport applicants can be grouped into several clusters with specific characteristics — for instance, clusters dominated by applicants of productive age with purposes of working or performing *Umrah*, as well as other clusters consisting mainly of applicants traveling for tourism or education; moreover, the analysis reveals spikes in applications during certain periods, such as before the *Hajj* season and long holidays.

**Conclusion:** Thus, it can be concluded that applying K-Means clustering provides added value beyond conventional data processing, enabling administrative data to be transformed into informative and predictive patterns that support resource planning, staff allocation, and more effective data-driven service policies.

---

**To cite this article:** Piranti, N. M., Khalish, F., & Martadireja, O. P. (2026). Passport Application Data Mining Implementation at The Muara Enim Non-TPI Class II Immigration Office. *INKUBIS: Jurnal Ekonomi dan Bisnis*, 8(1), 42–64. <https://doi.org/10.59261/inkubis.v8i1.141>

---

### INTRODUCTION

The development of technology that occurs in the current era has brought very significant changes in various aspects, including in public services. Digitalization allows government agencies to process and analyze large amounts of data more systematically and efficiently (Hapsari & Nataliani, 2024; Qing et al., 2020; Sun, 2019; Wulansari, 2019). With the use of digital technology, agencies can improve the quality of services and provide easier access for the public to receive information about the public services needed. Digitalization also allows for the integration of data from various agencies, making it possible to coordinate quickly and appropriately in policy-making. One form of digitalization that continues to grow is the application of data mining in government administration. Data mining is one of the techniques in artificial intelligence (AI) used to process, analyze, and find hidden patterns in large data sets (Li, 2022; Luo et al., 2022; Zhang, 2025). Data mining is also used to predict and make more accurate decisions. The application of this technique is expected to be able to identify historical trends, analyze community behavior patterns, and support sustainable service optimization based on data.

In various countries, data mining has been applied in the immigration system, which has an important role in managing population mobility, border surveillance, and administration of travel documents (Guohua & Francis, 2017; Hamid Mughal, 2018). One of its uses is to analyze the mobility trends of the population, both Indonesian citizens (WNI) who travel abroad and foreign citizens (WNA) who enter or leave Indonesian territory. By conducting this analysis, it is hoped that immigration officers can predict the increase in passport applications that will occur in the following year. In addition, data mining also allows for the optimization of immigration data management so that the system can automatically group applicants based on certain characteristics. With this technology, it is hoped that immigration services can develop to be more proactive, predictive, and more efficient in managing immigration administration data (Baghitz Hanan et al., 2025; Ryanindityo et al., 2024; Susaningsih & Puteri Andika, 2022).

The Directorate General of Immigration of the Ministry of Immigration and Corrections continues to digitize services to improve the efficiency, transparency, and quality of immigration services, one of which is through the Immigration Management Information System (SIMKIM), which allows centralized and accurate data management (Aditya, 2020; Alfarisi Akasy, 2021; Hafiz-AM, 2019). Passports as official documents for Indonesian travel are processed by the Immigration Office, including the Muara Enim Non-TPI Class II Immigration Office, which is responsible for passport services and immigration supervision in its work area. Along with the increase in public mobility for tourism, worship, work, and education purposes, there has been a significant surge in passport applications in the last three years, so a more effective and anticipatory service management strategy is needed to maintain timeliness, accuracy, and service quality (Assiroj et al., 2023; Azizah et al., 2025).

The selection of this issue was based on the results of an interview conducted with one of the employees of the Muara Enim Non-TPI Class II Immigration Office in the immigration traffic and residence permit section, stating that currently the data processing system carried out is still reactive, without predictive analysis that can help future planning. The increasing number of applicants every year requires a data-driven approach to identify passport application patterns in order to improve service quality and support more strategic policies. This research offers a new concept in the management of passport application data by applying K-Means-based clustering, which allows the grouping of passport applicant data based on their characteristics, such as gender, travel destination, and type of application.

The results of this clustering can be used to develop service improvement strategies, such as adjusting service schedules, allocating resources, and formulating data-based policies. The gap between current and future conditions lies in data management methods. Currently, passport services still depend on the number of applications that come in without a data-based strategy. With the implementation of data mining and clustering, the Immigration Office can transform toward a more modern and efficient approach, where data is used not only as an administrative archive but also as an analytical tool for predicting a surge in applications, identifying applicant trends, and increasing effectiveness in the field of services. Thus, this research is expected to make a real contribution in supporting the optimization of passport application management, facilitating decision-making, and improving the efficiency of immigration services as a whole.

Previous studies have explored the application of data mining in various sectors, but with different contexts and focuses. Research by Sianturi et al. (2025) applied the K-Means algorithm to cluster districts and cities in North Sumatra Province based on poverty indicators in 2023, demonstrating that clustering effectively identifies regional poverty patterns to support policy targeting; however, this study was limited to a single-year cross-sectional dataset without incorporating temporal trends or predictive dimensions for future planning. Meanwhile, Herman et al. (2022) conducted a comparative cluster analysis using K-Means versus K-Medoid methods in the context of financial performance evaluation, highlighting the strengths and trade-offs of each approach in producing meaningful groupings; yet the study focused exclusively on financial metrics and did not extend its methodology to public sector service data or operational decision-making contexts. On a related front, Marksteiner et al. (2024) explored the domain of passport control by mining and formally checking behavioral models of machine-readable travel documents in automated border systems, contributing to the security and verification dimension of immigration technology; nevertheless, this work did not address applicant-level data patterns, service demand clustering, or administrative trend analysis within immigration offices. Unlike

these previous studies, this research presents novelty along two dimensions: contextual novelty, as the pioneering application of K-Means clustering to passport application data in Indonesian immigration administration, a sector with distinct characteristics (high mobility rates, travel purpose variations, and seasonal fluctuations) that remains unexplored in existing data mining literature; and methodological novelty, through integrating a three-year temporal analysis (2022–2024) to identify a 45.4% application surge trend and predict future patterns based on cluster characteristics (gender, travel destination, and application type).

Based on the explanation that has been described, the researcher is interested in conducting research related to the use of passport applicant data at the Muara Enim Non-TPI Class II Immigration Office. This study aims to apply the data mining technique with the K-Means clustering method to analyze passport application patterns at the Muara Enim Non-TPI Class II Immigration Office, so that it can provide useful insights in the decision-making process and support the improvement of the efficiency of immigration services. The specific purpose is to analyze the process of processing passport application data using the clustering technique through the K-Means method at the Muara Enim Non-TPI Class II Immigration Office and to analyze the quality of the results of processing passport application data with the K-Means method and compare it with the data management conditions currently applied at the Muara Enim Non-TPI Class II Immigration Office.

This research provides both theoretical and practical benefits by advancing the application of unsupervised machine learning, specifically K-Means clustering, within the public administration and immigration service sectors. Theoretically, it enriches the literature on data mining implementation in government institutions and offers empirical evidence on the effectiveness of clustering validation metrics such as the Silhouette Score and Davies–Bouldin Index in identifying mobility patterns and forecasting passport demand trends. Practically, the findings support the Muara Enim Non-TPI Class II Immigration Office in transitioning from conventional administrative data management toward a predictive and analytical approach, enabling more accurate anticipation of application surges, optimized staff allocation, targeted service strategies for priority applicant segments, and improved operational efficiency.

## METHOD

This study employed a quantitative, applied data mining research design aimed at identifying passport application patterns using the K-Means clustering algorithm. The population of this research consisted of all passport application records submitted to the Muara Enim Non-TPI Class II Immigration Office from 2022 to 2024. The research sample included the entire accessible historical dataset within this period, applying a total sampling technique due to the use of secondary administrative data extracted from the SIMKIM (Immigration Management Information System) database. The primary research instrument was the structured dataset obtained from SIMKIM, containing attributes such as gender, travel destination, type of application, type of passport, and year of submission. Since the study used secondary quantitative data, construct validity was ensured through attribute selection aligned with clustering objectives, while data reliability was supported through system-generated records and consistency checks during preprocessing.

Data collection was conducted through official data extraction from SIMKIM under institutional authorization between March and September 2025. The research procedure followed the CRISP-DM framework, consisting of business understanding, data understanding, data preparation (including cleaning, transformation, encoding, and normalization), modeling, evaluation, and deployment. During the modeling stage, the optimal number of clusters was determined using the Elbow Method and Silhouette Score analysis. Data preprocessing ensured the removal of incomplete records and the normalization of categorical attributes into numerical form to improve clustering performance.

The analysis was implemented using the Python programming language, supported by the Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn libraries. Cluster quality was evaluated using the Davies–Bouldin Index (DBI) to measure compactness and separation between clusters.

The analytical results were interpreted descriptively and strategically to generate actionable insights for service optimization, including personnel allocation planning and demand forecasting, and were presented in the form of analytical reports and dashboards to support data-driven decision-making within the immigration office.

## RESULTS AND DISCUSSION

### Results

#### *Data Mining Implementation*

##### A. Business Understanding Phase

The Muara Enim Non-TPI Class II Immigration Office is a Technical Implementation Unit (UPT) under the Directorate General of Immigration, which is responsible for providing passport application services for Indonesian citizens, issuing residence permits, and supervising those domiciled within its working area. Along with the increase in community mobility, the number of passport applications at the Muara Enim Non-TPI Class II Immigration Office has also increased each year. From the data used in this study, an analysis will be carried out on the pattern of passport applications at the Non-TPI Class II Immigration Office in Muara Enim. The results of the processing and analysis carried out using the data mining technique are used as a basis for decision-making to support the effective management of passport application data and as an advancement of data science understanding in the context of immigration analysis.

The pattern obtained is used as a consideration for decisions regarding the application of data mining as a means of advancing immigration analysis, so that this study may serve as a reference for understanding the implementation of data mining using the k-means algorithm in the management of passport application data at the Non-TPI Class II Immigration Office in Muara Enim. The initial stage of this research process is data collection, in which passport application data is extracted for analysis. The data used in this study is drawn from passport application reports for the years 2022, 2023, and 2024.

##### B. Data Understanding Phase

In the data understanding phase, there are several stages which are as follows:

1. **Initial Data Collection.** This stage of data collection is carried out to learn the data structure that will be used. The data used is data retrieved from the Immigration Management Information System (SIMKIM). The results of the data retrieval are then stored in Excel format, with the total amount of data recorded from January 01, 2022 to December 31, 2024 being 30,690 passport application records, with a data distribution in which 2022 has 7,921 records, 2023 has 11,517 records, and 2024 has 11,252 records.
2. **Data Description.** The retrieved passport application data has several attributes. The attributes contained in the data are Application Number, Application Date, Application Type, Passport Purpose, Passport Type, Applicant Name, Gender, and Passport Number. These attributes are presented in the form of a table.
3. **Data Exploration.** This stage is carried out to explore and obtain data on passport applications. The data is the result of extractions carried out from the Immigration Management Information System (SIMKIM) with extraction periods covering 2022, 2023, and 2024. The results of the data collection are then stored in Excel format.
4. **Quality Verification.** Quality verification is a stage carried out to analyze the quality of the data obtained. This stage is carried out to ensure that the data retrieved through the Immigration Management Information System (SIMKIM) is complete and that no attributes are missing or data has failed to be extracted, so that the quality of the data resulting from the extraction process is consistent with the data contained in the system.

##### C. Data Preparation Phase

The data preparation phase is a stage carried out to prepare the passport application data used in this study. The stages of the data preparation phase are as follows:

1. **Data Selection.** In the data selection process, following the data extraction process, a data selection stage is carried out to determine the attributes that will be used in the data processing process. The data attributes taken from the primary dataset represent the data required for the

data processing process and will serve as the dataset used in the next stage. The attributes used in this study are Applicant Type, Passport Purpose, Passport Type, Gender, and Year.

2. Data Processing

**Table 1.** Nominal Type Attribute Display

<b>Types Applicant</b>	<b>Objectives Passport</b>	<b>Types Of Passports</b>	<b>Types Sex</b>	<b>Year</b>
Replace (Expired Validity Period)	<i>Hajj</i>	Ordinary Passport 48 H	P	2022
New	Tourism	Ordinary Passport 48 H	L	2022
Replace (Expired Validity Period)	Tourism	Ordinary Passport 48 H	L	2022
Replace (Expired Validity Period)	Tourism	Ordinary Passport 48 H	L	2022
New	Tourism	Ordinary Passport 48 H	L	2022
New	Tourism	Ordinary Passport 48 H	L	2022
New	Tourism	Ordinary Passport 48 H	P	2022
New	<i>Umrah</i>	Ordinary Passport 48 H	P	2022
New	Tourism	Ordinary Passport 48 H	L	2022
...	...	...	...	...
Replace (Expired Validity Period)	<i>Umrah</i>	Passport Electronics 48 H	P	2024
New	Tourism	Ordinary Passport 48 H	P	2024
New	Tourism	Ordinary Passport 48 H	P	2024

Data processing is used to improve the quality of the data that will be used in the analysis process. Data processing is used to convert raw data into useful and analyzable information. This process helps in identifying patterns, trends, and insights that can support better decision-making. In various fields, such as business, research, and technology, data processing enables the efficient handling of large amounts of data, resulting in accurate reports, predictions, and recommendations for strategy and policy development.

3. Transformation Data

Data transformation is the process of changing data from its original format or structure to a more suitable form for further analysis or processing. This involves a variety of techniques, such as normalization (scaling data to a standard range), aggregation (combining large amounts of data into a single unit), or discretization (e.g., converting numbers into categories). The goal of data transformation is to make data easier to understand, analyze, and use to make better decisions. At this stage, the data will be transformed according to the need to perform cluster analysis using k-means, where the nominal data will be transformed into numerical form. The following data transformations are performed based on data attributes:

a. Attributes of Application Type

Based on the data presented, the attributes of the type of applicant are categorized based on the type of passport application. The data transformation carried out on the attributes of the

applicant type is shown in Table 2.

**Table 2.** Transformation of Applicant Type Attribute Data

<b>Data Name</b>	<b>Quantity Record</b>	<b>Transformation</b>
New	21.723	1
Replace (Expires)	8706	2
Replace (Lost Because Kahur's Situation)	3	3
Replace (Lost)	191	4
Replace (Full / Page Full)	72	5
Replace (Damaged Due to Harsh Condition)	9	6
Replace (Broken)	16	7
Change	240	8

b. Attributes of the Purpose of the Passport

The transformation of data on the attributes of the passport purpose is presented based on the purpose of the applicant making a passport in accordance with the existing data and used in this study. The data transformation carried out on the passport destination attributes is shown in table 3.

**Table 3.** Passport Purpose Attribute Data Transformation

<b>Data Name</b>	<b>Number of Records</b>	<b>Transformation</b>
Formal Work	326	1
Learn	351	2
Treatment	272	3
<i>Hajj</i>	2570	4
TKI	568	5
<i>Umrah</i>	16.177	6
Tourism	10.696	7

c. Attributes of Passport Type

The transformation of passport type attributes is then carried out by analyzing the type of passport issued based on the needs of the applicant. The data transformation in the passport type attributes is shown in Table 4.

**Table 4.** Passport Type Attribute Data Transformation

<b>Data Name</b>	<b>Quantity Record</b>	<b>Transformation</b>
24 H Regular Passport	3	1
Ordinary Passport 48 H	23.557	2
Passport Ordinary Electronics 10 Years Lamination	51	3
Passport Ordinary Electronics 5 Years Lamination	169	4
48 H Electronic Passport	7180	5

d. Gender Attributes

For the Gender attribute, data transformation was carried out to encode Male as 1 and Female as 2. Table 5 shows the data transformation performed on the Gender attribute.

**Table 5.** Gender Attribute Data Transformation

<b>Data Name</b>	<b>Quantity Record</b>	<b>Transformation</b>
Male	13.447	1
Women	17.508	2

The following is a table of the results of the data transformation that has been carried out on the passport application data to be used, as shown in Table 6.

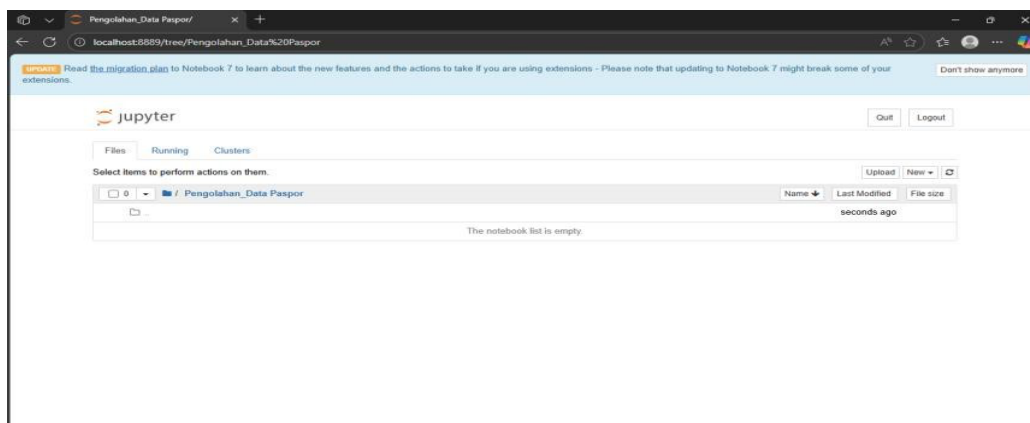
**Table 6.** Transformation Data Table

Applicant Type	Passport Purpose	Passport Type	Gender	Year
1	7	2	1	2022
2	7	2	1	2022
2	7	2	1	2022
1	7	2	1	2022
1	7	2	1	2022
1	7	2	2	2022
1	6	2	2	2022
1	7	2	1	2022
8	6	2	2	2022
...	...	...	...	...
2	6	5	2	2024
1	7	2	2	2024
1	7	2	2	2024

#### 4. Data Modeling Phase

This research uses the Python programming language with the Anaconda platform. The dataset collected in the previous stage is stored in a file named Data\_Set.xlsx, which has been processed so that it is ready to be analyzed using the k-means algorithm. In forming clusters, the k-means algorithm uses the centroid model. The centroid is the central point of the cluster in the form of a value. The similarity value of the data can be determined by calculating the distance between a data object and the centroid value. The data will be grouped into a cluster with the nearest central point, or in other words, the shortest distance to the centroid of the cluster. The starting centroid point is determined randomly based on the number of k clusters. The initial step of cluster testing in this study starts from k=2 to k=6. The following are the stages of data processing using the Python programming language with the Anaconda platform.

The initial stage is to open the initial view of Jupyter Notebook and then create a separate folder to save the files that will be used in the process. In this study, the folder created is the Pengolahan\_Data\_Paspor folder. Figure 1 is a view of Jupyter Notebook.



**Figure 1.** Jupyter Notebook Display

Then, upload the Excel file or dataset used in this study into the Pengolahan\_Data\_Paspor folder. The initial stage is to upload data by clicking the upload button.

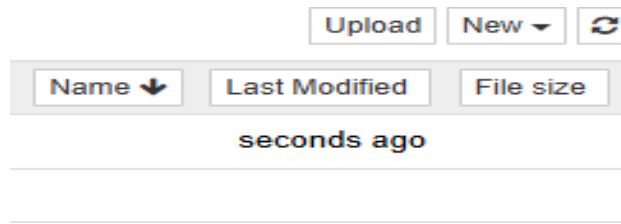


Figure 2. Button Upload

Then select the dataset that has been prepared in the form of an Excel file. Click Upload to upload data.



Figure 3. Data Upload Display

After clicking on upload, the data has been successfully uploaded into the folder.

The cluster test was carried out by trying the values  $k=2$  to 6. This test is done using the python programming language.

- a. The initial stage is to enter the python programming language text editor by clicking new and then selecting Python 3 (ipykernel).

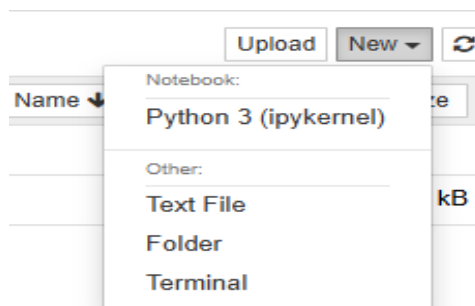


Figure 4. Creating a Python Text Editor

- b. Here's an Initial View of the Python Programming Language Text Editor.

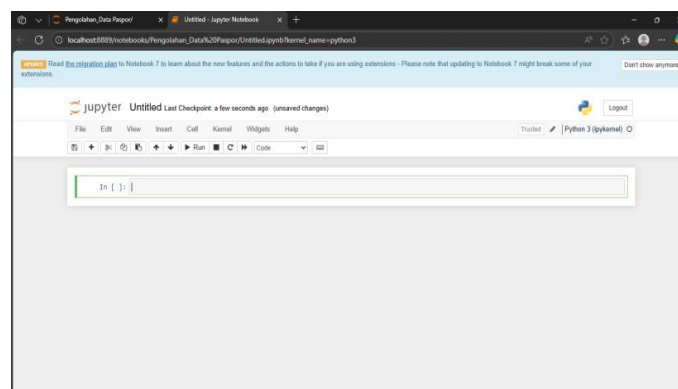


Figure 5. Initial Appearance of Python Programming Language Text Editor

- c. In the *Untitled* section, replace it with the name you want to create e.g. *Pengolahan\_Data\_Paspor\_k-means*.

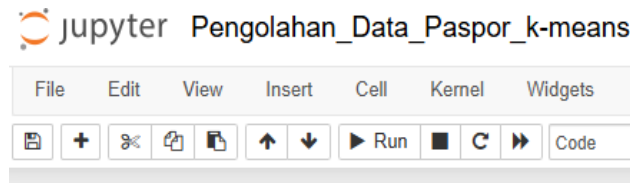


Figure 6. Naming Editor Files

d. Then do the initial coding is shown in figure 7.

```
In [ ]: import pandas as pd

# Load the uploaded Excel file
file_path = 'Data_Set.xlsx'
data = pd.read_excel(file_path)

# Show the first few rows of the data to understand its structure
data.head()
```

Figure 7. Read Data Initial Code

This code imports the pandas library, reads the data from the Excel file specified in the file\_path, and displays the first five lines using data.head() to give a preliminary idea of the data structure. After running, the output is shown in figure 8.

Out[1]:

	JENIS PEMOHON	TUJUAN PASPOR	JENIS PASPOR	JENIS KELAMIN	TAHUN
0	1	7	2	1	2022
1	2	7	2	1	2022
2	2	7	2	1	2022
3	1	7	2	1	2022
4	1	7	2	1	2022

Figure 8. Output Displaying Data

e. Data Clustering using the k-means Algorithm.

```
In [ ]: # Always show details
from sklearn.cluster import KMeans

# Select relevant features for clustering
features = data[['JENIS PEMOHON', 'TUJUAN PASPOR', 'JENIS PASPOR', 'JENIS KELAMIN', 'TAHUN']]

# Apply K-means clustering with a chosen number of clusters
kmeans = KMeans(n_clusters=2, n_init=10, random_state=42) # Adjust the number of clusters as needed
data['Cluster'] = kmeans.fit_predict(features)

# Save the resulting data with clusters to a new Excel file
output_file_path = 'clustering_results.xlsx'
data.to_excel(output_file_path, index=False)

# Get the number of iterations it took
iterations = kmeans.n_iter_

output_file_path, iterations
```

Figure 9. Data Clustering

This code clusters the data using the K-means algorithm from the sklearn library. Here's a quick rundown of each piece of code:

- 1) from sklearn. KMeans import cluster: Imports the KMeans algorithm from the sklearn library, which is used to perform data clustering.
- 2) features = data[['APPLICANT TYPE', 'PASSPORT PURPOSE', 'PASSPORT TYPE', 'GENDER', 'YEAR']]: Selects the relevant columns from the DataFrame data as the feature to be used for clustering. This feature includes categories related to the type of applicant, passport purpose,

- passport type, gender, and year.
- 3) `kmeans = KMeans(n_clusters=2, n_init=10, random_state=42)`: Initializes the K-Means model with two *clusters* (`n_clusters=2`), which will be randomly selected with 10 different initializations (`n_init=10`). `random_state=42` ensures consistent results each time the code is executed.
  - 4) `data['Cluster'] = kmeans.fit_predict(features)`: Applies the K-means algorithm to the selected feature and adds a new *Cluster* column to the DataFrame `data`, which contains *the cluster* label for each row of data based on the clustering results.
  - 5) `output_file_path = 'clustering_results.xlsx'`: Save the clustered result into a new Excel file with *the name clustering\_results.xlsx*.
  - 6) `data.to_excel(output_file_path, index=False)`: Saves the updated DataFrame (including columns *cluster*) into the Excel file specified by `output_file_path`.
  - 7) `iterations = kmeans.n_iter_`: Takes the number of iterations required by the K-means algorithm to converge (achieve the result).
  - 8) `output_file_path, iterations`: Displays the output file path and the number of iterations used by K-means to converge.

**Table 7.** Data clustering results

Type applicant	Purpose passport	Type passport	Type sex	Year	Cluster
1	7	2	1	2022	0
2	7	2	1	2022	0
2	7	2	1	2022	0
1	7	2	1	2022	0
1	7	2	1	2022	0
1	7	2	2	2022	0
1	6	2	2	2022	0
1	7	2	1	2022	0
8	6	2	2	2022	0
...	...	...	...	...	...
2	6	5	2	2024	1
1	7	2	2	2024	0
1	7	2	2	2024	0

f. Data clustering with Optimization with k-means Algorithm.

```
In [ ]: # Always show details
# Retry clustering with reduced number of iterations (max_iter) for faster processing
kmeans = KMeans(n_clusters=2, n_init=10, max_iter=100, random_state=42)

# Apply K-means clustering and add the cluster column
data['Cluster'] = kmeans.fit_predict(features)

# Save the results to a new Excel file
output_file_path = 'clustering_results_optimized.xlsx'
data.to_excel(output_file_path, index=False)

# Number of iterations for the k-means process
iterations = kmeans.n_iter_

output_file_path, iterations
```

**Figure 10.** Data clustering with Optimization

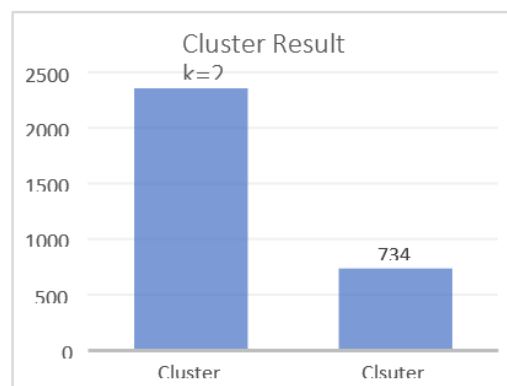
This code performs data clustering using the K-means algorithm with optimization to speed up the process. Here's an explanation for each piece of code:

- 1) `kmeans = KMeans(n_clusters=2, n_init=10, max_iter=100, random_state=42)`: Initializes the K-means model with two *clusters* (`n_clusters=2`). `n_init=10` specifies that the algorithm will run 10 different initializations to get more stabil. `max_iter=100` results limiting the maximum number of iterations that the algorithm can perform in a single run to 100 iterations, which helps speed up the process. `random_state=42` ensures consistent results every time the code is executed.'
- 2) `data['Cluster'] = kmeans.fit_predict(features)`: Applies K-means to the selected feature and

adds a new *Cluster* column to the DataFrame data, which contains *the cluster* label for each row of data.

- 3) `output_file_path = 'clustering_results_optimized.xlsx'`: Saves the updated clustering results into a new Excel file with *the clustering\_results\_optimized.xlsx* name.
- 4) `data.to_excel(output_file_path, index=False)`: Stores updated DataFrames, including columns *cluster*, into an Excel file in the location specified by the `output_file_path`.
- 5) `iterations = kmeans.n_iter_`: Keeps the number of iterations used by the K-means algorithm until it converges (reaches the result).
- 6) `output_file_path, iterations`: Displays the clustered result file path and the number of iterations used by K-means to achieve convergence.
- g. The cluster results use the k-means algorithm.
- 1) Clustering test results with  $k = 2$

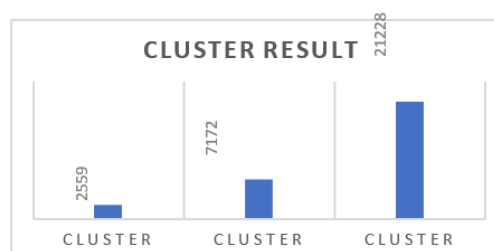
The results of the clustering test using the k-means algorithm with a value of  $k=2$  stopped in the 3rd iteration by being divided into 2 clusters, consisting of cluster 0 having 23,610 items while cluster 1 having 7,349 items. It can be seen in figure 11.



**Figure 11.** Cluster Results  $k=2$

- 2) Clustering test results with  $k = 3$

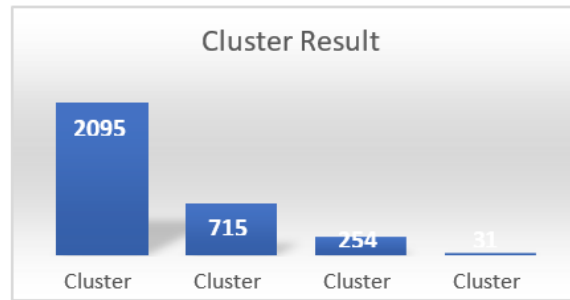
The results of the clustering test using the k-means algorithm with a value of  $k=3$  stopped in the 18th iteration by being divided into 3 clusters, consisting of cluster 0 having 2,559 items, cluster 1 having 7,172 items, and cluster 2 has 21,228 items. It can be seen in figure 12.



**Figure 12.** Cluster Results  $k=3$

- 3) Clustering test results with  $k=4$

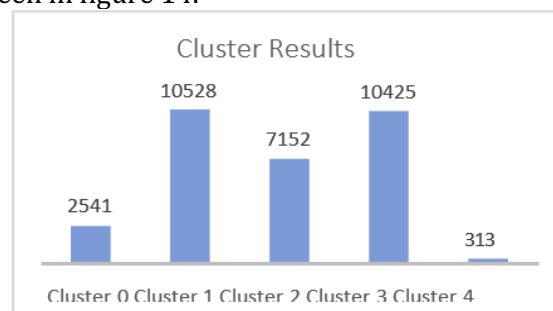
The results of the clustering test using the k-means algorithm with a value of  $k=4$  stopped in the 5th iteration by being divided into 4 clusters, which consisted of cluster 0 has 20,953 items, cluster 1 has 7,152 items, cluster 2 has 2,541 items, and cluster 3 has 313 items. It can be seen in figure 13.



**Figure 13.** Cluster Results k=4

4) Clustering Test Results with k=5

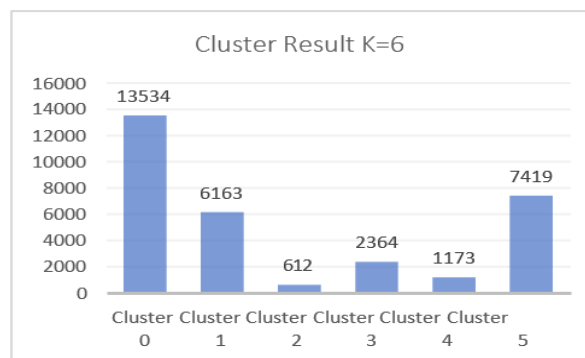
The results of the clustering test using the k-means algorithm with a value of k=5 stopped in the 5th iteration by being divided into 5 clusters, consisting of cluster 0 having 2541 items, cluster 1 having 10,528 items, cluster 2 has 7152 items, cluster 3 has 10,425 items, and cluster 4 has 313 items. It can be seen in figure 14.



**Figure 14.** Cluster Results k=5

5) Clustering Test Results with k=6

The results of the clustering test using the k-means algorithm with a value of k=6 stopped in the 5th iteration by being divided into 6 clusters, consisting of cluster 0 having 13,534 items, cluster 1 having 6,163 items, cluster 2 has 612 items, cluster 3 has 2,364 items, cluster 4 has 1,173 items, and cluster 5 has 7,419 items. It can be seen in figure 15.



**Figure 15.** Cluster Results k=6

*Evaluation Phase*

The clustering results that have been processed are then further evaluated in depth to match the goals that have been determined in the initial phase. In this study, the clustering evaluation phase was carried out by testing the quality of the results of cluster 2 to cluster 6 using the Davies Boulding Index (DBI). This DBI test was carried out to measure the performance of clustering that had been carried out previously so that the quality of the data processing in this study could be concluded

Based on the concept of the DBI value produced by clustering is getting smaller, which is closer to 0 or equal to 0, the better the quality of the cluster. And vice versa, large DBI results mean that the clustering process is not good or even bad. Here are the stages of testing a DBI using Python Programming to evaluate the results of clustering.

- A. The initial stage is to import clustering results from k2 to k=6 that have been stored in excel form before in the analysis process using k-means. Here is the code used to import data with python.

```
In [2]: import pandas as pd

# Fungsi untuk mengimpor file Excel
def import_excel(file_path):
    # Membaca file Excel
    df = pd.read_excel(file_path)
    return df

# File paths dari file yang diunggah
file_paths = [
    "clustering_results_optimized_k2.xlsx",
    "clustering_results_optimized_k3.xlsx",
    "clustering_results_optimized_k4.xlsx",
    "clustering_results_optimized_k5.xlsx",
    "clustering_results_optimized_k6.xlsx"
]

# Mengimpor semua file menjadi dataframe
dataframes = [import_excel(file_path) for file_path in file_paths]

# Menampilkan preview data dari file pertama
dataframes[0].head()
```

Figure 16. Cluster Result Import Code

- B. The following output displays data from the clustering results k=2.

```
Out[2]:
```

	JENIS PEMOHON	TUJUAN PASPOR	JENIS PASPOR	JENIS KELAMIN	TAHUN	Cluster
0	1	7	2	1	2022	0
1	2	7	2	1	2022	0
2	2	7	2	1	2022	0
3	1	7	2	1	2022	0
4	1	7	2	1	2022	0

Figure 17. Data Output Clustering Results k=2

- C. Next, it is to conduct a DBI test with the code in figure 18.

```
In [4]: from sklearn.metrics import davies_bouldin_score

# Function to calculate DBI for each clustering result
def calculate_dbi(df, features_columns):
    # Extracting features and clusters from the dataframe
    X = df[features_columns].values
    labels = df['cluster'].values
    # Calculate DBI
    dbi = davies_bouldin_score(X, labels)
    return dbi

# Feature columns for clustering
features_columns = ['JENIS PEMOHON', 'TUJUAN PASPOR', 'JENIS PASPOR', 'JENIS KELAMIN', 'TAHUN']

# Calculate DBI for each clustering result
dbi_results = {f'k={i+2}': calculate_dbi(df, features_columns) for i, df in enumerate(dataframes)}

# Convert DBI results into a DataFrame
dbi_df = pd.DataFrame(list(dbi_results.items()), columns=['K (Clusters)', 'DBI'])

# Save the DBI results to an Excel file
output_path = 'dbi_results_comparison.xlsx'
dbi_df.to_excel(output_path, index=False)

output_path
```

Figure 18. DBI Testing Codes

From the DBI test with the code figure 18, it generates the DBI test results which are stored in an excel file with the file name dbi\_results\_comparison.xlsx.

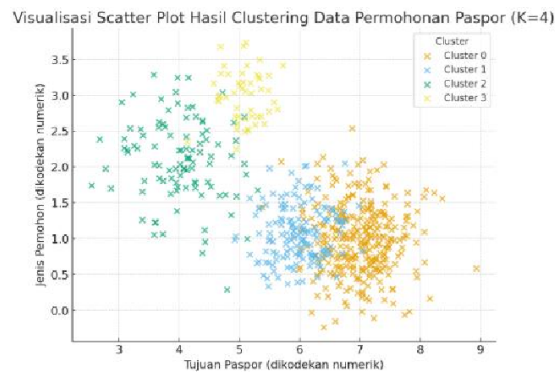
- D. DBI Test Results

The following DBI test results get the following values:

K (Clusters)	DBI
k=2	0,799
k=3	0,804
k=4	0,696
k=5	1,018
k=6	0,887

Based on the table 8, it can be concluded that cluster results with a value of k=4 have better and more optimal clustering quality than other clusters that have k=3 to k=6. The DBI value of cluster 2 is closest to 0 which is 0.696. To clarify the results of data grouping, the following is

presented a visualization of scatter plots resulting from the clustering process using the K-Means algorithm with the optimal number of clusters  $K=4$ . This visualization illustrates how passport application data is grouped based on the attributes of Applicant Type and Passport Purpose that have been converted to numerical form.



**Figure 19.** Scatter Plot Visualization

Figure 19 shows that the data of passport applicants is divided into four main groups. Cluster 0 and Cluster 1 are dominated by new applicants with the purpose of *Umrah*, while Cluster 2 depicts applicants for passport replacement for *Haji*, and Cluster 3 shows a small group of applicants with special administrative needs. This visualization strengthens the results of the analysis that data grouping with the K-Means algorithm can produce clear segmentation and can be used as a basis for decision-making in the field of immigration services.

#### Deployment Phase

This stage is the final stage of the data mining process in this research using clustering techniques with the K-Means algorithm. The result of this data mining process is in the form of patterns with distinct characteristics, making them information that can be utilized in various ways such as materials and guidelines for the Muara Enim Non-TPI Class II Immigration Office in considering the formulation of draft policies for passport applicant services in the passport application process, and as decision-making material concerning the supervision of passport usage by passport applicants.

#### Data Mining Performance

The processing of passport applicant data that has been processed using clustering produces a more optimal DBI value in cluster 4 which has a DBI value of 0.696 which shows that the performance and quality of the cluster is better and optimal than other clusters. The result of cluster 4 with a value of  $k=4$  produces 4 clusters divided into cluster 0 has 20,953 items, cluster 1 has 7,152 items cluster 2 has 2,541 items, and cluster 3 has 313 items. The processing of cluster results based on each attribute has been categorized as follows.

#### Attributes of Applicant Type

**Table 9.** Results of Processing Applicant Type Attributes

Data Name	C0	C1	C2	C3
New	72,31%	71,12%	60,54%	0,00%
Replace (Expires)	27,02%	28,07%	42,22%	0,00%
Replace (Lost Due to Kahur Condition)	0,014%	0,00%	0,00%	0,00%
Replace (Lost)	0,649%	0,475%	0,856%	0,00%
Replace (Full/Full Page)	0,00%	0,321%	0,040%	15,33%
Replace (Damaged by Force Majeure)	0,00%	0,00%	0,00%	2,875%
Replace (Broken)	0,00%	0,00%	0,00%	5,111%
Change	0,00%	0,00%	0,00%	76,67%

Table 9 shows the results of the grouping of passport applicant data based on the

attributes of the type of application in each cluster (C0, C1, C2, C3). In general:

1. Cluster 0 (C0) is dominated by new applicants (72.31%) and replaced due to expiry (27.02%). This shows that this cluster mostly contains applicants with regular passport needs.
2. Cluster 1 (C1) has a similar pattern to C0, with most new applicants (71.12%) and replacement due to expiry (28.07%).
3. Cluster 2 (C2) is more varied, with 60.54% new applicants and 42.22% of expired replacements. This figure indicates a higher tendency to extension applications.
4. Cluster 3 (C3) is very different from other clusters, because it is dominated by data change applicants (76.67%). In addition, there are also applicants with the type of Replace full page (15.33%), replace damaged due to force majeure (2.875%), and replace damaged (5.111%).
5. From these results, it can be concluded that Clusters 0, 1, and 2 tend to reflect regular applicants (new or expired extensions), while Cluster 3 more specifically contains applicants who make changes or replacements of passports due to damage and certain conditions.

#### Attributes of the Purpose of the Passport

**Table 10.** Results of Processing Passport Destination Attributes

Data Name	C0	C1	C2	C3
Formal Work	0,00%	0,00%	13,14%	1,278%
Learn	0,00%	0,00%	14,20%	0,958%
Treatment	0,00%	0,00%	11,06%	0,319%
Hajj	0,00%	13,42%	65,28%	2,875%
TKI	2,534%	0,503%	0,00%	0,319%
Umrah	58,10%	54,05%	0,00%	43,77%
Tourism	39,48%	32,01%	0,00%	50,48%

The results of the processing of passport destination attributes showed that most of the applicants in clusters 0, 1, and 2 applied for passports for tourist purposes, respectively by 39.48% in cluster 0, 32.01% in cluster 1, and 50.48% in cluster 2. In addition, in cluster 0 there are still applicants with the purpose of *Umrah* at 12.53%, while in clusters 1 and 2 there are also study categories with percentages of 10.94% and 14.20%. In contrast to the three clusters, cluster 3 showed the dominance of worship purposes, namely *Umrah* by 43.45% and *Hajj* by 26.73%. This indicates that the pattern of passport applications for tourism activities is more widespread in the initial cluster, while the purpose of worship is more focused on the last cluster.

#### Attributes of Passport Type

**Table 11.** Passport Type Attribute Processing Results

Data Name	C0	C1	C2	C3
24 H Regular Passport	0,014%	0,00%	0,00%	0,00%
Ordinary Passport 48 H	99,77%	0,00%	96,21%	93,61%
10 Year Laminated Electronic Ordinary Passport	0,215%	0,00%	0,245%	6,390%
Laminated Electronic Ordinary Passport 5 Years	0,00%	2,069%	0,857%	0,00%
48 H Electronic Passport	0,00%	97,93%	6,365%	0,00%

Based on Table 11, almost all applicants in the four clusters apply for a 48-page Ordinary Passport with a very high percentage, which is more than 97%. Meanwhile, other types of passports such as Electronic Passports and 24-page Passports only appear with a very small percentage, even less than 1% in each cluster. This shows that people still tend to choose ordinary passports over electronic passports, whether due to need, ease of process, or limited availability of other types of passports.

*Gender Attributes*

**Table 12.** Results of Gender Attribute Processing

Data Name	C0	C1	C2	C3
Male	42,50%	43,78%	52,06%	43,35%
Women	57,50%	56,22%	51,61%	56,55%

The results of the gender attribute processing showed that the distribution of passport applicants was relatively balanced between males and females in most clusters. In clusters 0 and 1, the proportion of women was slightly more dominant, with percentages of 57.50% and 56.22%, respectively, while men were in the range of 42–44%. In contrast, cluster 2 had a higher proportion of men, with a percentage of 52.06%. Meanwhile, cluster 3 was dominated by women, with a percentage of 66.55%. This shows that although the number of male and female applicants was generally almost evenly balanced, there was a variation in dominance based on the characteristics of the clusters formed.

*Year Attributes*

**Table 13.** Year Attribute Processing Results

Data Name	C0	C1	C2	C3
2022	35,41%	0,00%	15,50%	38,66%
2023	44,71%	4,978%	66,38%	43,13%
2024	19,88%	94,60%	21,79%	18,21%

In the attribute of the year of application, the number of passport applications increased quite significantly in 2023. This can be seen in cluster 1, which is dominated by 2023 data at 47.98%, and cluster 2, which reaches 66.38%. In cluster 0, 2023 is also still dominant, with a percentage of 44.71%, although 2022 also holds a fairly large percentage of 35.41%. In contrast, in cluster 3, 2022 was the most prominent, with a percentage of 39.86%. The data for 2024 is still relatively small because the research only uses data up to the middle of the current year. These findings show that 2023 marked the peak of passport applications compared to other years.

A. Characteristics of Each Cluster (k=4)

1. Cluster 0 (C0) – New Applicants with *Umrah* Destinations.

This cluster is the largest group with a total of 20,953 items (about 68.3% of the total data). The distinctive features of this cluster are:

**Table 14.** Detail Characteristics of Cluster 0 (k=4)

Attributes	Dominant Description	Percentage (%)
Applicant Type	New	72,31%
Purpose of Passport	<i>Umrah</i> (58.10%) and Tourism (39,48%)	—
Passport Type	Ordinary Passport 48 Pages	99,77%
Gender	Women	57,50%
Dominant Year	2023	44,71%

This cluster describes the public who are making passports for the first time, with the main purpose of *Umrah* pilgrimage or religious tourism. Most applicants are women, and the peak of submissions occurs in 2023. This shows an increase in post-pandemic mobility, especially for religious departures abroad.

2. Cluster 1 (C1) – New Applicants in 2024 with *Umrah* and Electronic Passport Purposes

This cluster contains 7,152 items (about 23.3% of the total data).

The distinctive features of this cluster are:

**Table 15.** Detail Characteristics of Cluster 1 (K=4)

Categories	Remarks	Percentage (%)
Applicant Type	New	71,12%
Purpose of Passport	<i>Umrah</i>	54,05%
Passport Type	48 H Electronic Passport	97,93%
Gender	Women	56,22%
Dominant Year	2024	94,60%

This cluster represents new applicants in 2024 who have begun to switch to the use of electronic passports (e-passports). Most of them make submissions for *Umrah* purposes. This pattern shows an increasing trend in the use of e-passports among *Umrah* pilgrims, especially women, in line with the digitization of immigration services.

### 3. Cluster 2 (C2) – Passport Replacement Applicant for *Hajj*

This cluster has 2,541 items (about 8.3% of the total data). Its distinctive features:

**Table 16.** Detail Characteristics of Cluster 2 (K=4)

Categories	Remarks	Percentage (%)
Applicant Type	Replace (Expires)	42,22%
Purpose of Passport	<i>Hajj</i>	65,28%
Passport Type	Ordinary Passport 48 H	96,21%
Gender	More men	52,06%
Dominant Year	2023	66,38%

This cluster describes applicants who extend or replace old passports for the purpose of *Hajj*. The majority are male. This pattern shows that there is an annual cycle of passport applications ahead of the *Hajj* season, especially in 2023.

### 4. Cluster 3 (C3) – Applicant Change Data and Specific Passport Type

This smallest cluster consists of 313 items (about 1% of the total data). Its distinctive features:

**Table 17.** Detail Characteristics of Cluster 3 (K=4)

Categories	Remarks	Percentage (%)
Applicant Type	Change	76,67%
Purpose of Passport	Tourism (50.48%) and <i>Umrah</i> (43.77%)	—
Passport Type	Electronic Ordinary Passport 10 Years Lamination	6,39%
Gender	Women	56,55%
Dominant Year	2022	38,66%

This cluster represents a small group of applicants who make changes to passport data (e.g., a change of name or status), with the primary purpose of travel or personal reasons. They also started using special types of passports with 10-year validity, and many came from 2022 applications. This cluster can be considered a minority segment with special administrative needs.

The results of the grouping analysis using the K-Means algorithm on passport application data at the Muara Enim Non-TPI Class II Immigration Office provide many strategic advantages that can be the basis for decision-making and improve the quality of public services. By knowing the patterns and characteristics of each group, the immigration authorities can better understand the behavior and needs of the community.

First, the results of the grouping can be used as a tool in the decision-making process for planning and managing service resources. For example, data shows that most new passport applicants have *Umrah* and tourism destinations, especially in 2023 and 2024. This information

can be a reference for immigration authorities to prepare a more suitable placement of service officers during periods with a high number of applicants, such as ahead of the *Umrah* and *Hajj* seasons. In this way, the service process can run faster, more orderly, and more efficiently.

Second, the results of the grouping can be used to predict an increase in passport applications in the future. Based on the patterns analyzed from the 2022–2024 data, there is a tendency toward increased applications among new applicants for *Umrah* and tourism purposes. This prediction can help immigration authorities in planning strategies to deal with surges in applications, either by increasing service quotas, adjusting operating hours, or providing a more effective online queuing system.

Third, the results of the grouping also contribute to increasing the efficiency and effectiveness of immigration services. By grouping applicants based on the type of application, passport purpose, and applicant characteristics, immigration authorities can formulate more targeted and data-based service policies. For example, for groups of applicants who frequently reapply, such as those replacing expired passports, a dedicated service line or an automatic notification system regarding passport validity periods can be provided.

In addition, this analysis also opens opportunities for the development of immigration information systems that are responsive and capable of anticipating needs. By utilizing the results of the grouping, the system can be developed to automatically provide service recommendations to applicants according to their profile and needs. For example, the system can suggest a specific type of passport or estimate the best time to apply to avoid long queues.

From a managerial perspective, the results of the grouping can be used to assess the performance of public services at the Immigration Office. By examining the distribution of data per group, management can evaluate how effective the policies that have been implemented are in meeting the needs of the community across each applicant group. This analysis also serves as the basis for compiling annual performance reports grounded in empirical and analytical data.

The use of data mining techniques through the grouping method has been proven to make a positive contribution to the management of immigration services. The resulting patterns and segmentation not only help in understanding passport application trends, but also support a faster, more precise, and evidence-based decision-making process, in line with the digital transformation in modern immigration management.

## Discussion

### *Data Mining Test Results*

**Table 18.** Comparison of Current Data Processing and Data Mining

Aspects	Conventional	Data Mining
Main objectives	As administrative records and routine reports	Find patterns, trends, and groups to support verdict
Processing properties	Reactive, only Respond after data is collected	Proactive and predictive, able to provide a forward trend
Resulting method	Manual or administrative recording with simple tabulation	K-means Clustering algorithm using Python and Libraries Supporters
Information generated	Number of applicants per period, passport type, and basic aggregate data	Pattern of grouping applicants based on age, purpose, type Application
Quality of analysis	Limited, unable to explain the relationship between Variable	Comprehensive, can show the characteristics of each cluster
Utilization of results	Only as an agency report to the upper level	Strategic planning, service efficiency, and Resource allocation

The comparison between the data processing carried out today and the processing using data mining shows a significant difference. So far, the management of passport application data at the Muara Enim Non-TPI Class II Immigration Office is still conventional, being limited to administrative records and routine reporting. The information generated is limited to the number of applicants per period, the type of passport applied for, as well as other aggregate data without in-depth analysis. The system is reactive because it can only respond after the data has been collected, and does not provide an overview of future trends or predictions.

On the other hand, through the application of data mining techniques with the K-Means algorithm using Python, passport application data can be processed into more comprehensive information. Data that was originally only in tabulated form can be grouped into several clusters based on certain characteristics, such as age, purpose of passport application, gender, and year of application. These results not only show the number of applicants, but also reveal trends, patterns, and dominant groups present within a given period.

With data mining-based processing, the quality of analysis has improved significantly. The information produced becomes more predictive and can be used to support strategic decision-making, for example in human resource planning, determining service allocation, and formulating public service policies. Therefore, the use of data mining has proven to provide significant added value compared to conventional methods, because it not only functions as a documentation tool, but also to improve the efficiency and quality of immigration services

### CONCLUSION

Based on the results of the study, the application of data mining techniques with the K-means clustering method on passport application data at the Class II Non-TPI Immigration Office Muara Enim was proven to be able to provide a more comprehensive understanding of the application patterns over the past three years through grouping based on gender, travel destination, type of application, type of passport, and year of application. The clustering results showed a variety of applicants' characteristics, including the dominance of applicants for work and *Umrah* purposes as well as significant increases in certain periods such as the *Hajj* season and school holidays, so that these findings can serve as the basis for planning officer allocation and setting more effective service schedules. Compared to conventional data management that is administrative in nature, the K-means approach with the evaluation of the Davies-Bouldin index provides a more strategic and predictive analysis. Therefore, the integration of data mining in the passport management system is highly recommended to improve service efficiency, accuracy of data-based decision-making, and support more responsive and transparent public services, as well as to open opportunities for further research to develop more in-depth and applicable analytical methods.

### ACKNOWLEDGEMENT

The authors would like to express their gratitude to the Class II Non-TPI Immigration Office Muara Enim for providing access to passport application data and supporting the research process. Appreciation is also extended to Politeknik Pengayoman Indonesia for academic guidance and institutional support that enabled the completion of this study.

### AUTHOR CONTRIBUTION STATEMENT

Nurul Maharani Piranti conceptualized the study, conducted data analysis using K-Means clustering, and prepared the manuscript draft. Fadhlulrahman Khalish contributed to methodology development, data validation, and technical implementation. Okky Pratama Martadireja supervised the research process, provided critical revisions, and finalized the manuscript for publication.

### REFERENCES

- Aditya, Y. R. (2020). Application Of E-Government In Immigration Management Information System (SIMKIM) At Directorate General Of Immigration. *TEMATICS: Technology ManagemenT and Informatics Research Journals*, 2(2). <https://doi.org/10.52617/tematics.v2i2.107>
- Alfarisi Akasy, A. H. (2021). Analysis Of Immigration Management Information System (SIMKIM)

- Version 2.0 For The Issuance Of A Residence Permit With The Cobit 5 Method At The Class I Immigration Office TPI Makassar. *TEMATICS: Technology Management and Informatics Research Journals*, 3(1). <https://doi.org/10.52617/tematics.v3i1.303>
- Assiroj, P., Pribadhi, I. A., & Rizal S, M. (2023). A K-Means Algorithm For The Immigration Residential Permit. *Jurnal Ilmiah Kajian Keimigrasian*, 6(2). <https://doi.org/10.52617/jikk.v6i2.468>
- Azizah, N., Henderi, H., & Raja, B. H. P. M. (2025). Implementation of the K-Nearest Neighbor Algorithm for Classifying Immigration Residence Permit Applicants at the Class I Special Immigration Office TPI Soekarno-Hatta. *CCIT Journal*, 18(2). <https://doi.org/10.33050/ccit.v18i2.3577>
- Baghitz Hanan, Virra Wirdhingsih, & Sri Kuncoro Bawono. (2025). Inovasi Administratif dalam Pelayanan Keimigrasian: Menuju Birokrasi Modern dan Responsif. *JOURNAL OF ADMINISTRATIVE AND SOCIAL SCIENCE*, 6(1). <https://doi.org/10.55606/jass.v6i1.2107>
- Guohua, W., & Francis, E. H. (2017). Data Mining: Concept, Applications and Techniques. *ASEAN Journal on Science and Technology for Development*, 17(1). <https://doi.org/10.29037/ajstd.134>
- Hafiz-AM, F. (2019). Improving The Quality Of Learning In Immigration Polytechnics Through Immigration Management Information System Facilities Directorate General Of Immigration. *TEMATICS: Technology Management and Informatics Research Journals*, 1(2). <https://doi.org/10.52617/tematics.v1i2.78>
- Hamid Mughal, M. J. (2018). Data mining: Web data mining techniques, tools and algorithms: An overview. *International Journal of Advanced Computer Science and Applications*, 9(6). <https://doi.org/10.14569/IJACSA.2018.090630>
- Hapsari, T. S., & Nataliani, Y. (2024). Text Mining With Latent Dirichlet Allocation for Analyzing Public Comments on the M-Passport Application Text Mining Dengan Latent Dirichlet Allocation Untuk Analisis Komentar Masyarakat Terhadap Aplikasi M-Paspor. *Jurnal Teknik Informatika (JUTIF)*, 5(4).
- Herman, E., Zsido, K. E., & Fenyves, V. (2022). Cluster Analysis with K-Mean versus K-Medoid in Financial Performance Evaluation. *Applied Sciences (Switzerland)*, 12(16). <https://doi.org/10.3390/app12167985>
- Li, L. (2022). Implementation of a Community Data Processing System Based on Data Mining. *Journal of Robotics*, 2022. <https://doi.org/10.1155/2022/8286007>
- Luo, J., Xu, J., Aldosari, O., Althubiti, S. A., & Deebani, W. (2022). Design and Implementation of an Efficient Electronic Bank Management Information System Based Data Warehouse and Data Mining Processing. *Information Processing and Management*, 59(6). <https://doi.org/10.1016/j.ipm.2022.103086>
- Marksteiner, S., Sirjani, M., & Sjödin, M. (2024). Automated Passport Control: Mining and Checking Models of Machine Readable Travel Documents. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/3664476.3670454>
- Qing, M., Ma, Z., & Hu, Y. (2020). Application analysis of data mining technology in administration. *IOP Conference Series: Materials Science and Engineering*, 750(1). <https://doi.org/10.1088/1757-899X/750/1/012042>
- Ryanindityo, M., Aji, K. P., Purnomo, A. S., Nurkumalawati, I., Bawono, S. K., Wiraputra, A. R., Wilonotomo, & Pujonggo, S. S. (2024). The role of collaborative governance and total quality management on immigration services quality. *Journal of Infrastructure, Policy and Development*, 8(9). <https://doi.org/10.24294/jipd.v8i9.6654>
- Sianturi, M. W., Sirait, K. F., Nurfadillah, D., Shopia, S., Lubis, M., Muliani, F., & Saumi, F. (2025). Penerapan Analisis Cluster K-Means untuk Pengelompokan Kabupaten/Kota di Provinsi Sumatera Utara Berdasarkan Indikator Kemiskinan Tahun 2023. *Jurnal Pendidikan Matematika Malikussaleh*, 5(3). <https://doi.org/10.29103/jpmm.v5i3.21682>
- Sun, H. (2019). Study on application of data mining technology in university computer network educational administration management system. *Journal of Intelligent and Fuzzy Systems*, 37(3). <https://doi.org/10.3233/JIFS-179133>
- Susaningsih, C., & Puteri Andika, N. (2022). Optimization Of Immigration Data Management-Temple Application At The Balikpapan Immigration Office. *Jurnal Abdimas Imigrasi*, 3(2). <https://doi.org/10.52617/jaim.v3i2.435>

Wulansari, Z. (2019). Application Of Data Mining To Determine Product Administration Using Apriori Algorithm. *JOSAR (Journal of Students Academic Research)*, 4(1). <https://doi.org/10.35457/josar.v4i1.676>

Zhang, Z. (2025). Research and Implementation of Campus Network Intrusion Detection System based on Data Mining and Image Processing. *Scalable Computing: Practice and Experience*, 26(2). <https://doi.org/10.12694/scpe.v26i2.3962>