JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 5, No. 3. June 2023 DOI: https://doi.org/10.34288/jri.v5i3.234
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

A Study on Enhanced Spatial Clustering Using Ensemble Dbscan and Umap to Map Fire Zone in Greater Jakarta, Indonesia

Silviya Hasana-1*), Devi Fitrianah-2

1*) Computer Science Program, School of Computer Science
2 Computer Science Department, Binus Graduate Program, Master of Computer Science
Bina Nusantara University
Jakarta, Indonesia
1*)silviya.hasana@binus.edu, 2devi.fitrianah@binus.edu
(*) Corresponding Author

Abstract

This research investigated ensemble clustering algorithms and dimensionality reduction for fire zone mapping, specifically DBSCAN + UMAP. We evaluated six clustering methods: DBSCAN, ensemble DBSCAN, DBSCAN + UMAP, ensemble DBSCAN + UMAP, HDBSCAN and Gaussian Mixture Model (GMM). We evaluated our results based on the Silhouette Score and the Davies-Bouldin (DB) index, emphasizing handling irregular cluster shapes, smaller clusters and resolving incompact clusters. Our findings suggested that ensemble DBSCAN + UMAP outperformed five other methods with zero noise clusters indicating clustering results are resistant to outliers, leading to a clearer identification of fire-prone areas, a high Silhouette Score of 0.971, indicating accurate cluster separation of distinct areas of potential fire hazards and an exceptionally low DB Index of 0.05 that indicates compact clusters to identify well-defined and geographically concentrated areas prone to fire hazards. Our findings contribute to the advanced techniques for minimizing the impacts of fires and improving fire hazard assessments in Indonesia.

Keywords: fires; clustering; ensemble, dbscan; umap

Abstrak

Riset ini berfokus pada investigasi algoritma pengelompokan ansambel dan reduksi dimensi untuk pemetaan zona api, khususnya menggunakan DBSCAN dan UMAP. Kami mengevaluasi enam metode pengelompokan: DBSCAN, DBSCAN ansambel, DBSCAN + UMAP, ansambel DBSCAN + UMAP, HDBSCAN, dan Model Campuran Gaussian (GMM). Penilaian didasarkan pada skor siluet dan indeks Davies-Bouldin (DB), dengan penekanan pada peningkatan kualitas klaster dalam mengatasi bentuk tidak beraturan dan klaster kecil dan tidak padat. Temuan kami menunjukkan ansambel DBSCAN + UMAP mengungguli lima metode pengelompokan lainnya dengan nol klaster outlier yang menunjukkan hasil pengelompokan tahan terhadap outlier, mengarah ke identifikasi yang jelas dari area rawan kebakaran, Skor Silhouette tinggi 0,971 menunjukkan pemisahan kluster yang akurat dari area berpotensi kebakaran dan Indeks DB yang sangat rendah sebesar 0,05 yang mengindikasikan kluster padat untuk mengidentifikasi area rentan kebakaran yang terdefinisi dengan baik dan terkonsentrasi secara geografis. Temuan kami berkontribusi pada teknik untuk meminimalkan dampak kebakaran dan meningkatkan penanggulangan bahaya kebakaran di Indonesia.

Kata kunci: kebakaran; pengelompokan; ansambel, dbscan; umap

INTRODUCTION

Fire hazards threaten both human lives and the environment (Abid, 2021; Courtwright, 2023), making good risk assessments and resource allocation critical for minimizing the impact of fires (Leiras et al., 2021). Indonesia presents unique problems with fire hazards due to its vast regions and diverse ecosystems. First, Indonesia is home to abundant peatland ecosystems. The Indonesian government estimates 149,056 km2 of peatlands on its three major islands: 28-32% in Kalimantan, 34%-43% in Sumatra, and 25%-38% in Papua (Harrison et al., 2020). Peatlands soil contains carbon-organic matter, which is prone to fire (Brasika, 2023).

Furthermore, the reaction between carbon and fire can smolder underground and reignite, making peatland fires difficult to extinguish (Lin et al., 2020). Next, Indonesia's unique geographical location makes it prone to El Niño Southern

Accredited rank 3 (SINTA 3), excerpts from the decision of the Minister of RISTEK-BRIN No. 200/M/KPT/2020

Oscillation (ENSO), which has been identified as a significant cause of fires. ENSO is a global phenomenon of periodic patterns of climate variability in the eastern Pacific Ocean (Nurdiati et al., 2022). An area influenced by ENSO suffers from periodic and extended droughts causing hotter and drier conditions, which are ideal for fire ignition and has increased burned areas in Indonesia (Nurdiati et al., 2022). Despite Indonesia's susceptibility to fire due to land composition and climate, certain areas are densely inhabited yet lack safe fire protection systems. Greater Jakarta, Indonesia's capital, is a prime example of an overpopulated city with many highrise buildings that still lack proper fire safety measures. Only 42% of Jakarta's highrise buildings have reliable fire protection systems, while 40% are less reliable and 18% are not (Rahardjo et al., 2020). This has led to an alarming 500 fire incidents annually in Jakarta from 2015 to 2020 (Rahardjo et al., 2020).

Current research focuses on investigating patterns of fire hazards by zone clustering. This direction provides effective fire assessments by detecting major clusters. Well-known clustering approaches have been used to investigate fire occurrences, each with its characteristics for finding hidden structures within datasets.

Murugesan et al. (Murugesan et al., 2021) and Artés et al. (Artés et al., 2019) investigated DBSCAN (Density-Based Spatial Clustering of Applications with Noise) to detect spatial clusters of fire occurrences. DBSCAN (Esther et al., 1996) is a density-based algorithm that separates high-density from low-density clusters based on data proximity. DBSCAN uses two tuning parameters: "*epsilon*'' and "*minPts*", that determine the radius and minimum number of points in a cluster. This approach effectively helps find high-density fire clusters for benchmarking (Murugesan et al., 2021) and monitoring systems (Artés et al., 2019).

Another approach involves Hierarchical Density-Based Spatial Clustering of Applications with Noise or commonly known as HDBSCAN (Campello et al., 2015), which is a hierarchical clustering as an extension of DBSCAN. Unlike DBSCAN, which is susceptible to noise, HDBSCAN is a high-density clustering that lessens noise by using two key parameters: "*min_cluster_size*" and "*min_samples*" as the smallest size of a group and the minimum number of neighbors to a core point. Huang et al. (Huang et al. 2022) conducted research to cluster fire distribution using HDBSCAN and obtained decent results in mapping the high-risk areas of fire hotspots to provide decision support for fire prevention. Research by Zerbe et al. (Zerbe et al., 2022) resonated with the same idea to utilize HDBSCAN for hotspot distribution to aid the development of natural hazard mitigation plans and risk reduction strategies in Washington, USA. HDBSCAN excels in detecting clusters of variable densities, allowing it to locate subclusters within larger fire-prone areas.

Gaussian Mixture Modeling (GMM) is another choice for fire hazard clustering. GMM is a probabilistic model that allows for thoroughly investigating fire data, revealing underlying patterns and assisting fire prediction. Júnior et al. (Júnior et al., 2020) proposed an automatic fire danger classification using GMM where each cluster corresponds to a different threat level for fire prevention and fire danger assessment.

Fire clustering research focuses on utilizing the best algorithm to improve our understanding of fire patterns, high-risk areas and effective fire risk assessments. Each clustering algorithm offers valuable strengths depending on the dataset characteristics. Fire hazards, for instance, are characterized by non-regular clusters, which DBSCAN, HDBSCAN and GMM are suitable to handle. However, GMM can run slowly and becomes highly expensive in terms of complexity. While both DBSCAN and HDBSCAN have their strengths, DBSCAN's simplicity and straightforward implementation are key strengths, making it a valuable tool for fire clustering. DBSCAN is set, which can be addressed with outlier removal and an effective dimension reduction method, namely UMAP, that offers better structure preservation compared to UMAP, which is a neighborhood-preserving reduction technique for constructing topological representations of high-dimensional data using manifold approximations. (McInnes et al., 2020). Well-chosen parameters tuning with a fitting dimension reduction strategy ultimately lead to DBSCAN as an effortless yet effective clustering tool for high dimensional data such as fire hazards.

We propose ensemble DBSCAN with UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) for fire-zone clustering. UMAP dimensionality reduction is utilized to address noise and increase mapping accuracy. We investigated features engineering before the ensemble. The ensemble technique will combine the results of numerous DBSCAN instances, each with distinct parameter settings. It allows us to obtain a thorough representation of fire clusters to enhance zone mapping for a more sophisticated knowledge of fire patterns.

RESEARCH METHODS

JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 5, No. 3. June 2023 DOI: https://doi.org/10.34288/jri.v5i3.234
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Following an experiment flow, we aim to map fire-prone zones using DBSCAN and UMAP, as shown in Fig 1. DBSCAN, with careful parameter selection, is used to identify dense regions representing fire clusters. Subsequently, UMAP is integrated to improve clustering by lowering dimensionality while maintaining data structure. We will measure cluster quality using metrics such as silhouette score and Davies-Bouldin (DB) index throughout numerous iterations of clustering by varying the parameters of DBSCAN and UMAP.

Methodology

We follow an experiment flow as shown in Figure 1. First is dataset collection, followed by preprocessing steps, including feature engineering, duplicates and outliers removal and normalization to ensure data quality. Next is feature selection, Consensus Function and results evaluation using the Silhouette and DB scores. Important steps are further explained below.

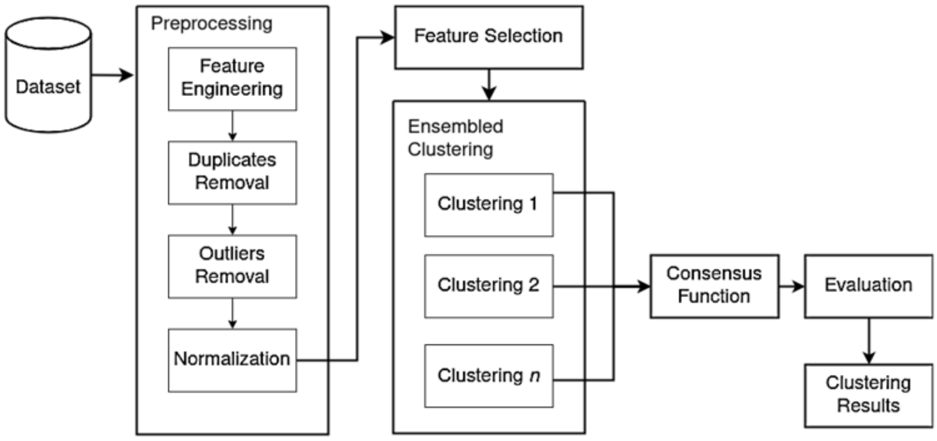

Figure 1. Experiment method

Dataset

Four datasets of reported fire hazards in Greater Jakarta from 2014 - 2017 were collected. The datasets were combined into a single dataset with 28 attributes and 5074 records. The list of attributes is seen in Table 1.

Table 1. Dataset of Fire Hazards in Greater Jakarta 2014-2017

No.	Feature Name	In English
1.	Tanggal	Date
2.	Status Kejadian	Event State
3.	Jenis Kejadian Bencana	Disaster Type
4.	Keterangan JKB	Number of Victims
5.	Sumber Informasi	Information Source
6.	Waktu	Time
7.	Jenis Waktu	Type of Time
8.	Alamat Kejadian	Address
9.	Kelurahan	Subdistrict
10.	Kecamatan	District
11.	Wilayah	Region
12.	Latitude	Latitude
13.	Longitude	Longitude
14.	Objek Kejadian Awal	Initial Event Object
15.	Kategori Objek Bencana	Disaster Object Category
16.	Unit Pertama Tiba	First Responder Arrival
17.	Waktu Respon	Response Time
18.	Menit Respon	Minutes to Response
19.	Sebab Kejadian	Cause of Fire
20.	Keterangan Sebab Kejadian	Cause Description
21.	Tanggal Selesai Kejadian	End Date
22.	Waktu Selesai	End Time
23.	Jenis Waktu Selesai	Type of End Time
24.	Luas Area Kejadian	Fire Range
25.	Taksiran Kerugian	Loss Estimation
26.	Kronologis Kejadian	Fire Chronology
27.	Keterangan Bencana	Disaster Description

Feature Engineering

A significant level of uncertainty and missing data in fire hazard datasets are common (Luo et al., 2020; St. Denis et al., 2020), making analysis and pattern clustering challenging. To address the same issue in our data, we attempted further processing, including features engineering, that has become crucial in fire hazards analysis.

The dataset separates the start and end dates and times of fire hazards. As a result, in addition to merging the start date and time of fire as *start_date*, we also merged the end date and time as *end_date*. We attempted to capture *fire_duration as* the interaction between *start_date* and *end_date*.

Given the sensitivity of fire hazards to time (e.g., a minute of fire would cause significant damage), we stored *fire duration* with minutes as the unit measure.

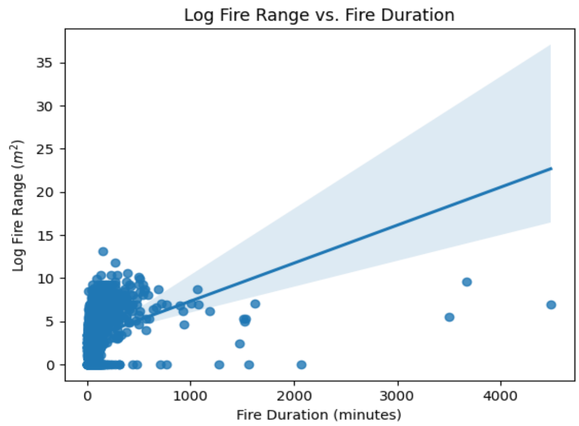

Figure 2. Scatter plot displaying a positive relationship between fire range and fire duration.

For subsequent processing, we encoded both *district* and *region*. We calculated *avg_fire_range by* dividing *fire_range* per *district.* Similarly, dividing *fire_intensity* by *district* provided us with *avg_fire_ints*. We chose the *district* feature because Indonesian *subdistrict* coverage could run too narrow, while *region* coverage could be too broad. The complete list of our newly engineered features is shown in Table 2.

The interplay from dividing the newly obtained *fire_range* by *fire_duration* gave us insight into the *fire intensity. ..*

Table 2. List of newly obtained features

No.	Feature Name	Obtained from
1.	fire_duration	start_date, end_date
2.	fire_range	regression model
3.	fire intensity	fire_range, fire_duration
4.	avg_fire_range	fire_range, district
5.	avg_fire_ints	fire_intensity, district

Feature Selection

Armed with 32 attributes, we continued removing low-variance features as they were less informative. We implemented a variance threshold method as our unsupervised feature selection to achieve this. Four best features achieved are: '*region*', '*district*', '*avg_fire_range*', '*avg_fire_ints*'. We proceeded with these features for clustering.

Ensemble Clustering

Fire hazards occur unevenly due to various external factors, such as temperature, wind speed, humidity, and precipitation (Liang et al., 2019). This renders fire hazards to have high dimensional characteristics, combined with outliers and unpredictable recurrence patterns, and we carefully considered irregular cluster shapes for fire hazards, leading us to choose DBSCAN. Ensemble clustering aims to deliver better results by combining multiple DBSCAN clustering models.

Before clustering, we removed duplicates and outliers that could negatively impact cluster results. We also considered a few dimensionality reduction techniques. PCA (Principal Component Analysis) is a common approach for dimensionality analysis, including studying the distribution of fire ignition areas (Arellano-del-Verbo et al., 2023). Knowing the presence of a linear positive relationship in our data, we attempted PCA, resulting in massive information loss in our dataset, including severe cluster loss. To preserve both global and local cluster structures that PCA might miss, we opted to implement UMAP.

Consensus Function

A single DBSCAN run will loop through the epsilon and min_samples parameters to find the best combination. Parameters tested were *epsilon*: 0.1, 0.2, 0.3, 0.4, 0.5 and *min_samples*: 10, 15, 20, 25. Subsequently, the consensus function operates on aggregating the results of individual clustering using different parameter combinations.

Evaluation

The goal is to assess the clustering quality, which is complicated since there are no universally accepted standards for 'good quality clusters'. We chose to consider three evaluations :

1. Visual Inspection.

   To assess if the clustering has resulted in good separation and to identify meaningful patterns in the data that the numerical metrics may not capture.

2. Silhouette score :

   It measures the cohesion-to-separation ratio, which ranges from -1 to 1, with the preferable metric maximization (near 1). A silhouette score (ss) formula for a data point r in cluster C_l is as shown in Equation 1 below (Mollaian, 2021) :

$$SS(r) = \frac{B-A}{max(AB)} \quad \text{....................} (1)$$

A, as seen in Equation 2, denotes the mean distance between a point and the other points within the same cluster.

$$A(r) = \frac{1}{|C_l|-1} \sum_{k \epsilon C_l, r \neq k} d(r,k) \quad \text{..........} (2)$$

JURNAL RISET INFORMATIKA **P-ISSN: 2656-1743 |E-ISSN: 2656-1735**
Vol. 5, No. 3. June 2023 DOI: https://doi.org/10.34288/jri.v5i3.234
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

And B, as seen in Equation 3, denotes the shortest mean distance between a point and the other points in the next closest cluster.

$$B(r) = min\ \frac{1}{|C_o|} \sum_{k \epsilon C_o} d(r,k) \quad (3)$$

Negative Silhouette scores indicate poor separation where points were assigned to incorrect clusters. Scores near zero indicate overlapping clusters. Scores near 1 are the best, indicating well-separated clusters, where each point well matches its cluster and poorly matches nearby clusters. To maximize the Silhouette score, A's value must be smaller than B's. While high scores indicate better clustering, they do not always align with visual inspection. Therefore, a combination with visual inspection is necessary.

3. Davies-Bouldin (DB) Index :

It calculates the distance between clusters relative to the size of the clusters to determine cluster compactness. T and M are computed. T is the mean distance between each cluster point and its centroid, as shown in Equation 4. (Mollaian, 2021).

$$T_l = \left(\frac{1}{|C_l|} \sum_{j=1}^{|C_l|} |x_j - e_l|^p\right)^{\frac{1}{p}} \quad (4)$$

where x_j is the cluster's feature vector, e_l is the cluster centroid, |Cl| is the cluster size, and p is generally 2, considering a Euclidean distance. M denotes the distance between cluster centroids, as seen in Equation 5:

$$M_{l,o} = \|e_l - e_o\|_p \quad (5)$$

The R_{ij} value is then determined for each pair of clusters.

$$R_{lo} = \frac{T_l + T_o}{M_{l,o}} \quad (6)$$

The maximum of R_{lo} from Equation 6 serves to calculate the Davies-Bouldin (DB) index as seen in Equation 7 below:

$$DB = \frac{1}{k} \sum_{l=1}^{k} max\ R_{lo} \quad (7)$$

Ideally, a high M value and low T value are desirable to obtain a low DB index, while index minimization near 0 is preferred to represent well-separated clusters.

RESULTS AND DISCUSSION

This research aimed to study the performance of the ensemble DBSCAN technique. Therefore, our first comparison is against various models of DBSCAN. We tested four clustering methods: DBSCAN, ensemble DBSCAN, DBSCAN + UMAP, and ensemble DBSCAN + UMAP. Table 3 shows our comparison results. Silhouette score evaluation focuses on the fit of individual data points to their clusters and the separation between fire zone clusters. At the same time, the DB index considers the overall compactness of fire zones.

Table 3. Silhouette score and DB index evaluation score for four clustering methods

No.	Method	Silhouette Score	DB index	Noise cluster
1.	DBSCAN	0.989	0.83	62
2.	Ensemble DBSCAN	0.553	0.25	33
3.	DBSCAN + UMAP	0.975	0.62	0
4.	Ensemble DBSCAN + UMAP	0.971	0.05	0

From the results, a single run of DBSCAN with the best parameter combination (*epsilon*: 0.1, *min_samples*: 10) achieved a high Silhouette Score of 0.989, indicating good fire zone separation, but obtained a DB Index of 0.83, suggesting sparse density in overall fire zones. Table 4 shows the results of the best single DBSCAN run with 46 fire zones and each data point count. As we can see from the table, this method generated a relatively large number of noise clusters (33). Figure 3 depicts the map visualization, and we can see that some dark blue clusters overlap with a purple cluster.

Table 4. List of clusters and each data point counts from mapping using DBSCAN

Cluster Label	Datapoint counts	Cluster Label	Datapoint counts
-1	33	22	166
0	190	23	106
1	92	24	160
2	75	25	44
3	66	26	106
4	133	27	178
5	84	28	104
6	120	29	119
7	149	30	59
8	132	31	75
9	99	32	112
10	126	33	74
11	87	34	131
12	121	35	50
13	114	36	17

Cluster Label	Datapoint counts	Cluster Label	Datapoint counts
14	75	37	12
15	163	38	137
16	82	39	145
17	55	40	84
18	180	41	74
19	208	42	155
20	218	43	58
21	85	44	104
			Total cluster: 46

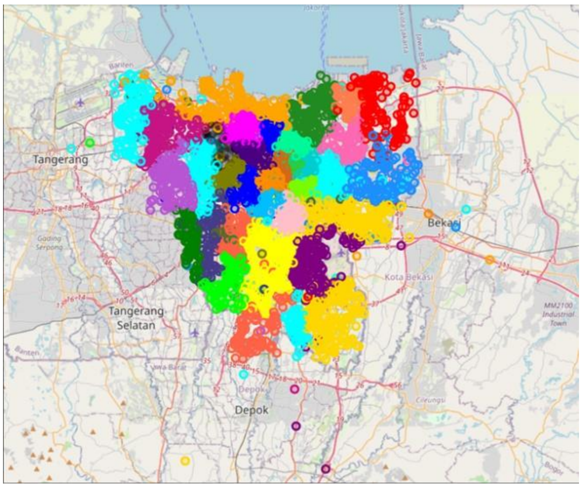

Figure 3. Fire zone map of Greater Jakarta Using DBSCAN.

The Ensemble DBSCAN approach showed a lower Silhouette Score of 0.553, suggesting moderate fire zone separation with an even higher number of 62 noise clusters. However, it achieved a significantly lower DB Index of 0.25, suggesting improved fire zone compactness by utilizing the ensemble approach. The best run of single DBSCAN + UMAP (*epsilon*: 0.2, *min_samples*: 15) achieved a high Silhouette Score of 0.975 and a moderate DB Index of 0.62 while successfully eliminating all noise clusters, indicating UMAP implementation is effective in handling outliers. Finally, the ensemble DBSCAN + UMAP outperformed all methods with zero noise clusters, a high Silhouette Score of 0.971 indicating excellent fire zone mapping separation and an exceptionally low DB Index of 0.05, near zero, indicating highly compact and well-defined fire zones, making ensemble DBSCAN + UMAP the preferable choice for clustering complex, high dimensional fire hazard datasets. Table 5 shows the cluster results with decently distributed data points from ensemble DBSCAN + UMAP. Figure 4 shows its map visualization.

Table 5. List of clusters and each data point counts from mapping using ensemble DBSCAN + UMAP

Cluster Label	Datapoint counts	Cluster Label	Datapoint counts
0	190	22	85
1	98	23	166
2	75	24	106
3	66	25	160
4	143	26	44
5	84	27	106
6	120	28	178
7	149	29	104
8	132	30	119
9	99	31	59
10	126	32	75
11	87	33	112
12	125	34	74
13	114	35	131
14	75	36	50
15	163	37	21
16	86	38	149
17	55	39	145
18	109	40	84
19	180	41	74
20	208	42	155
21	218	43	58
			Total cluster: 44

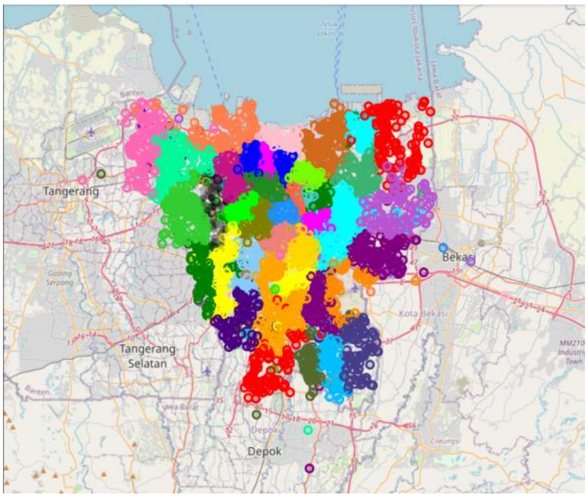

Figure 4. Fire zone map of Greater Jakarta using ensemble DBSCAN + UMAP.

There is a clear trade-off between fire zone separation (represented by Silhouette Score) and fire zone compactness (represented by DB index). It has been demonstrated that combining the ensemble technique with UMAP can effectively enhance DBSCAN clustering to improve fire zone separation, compactness and eliminate outliers. While the main study is to assess the performance of ensemble DBSCAN + UMAP, we evaluated our approach against other well-known clustering methods, such as HDBSCAN and GMM. Table 6 shows the evaluation results of ensemble DBSCAN + UMAP, HDBSCAN and GMM.

Table 6. Silhouette score and DB index evaluation score for ensemble DBSCAN + UMAP, HDBSCAN and GMM

No.	Method	Silhouette Score	DB index	Noise cluster
1.	Ensemble DBSCAN +	0.971	0.05	0
3.	UMAP	0.990	2.41	38
4.	HDBSCAN GMM	0.931	0.25	0

Among the three compared clustering methods, HDBSCAN delivered the highest Silhouette Score of 0.990, implying well-separated and distinct fire zones. However, the DB Index 2.4 raises concerns about potential overlapping or poorly defined clusters. The presence of 38 noise clusters further indicates a less ideal fire zone formation. On the other hand, GMM obtained a respectable Silhouette Score of 0.931, signifying decent fire zone separation, while its DB Index of 0.25 suggests some overlapping while successfully eliminating all noise clusters. Ensemble DBSCAN + UMAP showcased a competitive Silhouette Score of 0.971, denoting well-separated fire zones, and exhibited the most favorable DB Index of 0.05, indicating highly compact and well-defined zone mapping.

Additionally, the absence of noise clusters further strengthens its superiority in achieving more cohesive and interpretable fire zone mapping. While HDBSCAN performed well in Silhouette Score, noise clusters and a high DB Index raise concerns about its cluster quality. GMM performed reasonably well, but its DB Index indicates some overlap. In contrast, ensemble DBSCAN + UMAP demonstrated superior performance with the lowest DB index and the absence of noise clusters indicating geographically concentrated areas prone to a fire hazard that is clearer and well-separated, making it the most favorable choice for fire zone mapping.

CONCLUSIONS AND SUGGESTIONS

Conclusion

In this paper, we studied and analyzed ensemble clustering algorithms using DBSCAN + UMAP against five other methods: traditional DBSCAN, ensemble DBSCAN, DBSCAN + UMAP, HDBSCAN, and GMM. Our findings suggest that ensemble DBSCAN + UMAP outperformed five other clustering methods with zero noise clusters indicating clustering results are resistant to outliers, leading to a clearer identification of fire-prone areas, a high Silhouette Score of 0.971, indicating accurate cluster separation of distinct areas of potential fire hazards and an exceptionally low DB Index of 0.05 that indicates compact clusters to identify well-defined and geographically concentrated areas prone to fire hazards.

In the fire hazard dataset context, a clustering method with superior performance, like ensemble DBSCAN + UMAP, becomes particularly important due to its specific advantages. A lower DB index implies that the clusters are well-defined and compact, implying that they accurately represent distinct areas of potential fire hazards. The absence of noise clusters indicates that the clustering is resistant to outliers, resulting in the accurate identification of fire-prone locations. Additionally, a high Silhouette Score shows that the clusters are cohesive and well-separated, which supports the veracity of the detected fire-prone zones. Ensemble DBSCAN + UMAP makes a good choice for fire hazard datasets due to its accuracy and reliable clustering results, allowing for better identification and proactive management of potential fire-prone areas, which is critical for effective fire prevention and control strategies.

Valuable resources can be utilized efficiently by identifying and focusing on accurately distinct and compact fire-prone zones. It enables efficient resource allocation and targeted planning for fire response, increasing the effectiveness of fire mitigation efforts. Additionally, the cohesive and well-separated clusters provided by a low DB index and high Silhouette Score enhance the understanding of fire hazard patterns, enabling authorities to devise proactive measures and implement preventative measures to identify vulnerable areas. This approach facilitates better-informed decision-making, leading to more effective fire hazard management and preserving lives and properties.

Suggestions

Ensemble DBSCAN + UMAP is applicable to properly cluster most high-dimensional hazard datasets with significant information uncertainty, such as fire hazards. However, it is critical to acknowledge that the results may be context-dependent and not universally relevant to all scenarios. The experiments might have been limited to a specific geographical location or dataset characteristics, limiting the method's generalization. Further study is necessary to confirm this.

While the clustering analysis yielded valuable insights into identifying fire-prone zones, the conclusions do not explicitly provide practical recommendations or policy implications on fire hazards. Therefore, meaningful recommendations based on fire zone mapping results are necessary. For example, authorities and firefighting teams could use the fire zone mapping results to prioritize resource allocation for fire response. In addition, land planning and zoning policies should be adapted to account for high-risk locations, imposing tougher construction codes and fire safety measures. Furthermore, fire hazard management techniques such as increased patrols and public awareness campaigns could be targeted to areas with the highest fire risk. By incorporating practical recommendations based on the fire zone mapping analysis, this research can provide valuable assistance to policymakers in devising effective measures to prevent and mitigate fire hazards in vulnerable areas, ultimately improving public safety and minimizing fire-related damages.

REFERENCES

Abid, F. (2021). A Survey of Machine Learning Algorithms Based Forest Fires Prediction and Detection Systems. *Fire Technology*, *57*(2), 559–590. https:// doi.org/10.1007/s10694-020-01056-z

Courtwright, J. (2023). Prairie Fire: A Great Plains History. University Press of Kansas.

Leiras, A. B., Rodrigues, J. P. C., & Meacham, B. J. (2021). A performance-based fire risk analysis for buildings. Architecture, Structures and Construction, *1*(2), 143–175.https://doi.org/10.1007/s44150-021-00016-7

Harrison, M. E., Ottay, J. B., D'Arcy, L. J., Cheyne, S. M., Anggodo, Belcher, C., Cole, L., Dohong, A., Ermiasi, Y., Feldpausch, T., Gallego-Sala, A., Gunawan, A., Höing, A., Husson, S. J., Kulu, I. P., Soebagio, S. M., Mang, S., Mercado, L., Morrogh-Bernard, H. C., … van Veen, F. J. F. (2020). Tropical forest and peatland conservation in Indonesia: Challenges and directions. *People and Nature*, *2*(1), 4–28. https://doi.org/10.1002/pan3.10060

Brasika, I. B. M. (2023). Forest Fire Emissions in Equatorial Asia and Their Recent Delay Anomaly in the Dry Season. In K. P. Vadrevu, T. Ohara, & C. Justice (Eds.), *Vegetation Fires and Pollution in Asia* (pp. 447–462). Springer International Publishing.https://doi.org/10.1007/978-3-031-29916-2_26

Lin, S., Cheung, Y. K., Xiao, Y., & Huang, X. (2020). Can Rain Suppress Smoldering Peat Fire? *Science of The Total Environment*, *727*, 138468. https://doi.org/10.1016/j.scitotenv.2020.138468

Nurdiati, S., Bukhari, F., Julianto, M. T., Sopaheluwakan, A., Aprilia, M., Fajar, I., Septiawan, P., & Najib, M. K. (2022). The impact of El Niño Southern Oscillation and Indian Ocean Dipole on the Burned Area in Indonesia. *Terrestrial, Atmospheric and Oceanic Sciences*, *33*(1), 16.https://doi.org/10.1007/s44195-022-00016-0

Rahardjo, H. A., & Prihanton, M. (2020). The Most Critical Issues and Challenges of Fire Safety for Building Sustainability in Jakarta. *Journal of Building Engineering*, *29*, 101133. https://doi.org/10.1016/ j.jobe.2019.101133

Murugesan, N., Cho, I., & Tortora, C. (2021). Benchmarking in Cluster Analysis: A Study on Spectral Clustering, DBSCAN, and K-Means. In T. Chadjipadelis, B. Lausen, A. Markos, T. R. Lee, A. Montanari, & R. Nugent (Eds.), *Data Analysis and Rationality in a Complex World*. Springer International Publishing. https://doi.org/10.1007/ 978-3-030-60104-1_20

Artés, T., Oom, D., de Rigo, D., Durrant, T. H., Maianti, P., Libertà, G., & San-Miguel-Ayanz, J. (2019). A Global Wildfire Dataset for the Analysis of Fire Regimes and Fire Behaviour. *Scientific Data*, *6*(1), Article 1. https://doi.org/ 10.1038/s41597-019-0312-2

Ester, M., Kriegel, H., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Knowledge Discovery and Data Mining (Vol. 96, pp. 226–231).

Campello, R. J., Moulavi, D., Zimek, A., & Sander, J. (2015). Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *10*(1), 1-51.

Huang, J., Xu, Z., Yang, F., Zhang, W., Cai, S., Luo, J., Xie, G., & Li, T. (2022). Fire Risk Assessment and Warning Based on Hierarchical Density-Based Spatial Clustering Algorithm and Grey Relational Analysis. *Mathematical Problems in Engineering*, *2022*. https:// doi.org/10.1155/2022/7339312

Zerbe, K., Polit, C., McClain, S., & Cook, T. (2022). Optimized Hot Spot and Directional Distribution Analyses Characterize the Spatiotemporal Variation of Large Wildfires in

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v5i3.234
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Washington, USA, 1970–2020. *International Journal of Disaster Risk Science*, *13*(1), 139–150. https:// doi.org/10.1007/s13753-022-00396-4

Júnior, J. S. S., Paulo, J. R., Mendes, J., Alves, D., Ribeiro, L. M., & Viegas, C. (2022). Automatic Forest Fire Danger Rating Calibration: Exploring clustering techniques for regionally customizable fire danger classification. Expert Systems with Applications, 193, 116380. https://doi.org/10.1016/j.eswa.2021.116380

McInnes, L., Healy, J., & Melville, J. (2020). (2018). UMAP: Uniform Manifold Approximation and Projection. Journal of Open Source Software, 3(29), 861, https://doi.org/10.21105/joss.00861

Luo, J., Han, Y., Zhao, Y., Huang, Y., Liu, X., Tao, S., Liu, J., Huang, T., Wang, L., Chen, K., & Ma, J. (2020). Effect of northern boreal forest fires on PAH fluctuations across the arctic. *Environmental Pollution*, *261*, 114186.https://doi.org/10.1016/j.envpol.2020.114186

St. Denis, L. A., Mietkiewicz, N. P., Short, K. C., Buckland, M., & Balch, J. K. (2020). All-hazards dataset mined from the US National Incident Management System 1999–2014. *Scientific data*, *7*(1), 64.

Liang, H., Zhang, M., & Wang, H. (2019). A Neural Network Model for Wildfire Scale Prediction Using Meteorological Factors. *IEEE Access*, *7*, 176746-176755.

Arellano-del-Verbo, G., Urbieta, I. R., & Moreno, J. M. (2023). Large-Fire Ignitions Are Higher in Protected Areas than Outside Them in West-Central Spain. Fire, 6(1), 28.

Mollaian, M. (2021). Application of Dimension Reduction and Clustering Methods for Detection of Faulty Operations in Process Systems. University of California, Davis.

P-ISSN: 2656-1743 | E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v5i3.234

Accredited rank 3 (SINTA 3), excerpts from the decision of the Minister of RISTEK-BRIN No. 200/M/KPT/2020

This sheet is intentionally left blank