Journal of Information System and Technology Research

journal homepage: https://journal.aira.or.id/index.php/jistr/

OPTICS-Based Clustering of East Java Regencies and Cities by Divorce Factors

Cesaria Deby Nurhalizah[1], Aviolla Terza Damaliana[2,*], Dwi Arman Prasetya[3]

[1,2,3]Department of Data Science, Pembangunan National University Veteran of East Java, Indonesia

ARTICLE INFO

Article history:

Received July 10, 2025
Accepted Sept 27, 2025
Available online Sept 30, 2025

Keywords:

Divorce,
Divorce Factor,
OPTICS,
Clustering

ABSTRACT

Divorce is a social phenomenon that occurs when a married couple decides to legally end their marriage. This decision is influenced by various factors such as conflict, economic pressure, domestic violence, and deviant behavior. The aim of this study is to group regencies and cities in East Java Province that share similarities in the main causes of divorce, in order to understand the patterns that emerge across regions. The OPTICS (*Ordering Points to Identify the Clustering Structure*) clustering method was chosen for its ability to identify cluster structures with varying densities. The modeling process was conducted using a proportion-based approach for each causal factor, with optimal parameters obtained through manual grid search using *min_samples* = 2, *xi* = 0.05, and *min_cluster_size* = 0.1. The analysis identified three main clusters, each dominated by conflict, economic hardship, and deviant behavior, respectively. The quality of the clustering was evaluated using a Silhouette Score of 0.588, indicating reasonably good results. These findings are expected to serve as an initial understanding of divorce causes in East Java and can be used as input for the formulation of more targeted social policies.

Corresponding Author:

Aviolla Terza Damaliana,
Department of Data Science, Pembangunan National University Veteran of East Java,
1 Rungkut Madya St., Gunung Anyar, Surabaya, East Java 60294, Indonesia.
Email: aviolla.terza.sada@upnjatim.ac.id

1. INTRODUCTION

Divorce is defined as the legal dissolution of a marital union between a husband and wife, declared by the court and enforced once it obtains permanent legal status, effective from the date of marriage registration [1]. It constitutes a complex social phenomenon with far-reaching consequences, affecting not only the couple and their families but also the broader social and economic fabric of society. According to the Central Bureau of Statistics (*Badan Pusat Statistik* or BPS), the number of divorce cases in Indonesia reached 463,654 in 2023. East Java recorded the second-highest number of cases, with 88,213 divorces. This alarming figure reflects a growing trend that may be attributed to a weakening perception of the sanctity of marriage and the relatively simple procedures required to obtain a divorce [2].

Each regency and city exhibit distinct characteristics in terms of divorce causation, such as prolonged conflict, financial hardship, domestic violence, and abandonment [3]. These variations arise from the unique socioeconomic and interpersonal conditions faced by each household [4]. Increasing awareness among couples regarding the state of their marital relationships also contributes to decisions to terminate the union when issues remain unresolved. However, premarital education, which is ideally intended to serve as a foundation for family resilience, is often perceived merely as an administrative formality [5]. As a result, many couples are inadequately prepared to handle conflicts within the household and may perceive divorce as the only viable solution.

To address this issue, the government has introduced several intervention measures, including the establishment of the Advisory Board for Marriage and Divorce Disputes (*Badan Penasehat Perselisihan Perceraian dan Perkawinan* or BP4) under the Office of Religious Affairs (*Kantor Urusan Agama* or KUA), and the implementation of family economic empowerment programs, recognizing that economic stress is a major contributing factor to divorce [6]. Despite these efforts, the effectiveness of such policies remains limited, as evidenced by the continuing rise in divorce rates over the years [7].

Several previous studies have explored the application of clustering algorithms in social and health domains. For example, a study by Hastuti *et al*. applied the OPTICS (Ordering Points to Identify the Clustering Structure) algorithm to spatial health data. Using a minimum points (*MinPts*) parameter of 2 and a ξ value of 0.5, the algorithm successfully generated five distinct clusters along with several noise points, achieving a Silhouette Score of 0.607. These findings demonstrate the capability of OPTICS to uncover complex cluster structures, particularly in settings characterized by heterogeneous distributions, such as healthcare worker deployment. In comparison, the ST-DBSCAN algorithm used in the same study resulted in only two clusters and a significantly lower Silhouette Score of 0.329. Conceptually, the strength of OPTICS lies in its use of *reachability distance*, which does not rely on a predefined threshold such as *epsilon*. This allows for greater flexibility in detecting clusters within datasets that exhibit varying densities [8].

In the context of divorce in Indonesia, Nurhayati *et al*. employed the K-Means clustering algorithm to analyze divorce causative factors at the provincial level. The study utilized four main variables, namely ongoing disputes, economic hardship, domestic violence, and abandonment, and successfully grouped 29 provinces into two clusters using RapidMiner software. The results distinguished between three provinces with high divorce intensity and 26 provinces with lower levels of causation.

Building on the insights from both studies, the present research applies the OPTICS algorithm in a different social context, namely the clustering of administrative regions based on divorce data at the regency and city level in East Java for the year 2024. Utilizing the same four primary divorce factors, this study aims to identify clusters of regions that exhibit similar divorce patterns. The results are expected to contribute valuable insights into regional divorce dynamics and support the development of more targeted and contextually appropriate intervention strategies tailored to the characteristics of each region [1].

2. RESEARCH METHOD

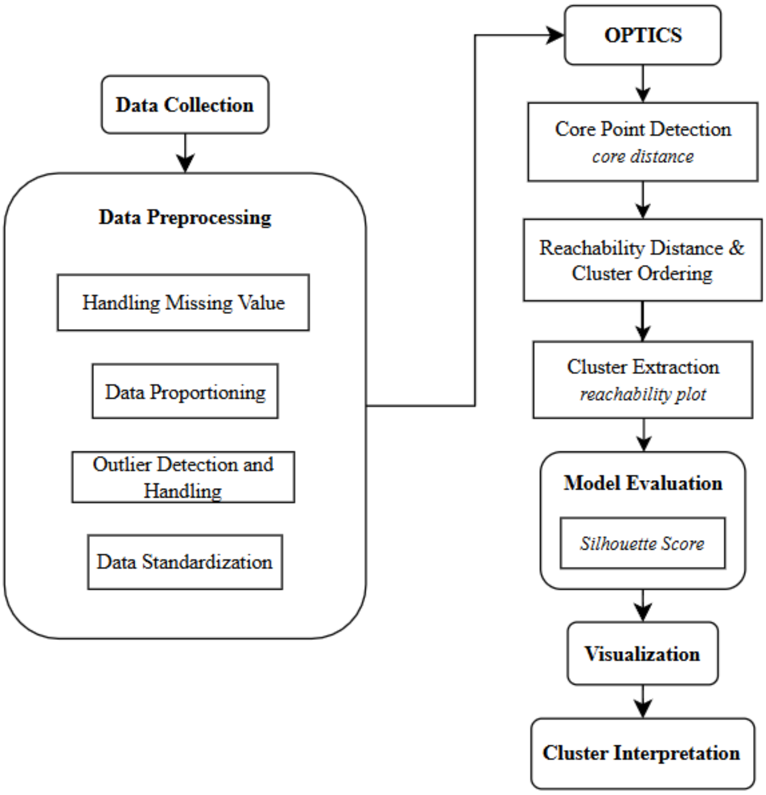

Figure 1. Research Method Diagram

Figure 1 illustrates the overall workflow of this research, which is organized systematically, starting from data collection to cluster interpretation. This study applies a clustering approach within the field of data mining to group regencies and cities in East Java Province based on the similarity in the distribution of divorce causes. Clustering is an exploratory technique used to uncover natural structures in unlabeled data by organizing objects into groups that share common characteristics [9] [10]. In this research, the OPTICS (*Ordering Points to Identify the Clustering Structure*) algorithm is employed as a density-based clustering method that was developed to overcome the limitations found in the DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) algorithm [11]. OPTICS functions by arranging data points based on their local density and calculating two key parameters, namely *core distance* and *reachability distance* [12].

In contrast to DBSCAN, OPTICS does not require the predefined number of clusters and is capable of detecting non-globular cluster structures while accommodating variations in density between clusters. The result of the OPTICS algorithm includes a *reachability plot* that can be used for visual cluster extraction or further analysis through auxiliary methods. These characteristics make OPTICS a more flexible and robust tool, especially for analyzing complex social data such as the distribution of divorce causes, which may vary in nonlinear and heterogeneous ways across regions.

The clustering results obtained in this study are expected to provide meaningful insights that can support government efforts in addressing divorce issues in East Java. Each regency or city may then benefit from more targeted and context-specific intervention programs that reflect the unique characteristics of the region.

2.1 Data Collection

The data used in this study are secondary data obtained from the Central Bureau of Statistics (Badan Pusat Statistik or BPS). The dataset includes the number of divorce cases along with their causal factors across all regencies and cities in East Java Province for the year 2024. These factors include disputes and quarrels (X_1), economic problems (X_2), domestic violence (X_3), and abandonment by one party (X_4). The data are presented in tabular form, where each row represents a specific regency or city, and each column represents the value of each factor as well as the total number of divorce cases. The table format is illustrated in Table 1 below.

Tabel 1. Data Structure

Regency/City	X_1	X_2	X_3	X_4	**Total Divorce**
[Region 1]	$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	983
[Region 2]	$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	983
….	…	…	…	…	…
[Region 38]	$X_{38,1}$	$X_{38,2}$	…	$X_{38,4}$	…

2.2 Data Preprocessing

Data preprocessing is conducted to ensure the quality and consistency of the data prior to the clustering modeling process [13]. The following steps are involved in this stage:

1. Handling Missing Values

The handling of missing values begins with filtering out data entries that do not report any divorce cases [14]. All rows in the divorce count attribute with a value of zero are removed, as these rows do not represent relevant divorce events for analysis. Subsequently, missing values in the divorce factor variables are assumed to indicate the absence of cases and are imputed with a value of zero. This approach is employed to preserve data integrity without introducing bias or eliminating important informational structures.

2. Data Proportioning

Following the data cleaning process, each variable representing a divorce cause is expressed in proportional form by dividing the count of each factor by the total number of divorce cases in the respective administrative region.

$$\text{Factor Proportion} = \frac{\text{Number of Cases by Factor (X)}}{\text{Total Divorce Cases}} \tag{1}$$

In other words, the proportion is calculated by dividing the number of cases attributed to a specific divorce factor by the total number of divorce cases within the same region. This process is implemented to standardize the data across regencies and cities, considering that a single divorce case may involve multiple causes and that the total number of divorce cases varies significantly between regions. By using proportions, interregional comparisons become more representative and objective, while reducing bias caused by differences in absolute scale [15].

3. Outlier Detection and Handling

The next step involves detecting and handling outliers that may compromise the validity of the clustering results. Detection is carried out using a statistical approach based on the interquartile range (IQR), where the upper and lower thresholds are determined according to the quartile distribution of each variable, calculated using the following formula [16]:

$$\text{IQR} = Q_3 - Q_1 \quad (2)$$

where Q_1 and Q_3 represent the first and third quartiles, respectively. A value is classified as an outlier if it falls outside the following range:

$$\text{Lower Bound} = Q_1 - 1.5 \times IQR \quad (3)$$

$$\text{Upper Bound} = Q_3 + 1.5 \times IQR \quad (4)$$

Explanation:
Q_1 (first quartile) : the value that marks the lowest 25% of the data.
Q_3 (third quartile) : the value that marks the lowest 75% of the data.

To preserve the stability of the distribution without removing any observations, the winsorization technique is applied by limiting extreme values to within statistically acceptable boundaries. This process ensures that the data distribution remains representative and free from distortions caused by excessive outliers [17].

4. Data Standardization

This process is carried out using the *z*-score normalization method through the use of *StandardScaler*, which transforms the data to have a distribution with a mean (μ) of 0 and a standard deviation (σ) of 1, thereby eliminating scale bias across variables [16]. The formula is defined as follows:

$$z = \frac{x - \mu}{\sigma} \quad (5)$$

Explanation:
z : standardized value
x : original variable value
μ : mean of the variable
σ : standard deviation of the variable

Standardization is essential to prevent certain variables from dominating distance calculations during the clustering process and to ensure the stability of the model in high-dimensional space [14].

2.3 OPTICS Algorithm

OPTICS (*Ordering Points to Identify the Clustering Structure*) is a density-based clustering algorithm that extends DBSCAN to better detect clusters in datasets with non-uniform density distributions [12]. Unlike DBSCAN, which directly assigns cluster labels based on a predefined distance threshold (ε) and minimum number of points (*MinPts*), OPTICS does not produce a fixed clustering result. Instead, it creates an augmented ordering of the dataset and stores two key distance metrics for each point to reveal the inherent clustering structure [18]. The two primary metrics used in OPTICS are:

1) *Core distance*: The smallest radius required to include *MinPts* neighbours around a data point. Formally, it is the distance to the *MinPts*-th nearest neighbour. If a point has fewer than *MinPts* neighbours within any radius, its core distance is undefined, meaning the point cannot serve as a core point.

$$[\text{core\_dist}(p) = \text{dist}(p, o_{(k)}) \quad \text{where } k = \text{minPts}] \quad (6)$$

2) *Reachability distance*: Defined between two points, it expresses how strongly a point is density-reachable from a core point. Specifically, for a point p and its neighbour o, the reachability distance is given by

$$\text{reachability-dist}(p, o) = \max(\text{core-dist}(o),\ \text{dist}(p, o)) \quad (7)$$

This metric ensures that the reachability from a point depends both on the local density around the core point and the actual distance to the target point.

OPTICS not only calculates distance metrics, but also produces a cluster ordering, which is a linear sequence of the data points reflecting the underlying clustering structure. This ordering preserves spatial relationships between points with similar density levels and facilitates cluster extraction, either visually through the reachability plot or via algorithmic methods.

An important feature in OPTICS is the ξ (xi) parameter, which is used during the extraction phase to automatically detect significant drops in reachability distances that indicate cluster boundaries. By adjusting the ξ value, users can fine-tune the sensitivity of cluster detection and control the level of detail in the resulting clusters.

The output of the OPTICS algorithm is a reachability plot, a 1-dimensional representation of the reachability distances ordered according to the processing sequence. Valleys in the plot correspond to clusters, and their depth and width reflect the cluster's density and size. Clusters can then be extracted either visually or through post-processing methods.

The key advantage of OPTICS lies in its ability to detect clusters of arbitrary shape and varying density without a strict dependence on global parameters. This flexibility makes it well-suited for analysing real-world data, such as social or behavioural datasets, which often exhibit heterogeneous structures. The workings of the OPTICS algorithm can be concisely described using the following pseudocode, which outlines the step-by-step procedure implemented in the clustering process:

Algorithm 1: OPTICS Algorithm

```
Input   : D (set of data points), ε (maximum radius), MinPts (minimum points)
Output  : Ordered list of points with core distances and reachability distances

Initialize all points in D as unprocessed
Initialize an empty list called ordered_list

for each unprocessed point p in D do
   mark p as processed
   core_dist_p ← CoreDistance(p, ε, MinPts)     ← [Core Point Detection]
   append p to ordered_list
   if core_dist_p ≠ UNDEFINED then
      Seeds ← empty priority queue
      update(Seeds, p, ε, MinPts)            ← [Reachability Distance & Cluster Ordering]
       while Seeds is not empty do
          q ← Seeds.pop()                  ← next point with smallest reachability distance
          mark q as processed
          core_dist_q ← CoreDistance(q, ε, MinPts)
          append q to ordered_list
          if core_dist_q ≠ UNDEFINED then
             update(Seeds, q, ε, MinPts)
Return ordered_list with reachability distances  ← [Cluster Extraction via Reachability Plot]
```

The presented pseudocode outlines the main steps of the OPTICS algorithm, including core point detection, reachability distance computation, and the generation of cluster ordering. Each data point is processed to compute its core distance, which determines whether it qualifies as a core point based on the *MinPts* parameter. The algorithm then updates reachability distances for neighboring points and constructs an ordered list that preserves the spatial density relationships among points. This ordering facilitates the extraction of clusters from the resulting reachability plot. In this study, the pseudocode implementation was adapted to analyze divorce causation data across East Java, allowing the identification of clusters based on regional similarities in the distribution of divorce factors.

2.4 Model Evaluation

The silhouette score is an evaluation metric used to measure the accuracy and consistency of clustering results [19]. It assesses how well each data point fits within its assigned cluster compared to other clusters. A higher silhouette value indicates that the point is well matched to its own cluster and poorly matched to neighboring clusters [17]. Conversely, negative values suggest that a point may have been incorrectly or ambiguously assigned to a cluster. For each data point *i*, the silhouette value is computed using the following formula:

$$s = \frac{b(i)-a(i)}{max(a(i),b(i))} \tag{7}$$

Explanation:
s : silhouette score
$a(i)$: the average distance between point *i* and all other points in the same cluster (cluster cohesion).
$b(i)$: the average distance between point *i* and the nearest neighboring cluster (cluster separation).

In this study, the silhouette score is employed to evaluate the quality of clusters generated by the OPTICS algorithm. Although OPTICS does not explicitly determine the number of clusters, this metric remains relevant for assessing the coherence and separation of the resulting clusters. It also provides insights into the reliability of the clustering structure.

3. RESULTS AND DISCUSSION

The initial stage of analysis began with data preprocessing to ensure the validity and distribution of each divorce causation variable. Figure 2 displays histograms for each divorce causation variable, aimed at visually observing the distribution patterns [20]. Most variables exhibit a positively skewed distribution, indicating that extreme values occur more frequently in the higher range of the data.

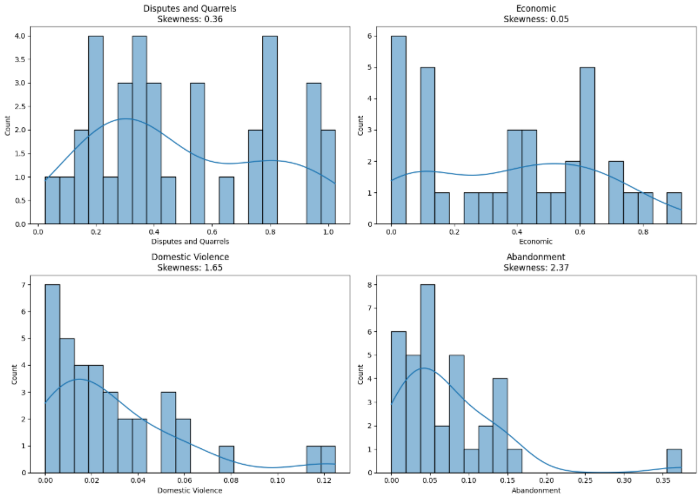

Figure 2. Histograms of Divorce Causation Factors with Skewness Indicators

Following the distribution analysis, boxplot visualizations were used to detect outliers in each variable, as shown in Figure 3. Outliers were particularly observed in the domestic violence and abandonment variables.

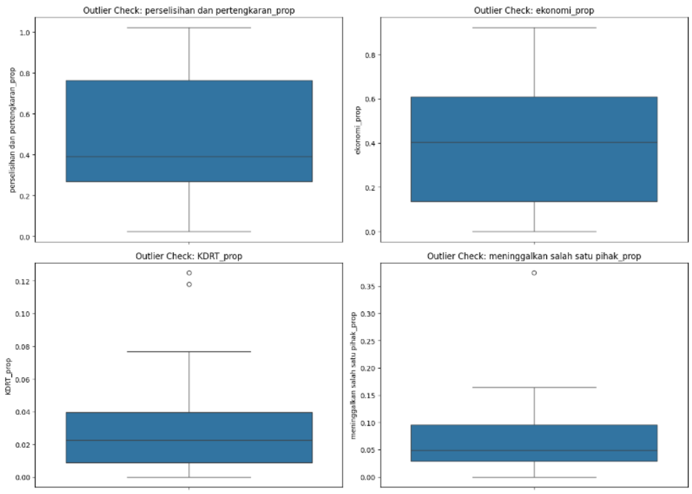

Figure 3. Boxplots of Divorce Causation Factors Before Winsorization

These outliers have the potential to affect the modeling outcomes; therefore, a Winsorization technique was applied to reduce their impact without removing any data points. After preprocessing, Figure 4 shows boxplots that reflect a more concentrated and symmetrical data distribution.

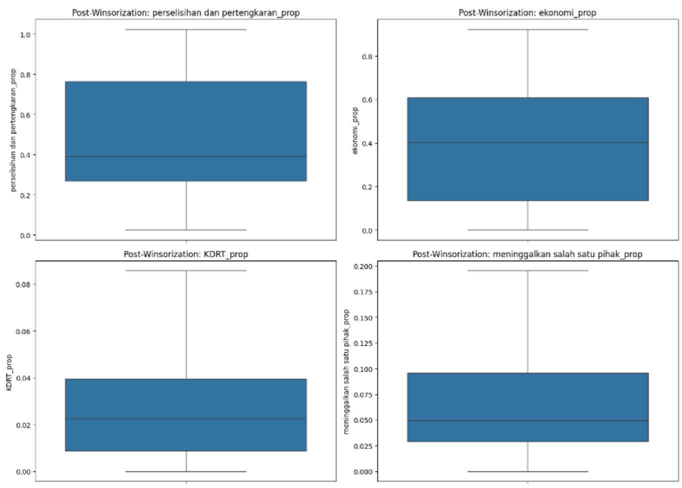

Figure 4. Boxplots of Divorce Causation Factors After Winsorization

Clustering in this study was performed using the OPTICS algorithm with a minimum points (*MinPts*) parameter of 2 and a ξ value of 0.5. The *MinPts* parameter defines the minimum number of points required to form a cluster, while ξ indicates sensitivity in identifying changes in data density structure. The initial analysis focuses on two key metrics in the OPTICS algorithm, namely *core distance* and *reachability distance*.

Table 2. *Core Distance* dan *Reachability Distance Statistics*

Statistic	***Core Distance***	***Reachability Distance***
Number of finite values	35	34
Number of infinite values	0	1
Mean	0.7241	0.8175
Standard deviation	0.3509	0.3465
Minimum	0.2184	0.2184
Maksimum	1.6694	1.6694

Table 2 presents the statistical summary of the core distance and reachability distance values obtained through the OPTICS algorithm. All 35 data points recorded finite core distances, indicating that each point had a sufficient number of neighbouring data points within the specified radius to be considered for clustering. This suggests a relatively dense distribution across the dataset. In contrast, only one point exhibits an infinite reachability distance, which is commonly observed in the initial point of the processing sequence, as it has no preceding reference for comparison.

The mean core distance is 0.7241, and the mean reachability distance is slightly higher at 0.8175, implying that the spatial reachability between points is generally broader than the minimum local density required for forming a core point. This difference reflects the nature of how OPTICS prioritizes connectivity across varying density levels. The standard deviations for both metrics are relatively similar (0.3509 for core distance and 0.3465 for reachability distance), suggesting a comparable spread in the distribution of local density and inter-point reachability [21]. Furthermore, the identical minimum and maximum values for both core and reachability distances (0.2184 and 1.6694, respectively) demonstrate consistency in the density and distance structure of the dataset. This consistency supports the robustness of OPTICS in identifying underlying clusters within data exhibiting heterogeneous density characteristics.

The spatial and density-based relationships observed in these statistics are further visualized through the reachability plot in Figure 2, which highlights the distribution of reachability distances in sequential order. This plot serves as a diagnostic tool for interpreting the clustering structure formed by the algorithm.

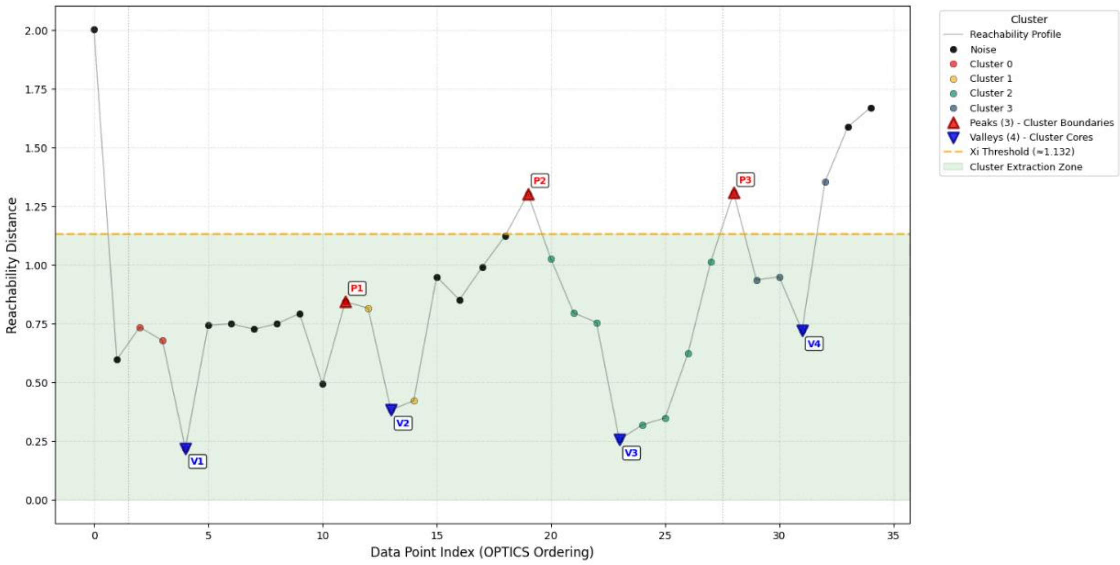

Figure 5. *Reachability Plot*

Figure 5 presents the reachability plot, which visualizes the reachability distances of each data point based on the processing order of the OPTICS algorithm. The resulting visual pattern reveals four major valleys (V1–V4) representing high-density cluster cores, and three prominent peaks (P1–P3) that signify the boundaries between clusters. The orange horizontal line indicates the cluster extraction threshold set at $\xi = 1.132$, which defines the cluster extraction zone shaded in green. Data points exceeding this threshold are classified as noise, as they do not meet the minimum density requirements.

This visualization reinforces the clustering results, which identified four main clusters and classified 15 data points as noise, represented by black dots in the graph. The structure clearly demonstrates the effectiveness of the OPTICS algorithm in detecting complex and heterogeneous data distributions and offers an insightful depiction of the local density structure and the separation between regions. These clusters reflect meaningful groupings based on the distribution of divorce factors in East Java. The details of regency and city members within each cluster are presented in Table 3.

Table 3. OPTICS Cluster Labels

Cluster Label	Number of Members	Cluster Members
Noise	15	Ngawi, Bangkalan, Madiun, Situbondo, Kota Malang, Mojokerto, Trenggalek, Bojonegoro, Lamongan, Sumenep, Kediri, Kota Kediri, Pacitan, Tuban, Malang
Cluster 0	3	Ponorogo, Jember, Banyuwangi
Cluster 1	3	Bondowoso, Gresik, Kota Pasuruan
Cluster 2	9	Probolinggo, Pasuruan, Sidoarjo, Jombang, Magetan, Sampang, Pamekasan, Kota Probolinggo, Kota Surabaya
Cluster 3	5	Tulungagung, Blitar, Lumajang, Nganjuk, Kota Madiun

Based on the clustering results presented in Table 3, five regional groups were identified, including one group classified as noise. The noise cluster consists of 15 regencies and cities that do not exhibit sufficiently strong or consistent patterns in the distribution of divorce factors to be assigned to any of the main clusters. Cluster 0 comprises three regions: Ponorogo, Jember, and Banyuwangi. Cluster 1 includes Bondowoso, Gresik, and Pasuruan City. Cluster 2, which has the second-highest number of members, contains nine regions, including Surabaya City, Probolinggo, and Sidoarjo. Meanwhile, Cluster 3 includes five regions such as Tulungagung, Blitar, and Madiun City. These findings indicate the presence of spatial variation in divorce causation patterns across East Java Province, successfully revealed through the application of the OPTICS algorithm.

Table 4. OPTICS Cluster Means

Cluster Label	Disputes and Quarrels	Economic	Domestic Violence	Abandonment
Cluster -1	-0.511	0.470	0.147	0.185
Cluster 0	-1.150	1.412	-0.293	-0.317
Cluster 1	-0.425	0.116	2.261	-0.581
Cluster 2	1.366	-1.132	-0.744	-8.852
Cluster 3	0.018	-0.288	-0.281	1.518

Table 4 presents the standardized mean values of each divorce causation variable based on the clustering results obtained through the OPTICS algorithm. The results show that each cluster exhibits distinct characteristics in terms of dominant causative factors. Cluster 0 is strongly dominated by economic issues, indicating that the regions within this group experience financial pressure as the primary cause of divorce. Cluster 1 is characterized by a predominance of domestic violence, suggesting a high level of physical or emotional conflict within households in these areas. Meanwhile, Cluster 2 shows the highest value in disputes and quarrels, pointing to persistent interpersonal conflict as the main driver of divorce. Cluster 3 stands out for abandonment, which may reflect a lack of commitment or physical presence of a partner within the marriage. The noise cluster (labeled -1) displays relatively balanced values across all variables, without a single dominant cause. This indicates that the regions categorized as noise do not exhibit consistent divorce patterns and therefore cannot be classified into any of the main clusters. These findings support the interpretation that the clustering process effectively groups regions according to differing underlying patterns of divorce causation.

Tabel 5. Silhouette Score of Each Cluster

Cluster Label	***Silhouette Score***
Overall Average	0.588
Cluster 0	0.807
Cluster 1	0.850
Cluster 2	0.620
Cluster 3	0.241

The results presented in Table 5 indicate that the overall quality of the clustering is at a reasonably good level, with an average Silhouette Score of 0.588. Clusters 1 and 0 achieved the highest scores, indicating well-separated and clearly defined structures, while Cluster 3 recorded the lowest score, suggesting that this group is less optimally defined compared to the others. Table 5 presents the Silhouette Scores for each cluster generated by the OPTICS algorithm. The Silhouette Score is a metric used to evaluate the quality and validity of clustering results, ranging from -1 to 1. A score close to 1 indicates that a data point is well-matched to its own cluster and distinctly distant from other clusters, suggesting strong separation. Conversely, values near zero or negative suggest suboptimal clustering.

Overall, the average Silhouette Score of 0.588 indicates that the clustering result is reasonably good and coherent. Cluster 1 achieved the highest score of 0.850, followed by Cluster 0 with a score of 0.807. These two clusters exhibit the clearest structure and strongest separation. Cluster 2 also showed a fairly good performance with a score of 0.620, though not as strong as the previous two.

On the other hand, Cluster 3 recorded the lowest score of 0.241. This suggests that the data points within this cluster are less clearly grouped and may overlap with other clusters. In other words, Cluster 3 is less well-defined compared to the others. This outcome provides valuable insight for further evaluation, such as adjusting clustering parameters or considering alternative clustering approaches.

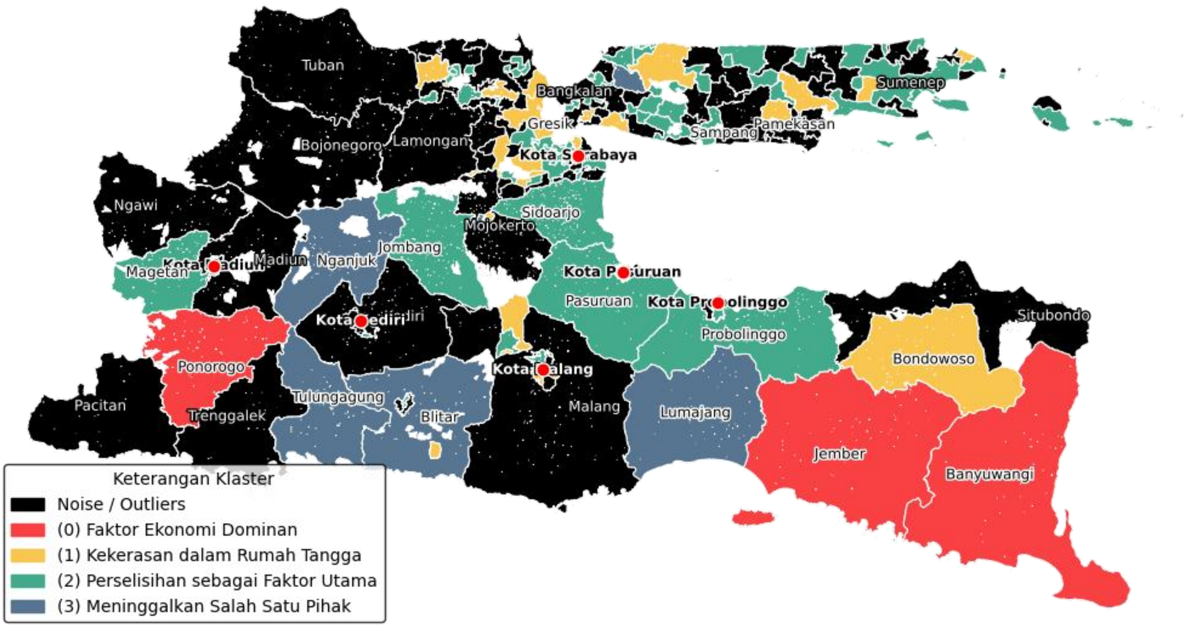

Figure 6. Clustering Map of Divorce Factors in East Java

The visualization in Figure 6 presents the clustering results of regencies and cities in East Java Province based on divorce causation patterns using the OPTICS algorithm. Visual tools such as maps and graphs help represent complex data structures for easier interpretation [22]. Each color represents a different cluster, where red indicates regions with a high proportion of divorces due to economic factors, such as Jember and Banyuwangi, yellow represents areas with a significant number of divorces caused by domestic violence, such as Bondowoso, green refers to regions like Surabaya and Probolinggo where disputes and quarrels are the dominant cause, while blue highlights areas such as Tulungagung and Blitar where abandonment is the main factor. Meanwhile, black marks regions classified as noise or outliers, which do not exhibit a dominant divorce pattern. This spatial visualization provides a comprehensive overview of divorce characteristics across regions and serves as a foundation for developing more targeted and region-specific policy interventions.

4. CONCLUSION

This study successfully clustered regencies and cities in East Java based on the distribution of divorce causative factors using the OPTICS algorithm. The method was applied with the parameters *min samples* set to 2 and *xi* set to 0.5, resulting in four main clusters and identifying 15 instances as noise. Each cluster demonstrated a distinct dominant characteristic. Cluster 0 was characterized by economic issues, Cluster 1 by domestic violence, Cluster 2 by disputes and quarrels, and Cluster 3 by abandonment. The clustering quality was evaluated using the Silhouette Score, which produced a value of 0.588. This indicates a reasonably good level of separation between clusters. These findings confirm that the OPTICS algorithm is capable of effectively identifying clustering structures based on the distribution of divorce causes in regions with diverse characteristics.

However, this study has certain limitations, particularly regarding the relatively limited dataset and the exclusive use of a single clustering method. Therefore, future research is encouraged to compare or combine OPTICS with other clustering techniques to obtain more comprehensive results. Additionally, expanding the dataset in terms of both temporal and spatial scope is expected to enhance the accuracy and reliability of the analysis, as well as enrich the understanding of divorce patterns across different regions.

5. ACKNOWLEDGEMENTS

The author would like to express sincere gratitude to all parties who have contributed to the successful completion of this study. Special thanks are extended to the academic supervisors for their valuable guidance, constructive feedback, and continuous support throughout the research process. Without their assistance, this study would not have been completed effectively. The researcher also wishes to thank the institution that provided access to essential resources and research data, which served as the foundation for the analysis. This support has been instrumental in ensuring the smooth execution and success of this research. It is hoped that the findings of this study will contribute positively to the development of scientific knowledge and offer meaningful benefits to the wider community.

6. REFERENCES

[1] N. Nurhayati, F. Azzahra, S. Ramadani, and ..., “Analisis Metode Klastering Pada Kasus Penyebab Perceraian Berdasarkan Provinsi Dengan Teknik K-Means,” *KOMIK (Konferensi ...*, vol. 4, pp. 278–284, 2020, doi: 10.30865/komik.v4i1.2699.

[2] G. Salsabila and A. Rofi, “Analisis Konteks Wilayah Terhadap Perceraian Di Provinsi Jawa Timur,” *J. Ilmu Kel. dan Konsum.*, vol. 15, no. 1, pp. 1–13, 2022, doi: 10.24156/jikk.2022.15.1.1.

[3] N. Januari, “MENGGALI AKAR MASALAH: Analisis Kasus Perceraian di Indonesia,” *Akad. J. Mhs. Humanis*, vol. 3, no. 3, pp. 120–130, 2023, doi: 10.37481/jmh.v3i3.613.

[4] S. A. Izzah, I. Zain, and E. O. Permatasari, “Pengelompokan Kabupaten/Kota Berdasarkan Faktor Penyebab Perceraian di Provinsi Jawa Barat,” *J. Sains dan Seni ITS*, vol. 11, no. 3, 2022, doi: 10.12962/j23373520.v11i3.77018.

[5] N. S. Manna, S. Doriza, and M. Oktaviani, “Cerai Gugat: Telaah Penyebab Perceraian Pada Keluarga di Indonesia,” *J. Al-AZHAR Indones. SERI Hum.*, vol. 6, no. 1, p. 11, 2021, doi: 10.36722/sh.v6i1.443.

[6] Sumiati, “Peranan Badan Penasihat Pembinaan Dan Pelestarian Perkawinan (Bp4) Dalam Memberikan Penataran Dan Bimbingan Perkawinan Di Kantor Urusan Agama (Kua) Kabupaten Maros,” *Visipena J.*, vol. 9, no. 2, pp. 342–357, 2018, doi: 10.46244/visipena.v9i2.464.

[7] M. Mahanani, U. Dwi Safiroh, and M. A. Tubagus, “Rekomendasi Solusi Atas Kasus Perceraian Di Kabupaten Lamongan Berbasis K-Means Clustering,” *Semin. Nas. Teknol. Inf. dan Komun. (SeNTIK STI&K) STMIK Jakarta STI&K*, vol. 12, p. 2024, 2024.

[8] S. H. Hastuti, A. Septiani, H. Hendrayani, and W. P. Nurmayanti, “Penerapan Metode OPTICS dan ST-DBSCAN untuk Klasterisasi Data Kesehatan,” *Edumatic J. Pendidik. Inform.*, vol. 8, no. 1, pp. 252–261, 2024, doi: 10.29408/edumatic.v8i1.25765.

[9] S. Hidayati, A. T. Darmaliana, and R. Riski, “Comparison of K-Means, Fuzzy C-Means, Fuzzy Gustafson Kessel, and DBSCAN for Village Grouping in Surabaya Based on Poverty Indicators,” *J. Pendidik. Mat.*, vol. 5, no. 2, p. 185, 2022, doi: 10.21043/jpmk.v5i2.16552.

[10] Elvi, “Analisis Algoritma K-Means Dalam Pengelompokan Perkara Perceraian Berdasarkan Kelurahan Di Kota Jambi,” *J. Process.*, vol. 16, no. 1, pp. 9–19, 2021, doi: 10.33998/processor.2021.16.1.920.

[11] M. A. Ahmed, H. Baharin, and P. N. E. Nohuddin, “Analysis of K-means, DBSCAN and OPTICS Cluster algorithms on Al-Quran verses,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 8, pp. 248–254, 2020, doi: 10.14569/IJACSA.2020.0110832.

[12] K. Dan, S. Dengan, S. Optics, and D. A. N. Dbscan, "ALGORITMA ORDERING POINTS TO IDENTIFY THE CLUSTERING".

[13] A. Salsabiela, A. P. Kuncoro, P. Subarkah, and P. Arsi, "Rekomendasi Restock Barang di Toko Pojok UMKM Menggunakan Algoritma K- Means Clustering," vol. 5, no. 2, pp. 3–8, 2024, doi: 10.37802/joti.v5i2.554.

[14] A. T. Damaliana, A. Muhaimin, and D. A. Prasetya, "Forecasting the Occupancy Rate of Star Hotels in Bali Using the Xgboost and Svr Methods," *J. Stat.*, vol. 12, no. 1, pp. 24–33, 2024, doi: 10.14710/JSUNIMUS.

[15] A. T. Damaliana, A. Muhaimin, and P. A. Riyantoko, "Peramalan Lonjakan Kasus Harian Covid-19 Di Indonesia Dengan Model Arima," *Pros. Semin. Nas. Sains Data*, vol. 3, no. 1, pp. 184–189, 2023, doi: 10.33005/senada.v3i1.112.

[16] J. Anggara, F. R. Ramadhan, and A. F. Syofian, "Pengembangan Sistem Prediksi Harga dan Rekomendasi Mobil Bekas Berbasis Machine Learning," vol. 7, no. 1, pp. 11–23, 2025.

[17] D. A. Prasetya, A. P. Sari, M. Idhom, and A. Lisanthoni, "Optimizing Clustering Analysis to Identify High-Potential Markets for Indonesian Tuber Exports," *Indones. J. Electron. Electromed. Eng. Med. Informatics*, vol. 7, no. 1, pp. 113–122, 2025, doi: 10.35882/skzqbd57.

[18] Mustikasari and N. Salman, "Analisis klaster berbasis kepadatan dengan dbscan dan optics," *Insypro Inf. Syst. Process.*, vol. 8, no. 1, pp. 1–8, 2023, [Online]. Available: http://journal.uinalauddin.ac.id/index.php/insypro

[19] I. N. Simbolon and P. D. Friskila, "ANALISIS DAN EVALUASI ALGORITMA DBSCAN SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE) PADA TUBERKULOSIS," vol. 12, no. 3, 2024.

[20] S. Muhajannah, A. R. Kadisi, and N. Wakhidah, "Penerapan Regresi Linear Terhadap Tingkat Perceraian Di Jawa Tengah Menggunakan Bahasa Python," *J. Tekno Kompak*, vol. 19, no. 2, pp. 76–90, 2025, doi: 10.33365/teknokompak.v19i2.54.

[21] M. A. Novianta, S. Syafrudin, and B. Warsito, "K-Means Clustering for Grouping Rivers in DIY based on Water Quality Parameters," *JUITA J. Inform.*, vol. 11, no. 1, p. 155, 2023, doi: 10.30595/juita.v11i1.16986.

[22] D. A. Prasetya, P. T. Nguyen, R. Faizullin, I. Iswanto, and E. F. Armay, "Resolving the shortest path problem using the haversine algorithm," *J. Crit. Rev.*, vol. 7, no. 1, pp. 62–64, 2020, doi: 10.22159/jcr.07.01.11.