The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 2, July 2025, Page 60-71
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i2.8900

Artificial Intelligence Analysis of Recommendations for Granting Business Licenses to Determine the Priority of Business Supervision and Control Using the DBSCAN Method (Case Study: DPMPTSP Langkat Regency)

Suhardiansyah*, Zulham Sitorus, Muhammad Iqbal

Postgraduate, Master of Information Technology, Panca Budi University of Development , Medan, Indonesia
Email: [1*]suhardiansyah16@gmail.com, [2]zulhamsitorus@dosen.pancabudi.ac.id, [3]muhammadiqbal@dosen.pancabudi.ac.id
(*:suhardiansyah16@gmail.com)
Submitted: **08/07/2025**; Accepted: **20/07/2024**; Published: **23/07/2025**

Abstract
In facing the challenges of limited resources and business complexity, the Investment and One-Stop Integrated Services Office (DPMPTSP) of Langkat Regency requires a data-driven approach to determine priorities for business supervision and enforcement. This study applies the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm to cluster business entities based on three main psarameters: risk level, business scale, and licensing status. Secondary data from 3,748 companies were collected, processed through label encoding and normalization, and analyzed in a three-dimensional space (X1_Risk, X2_Scale, X3_License). The clustering results revealed the formation of clusters and a Silhouette Score value, indicating optimal cluster structure and separation between groups. Each cluster was interpreted as a representation of recommendation categories such as Routine Monitoring and Evaluation, Intensive Monitoring and Evaluation, Administrative Warning, Temporary Operational Suspension, and Permanent Operational Termination. The resulting visualizations enhanced the understanding of spatial mapping and clustering patterns comprehensively. This demonstrates that DBSCAN is effective as a decision-support tool for automated and objective priority mapping in business supervision, and capable of detecting business entities that deviate from general norms (outliers). This approach significantly contributes to improving the efficiency and accuracy of decision-making in business license supervision and enforcement at the regional level.

Keywords:DBSCAN, Business Supervision, Clustering, Data Mining, Licensing

1. INTRODUCTION

In the digital era, the use of information and communication technology has become an integral part of various sectors, including in the business licensing process. The Government of Indonesia encourages the digitization of public services to improve efficiency and ease of doing business through *the Online Single Submission* (OSS) system. However, in practice, the management of business licensing still faces various obstacles, such as limited resources in carrying out supervision and control. The increasing complexity of businesses, both in terms of type, risk category, and business location, makes the supervisory process more difficult and requires a more sophisticated data-driven approach. Various previous studies have shown that the use of artificial intelligence technology (*Artificial Intelligence)* and data analysis can be a solution in increasing the effectiveness of business supervision [1], [2]. AI is able to provide data-driven recommendations to set surveillance priorities. Research by Fitriana et al. (2021) added that the right analytical method can help in grouping businesses based on certain characteristics, so that the supervision process can be more directed and efficient [3], [4].

Data mining is the process of extracting meaningful information or patterns from large data sets using statistics, mathematics, and artificial intelligence techniques [5], [6], [7], [8], [9], [10], [11]. The main goal is to find hidden patterns, relationships, trends, or anomalies in the data that can be used for more effective decision-making. Data mining is often used in a variety of fields, such as business, finance, healthcare, and government, to support predictive analysis, segmentation, fraud detection, and data-driven policymaking [12], [13], [14], [15], [16]. One of the relevant algorithms in this context is DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*). DBSCAN is a density-based clustering method that is effective in handling irregular data and is able to detect *noise* or outlying data [17], [18]. The DBSCAN method has the advantage that it does not require determining the number of clusters at the beginning and is able to identify hidden patterns in large-scale datasets. This advantage makes DBSCAN very suitable for analyzing business licensing data which generally has uneven distribution characteristics and high complexity. DBSCAN can form clusters naturally based on data density, so that business groupings become more representative and relevant to support the supervisory function.

Various studies have shown the effectiveness of the DBSCAN algorithm in complex spatial data segmentation. Research by Permatasari et al. (2019) applied DBSCAN for segmentation of COVID-19 cases in West Java with two parameter scenarios, where the configuration of Eps = 0.5 and MinPts = 15 resulted in 7 clusters with higher prediction accuracy than the second configuration despite having *a lower* silhouette value[19]. The study by Risman et al. (2023) grouped districts/cities in Kalimantan based on people's welfare indicators, produced 5 optimal clusters with a DBI of 0.605, and was able to identify 33 areas as outliers with extreme characteristics [20]. Meanwhile, Utami et al. (2023) showed that DBSCAN was able to detect the concentration of forest and land fire hotspots very well, with the parameters Eps = 1.0 and MinPts = 5 resulting in a *silhouette* value of 0.876 reflecting a strong cluster structure. Overall, DBSCAN

Vol 9 No 2, July 2025, Page 60-71
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i2.8900

has proven to be reliable in identifying spatial patterns and effectively separating outlier data across different application domains [21].

Based on this background, this study aims to analyze business licensing data using the DBSCAN algorithm to provide recommendations for business supervision and control priorities. This study was conducted at the Investment and Integrated One-Stop Service Agency (DPMPTSP) of Langkat District, which faces challenges in determining supervision priorities amid rapid business growth. It is hoped that the results of this study will assist the DPMPTSP in optimizing resource utilization and creating a more conducive business and investment climate in the region.

2. RESEARCH METHODOLOGY

2.1 Research Stages

The proposed research stages can be seen in the following figure 1:

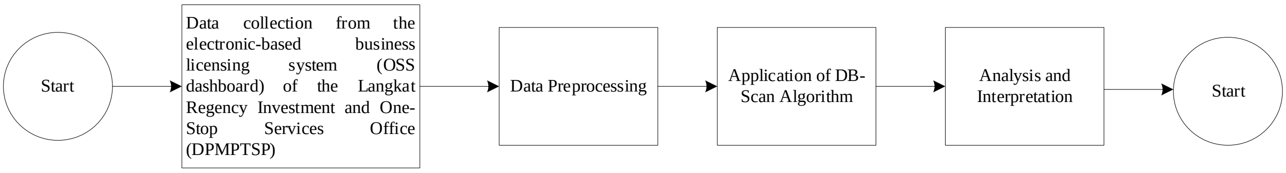

Figure 1. Stages of Research

1. Start the Research Process
   The research began with planning the steps to be taken. At this stage, the researcher determines the purpose of the research, develops a research design, and prepares the resources needed. This research focused on analyzing business licensing data available in Langkat district to classify supervision and control recommendations.
2. Data Collection
   At this stage, data was collected in accordance with the research needs. The data used is secondary data which includes information related to Limited Liability Companies (PT) and *Commanditaire Vennootschap* (CV) obtained from the Investment Office and One-Stop Integrated Licensing Services of Langkat Regency, with an observation period from January to December 2024.
3. Data Preprocessing
   After the data is collected, a preprocessing stage is carried out to clean the data from errors or irrelevant information, such as duplicates, missing characters, or elements that are not needed. This process also includes data normalization, such as conversion to lowercase, removal of punctuation marks, as well as data formatting into a form that is ready for analysis.
4. DBSCAN Algorithm Implementation
   The next stage is the application of the DBSCAN algorithm. This algorithm is used to group priority data for business supervision and control, the DBSCAN algorithm works based on density and does not require determining the number of clusters at the beginning.
5. Analysis and Interpretation of Results
   After the algorithm implementation is complete, the classification results are analyzed and interpreted. At this stage, the researcher checks the accuracy of the algorithm, identifies trends or patterns in the clustering, and understands the data model used. The results of this analysis aim to provide in-depth insights into providing recommendations for business supervision and control.
6. Completion and Documentation
   The final stage is the completion of the analysis process. All results obtained during the research are systematically documented and compiled in the form of a research report that includes all steps, analysis, and interpretation of results. At this stage, the researcher ensures that the research objectives have been achieved in accordance with the initial plan. The results of the analysis are not only the basis for providing recommendations for business supervision and control, but also as an evaluation strategy in the implementation of monitoring and evaluation at the investment office and one-stop integrated licensing services in Langkat district, so as to optimize limited resources. This information can be utilized by related parties, such as technical agencies, ministries/institutions in the implementation of monitoring and evaluation of business licensing in Langkat District.

2.2 Research Data

Vol 9 No 2, July 2025, Page 60-71
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i2.8900

Table 1. Langkat Regency Business Licensing Data Year 2024

No	Nib	Company name	KBLI	Description of project risk	Description of business scale	License status
1	1101240054533	Harapan makmur lestari	46201	Low	Micro business	Already have a license
2	1101240054533	Harapan makmur lestari	46202	Low	Micro business	Already have a license
3	1101240054533	Harapan makmur lestari	46314	Low	Micro business	Already have a license
4	1101240054533	Harapan makmur lestari	46693	Low	Micro business	Already have a license
5	1201240064239	Kaysan Construction	46100	Low	Small business	No license yet
6	1201240064239	Kaysan Construction	46319	Low	Small business	No license yet
7	1201240064239	Kaysan Construction	46412	Low	Small business	No license yet
8	1201240064239	Kaysan Construction	46421	Low	Small business	No license yet
9	1201240064239	Kaysan Construction	46422	Low	Small business	No license yet
10	1201240064239	Kaysan Construction	41011	Medium High	Small business	Not yet licensed
...	...	...	...	..	..	...
3746	2511240077199	JAKIDA USIBA TAMA	47845	Low	Micro Business	Already have a license
3747	2511240077199	JAKIDA USIBA TAMA	47845	Low	Micro Business	Already have a license
3748	2711240012228	HASYA SEHAT SEJAHTERA	47721	High	Micro Business	Already have a license

2.3 Business Scale

Business scale is a grouping of business activities based on their size or capacity, determined by certain criteria such as the number of workers, capital (assets), and/or annual turnover. In Indonesia, this classification is legally regulated and used in various policies, including risk-based business licensing (OSS RBA), taxation, business assistance, and public procurement.

Table 2. Business Scale Categories in Indonesia (based on PP No. 7 of 2021)

Category	Assets (excluding land & building)	Turnover per year
Micro Enterprises	≤ IDR 1 billion	≤ Rp 2 billion
Small Business	> Rp 1 billion - Rp 5 billion	> Rp 2 billion - Rp 15 billion
Medium Enterprises	> Rp 5 billion - Rp 10 billion	> Rp 15 billion - Rp 50 billion
Large Business	> Rp 10 billion	> Rp 50 billion

2.4 Project Business Risk (Business Risk)

Based on the provisions of Government Regulation of the Republic of Indonesia Number 5 of 2021 concerning the Implementation of Risk-Based Business Licensing, which is a derivative regulation of Law Number 11 of 2020 concerning Job Creation, every business activity in Indonesia is categorized based on the level of risk inherent in its business field. This classification uses the KBLI (Indonesian Standard Business Field Classification) reference which becomes the formal identity of the type of business activity. The risk level classification aims to adjust the level of supervision and licensing requirements to the level of danger and impact caused by business activities on society, the environment, and the economy. The classification is divided into four main categories as follows:

Vol 9 No 2, July 2025, Page 60-71
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i2.8900

1. Low Risk
   Business activities included in this category are only required to have a Business Identification Number (NIB) as a single form of licensing. Business actors can immediately start activities after obtaining an NIB without the need to fulfill additional requirements.
2. Medium Low Risk
   In addition to the NIB, business actors are also required to have a Standard Certificate issued through a *self-declared* mechanism, meaning that the business actor declares that his/her business has met the stipulated technical standards.
3. Medium High Risk
   Businesses in this category must have an NIB and Standard Certificate, but the issuance must be preceded by the fulfillment of technical requirements documents and through a verification process by the authorized technical Ministry or Institution (K/L).This process ensures that businesses that have potential impacts are carrying out practices in accordance with the relevant sector provisions.
4. High Risk
   For this category, businesses are required to have an NIB and License obtained after fulfilling the required documents and undergoing strict verification by the relevant technical agencies. This is because high-risk business activities have the potential to cause major impacts on health, safety, the environment, and public order.

2.5 Permit Status

As previously described, each business activity classified in the Indonesian Standard Industrial Classification (KBLI) has a different level of risk and type of license, depending on the sectoral authority of the ministry or technical institution (K/L) that regulates it. In the Risk-Based Business Licensing system, business actors are required to fulfill documents and technical requirements tailored to the risk level of each KBLI they run. The legality status of a business, in this case referred to as license status, is basically determined by the extent to which business actors have completed and fulfilled the administrative and technical requirements of each of their business fields. If all documents and technical standards have been fulfilled in accordance with the provisions, then the status is categorized as "existing" or "licensed". However, in practice in the field, it is not uncommon to find companies applying for licenses with more than one KBLI but only fulfilling some of the required requirements. This results in some KBLI listed in the business legal documents not fulfilling the technical documents as required. As a result, even though the company has obtained a Business Identification Number (NIB), some business fields in it still have the status of "license not yet issued" or even considered substantively "non-existent". This discrepancy can occur due to various factors, such as business actors' ignorance of sectoral licensing obligations, lack of socialization from technical agencies, or the unpreparedness of the online verification system in harmonizing data between agencies. Therefore, license status becomes one of the important indicators in determining the level of compliance of business actors and the priority of supervision that needs to be carried out by local governments.

2.6 Recommendation Criteria

This research is designed to produce five categories of supervision recommendations that are adaptive and data-driven, to support the effectiveness of decision-making by the One-Stop Investment and Integrated Services Agency (DPMPTSP). Through a clustering approach using the DBSCAN algorithm, each business entity is classified based on similar characteristics of risk level, business scale, and permit status.

The results of this analytic process will group companies into five categories of recommended actions, which can be the basis for developing a more proportional supervision and control strategy. The five recommendations include:

1. Routine Monitoring and Evaluation
   Aimed at companies that are considered to have low risk and have fulfilled their licensing obligations, as well as showing a high level of compliance. Supervision activities are carried out regularly in order to maintain compliance stability.
2. Intensive Monitoring and Evaluation
   Given to companies that have a high level of risk, but have fulfilled licensing requirements. This category requires stricter supervision and a higher frequency of monitoring, given the potential impact of the business on the environment or community.
3. Administrative Warning
   Given to companies that have not fulfilled their licensing obligations, despite being in the low or medium-low risk category. This action is persuasive in the form of a warning letter or administrative coaching.
4. Temporary Suspension of Operational Activities
   Aimed at medium-high risk companies that do not have a license, and have not fulfilled the stipulated technical requirements documents. This recommendation is made as a rule enforcement effort to encourage license fulfillment before business activities resume.
5. Permanent Suspension of Operational Activities

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 2, July 2025, Page 60-71
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i2.8900

Given to companies with[high risk categories that do not have licenses and are considered to endanger public safety, the environment, or cause significant losses. This action is carried out as the last step in policing.

The five recommendation criteria can be seen in Table 3 below:

Table 3. Recommendation Criteria

Description of Business Scale	Risk Description	License Status	Recommendation
micro	low	available	Regular monitoring and evaluation
micro	Medium low	available	Routine monitoring and evaluation
micro	Medium high	available	Routine monitoring and evaluation
micro	high	available	Intensive monitoring and evaluation routine
micro	low	None	Administrative warning
micro	Medium low	None	Administrative warning
micro	Medium high	None	Temporary suspension of activities
micro	high	None	Permanent suspension of activities
small	low	Yes	Regular monitoring and evaluation
small	Medium low	available	Routine monitoring and evaluation
small	Medium high	available	Routine monitoring and evaluation
small	high	available	Intensive monitoring and evaluation routine
small	low	None	Administrative warning
small	Medium low	None	Administrative warning
small	Medium high	None	Temporary suspension of activities
small	high	None	Permanent suspension of activities
medium	low	available	Regular monitoring and evaluation
medium	Medium low	available	Routine monitoring and evaluation
medium	Medium high	available	Routine monitoring and evaluation
medium	high	available	Intensive monitoring and evaluation routine
medium	low	None	Administrative warning
Medium	Medium low	None	Administrative warning
medium	Medium high	None	Temporary suspension of activities
medium	high	None	Permanent suspension of activities
large	low	available	Regular monitoring and evaluation
large	Medium low	available	Routine monitoring and evaluation

2.7 DBScan (*Density-Based Spatial Clustering With Noise*) Algorithm

DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) is a density-based clustering algorithm developed by Ester, Kriegel, Sander, and Xu in 1996. It groups data based on local density, by identifying areas with high data concentration as *clusters*, and areas with low density as *noise or outliers*. Unlike other methods, DBSCAN does not require determining the number of clusters at the beginning, so it is able to recognize clusters of varying shapes and sizes flexibly and efficiently. The DB-SCAN algorithm can find high-density core samples and expand clusters from them. There are two main parameters of the algorithm that determine the *cluster* minimum sample size and ε. *Core Point* is a point with ≥ min_samples or MinPts (including the point itself) within the radius eps (ε) and *Density Reachable* is a point within the range eps (ε) of a *core point* will be included in the same cluster. *Border Point* is a *non-core* point (its neighbors are less than *min_samples* (MinPts)) but within the radius eps (ε) of a *core point*, so it is included in the cluster. The first parameter defines shared as a core sample. The stages of applying the DBSCAN algorithm in this study are:

1. Data collection
   The process of collecting data using both primary and secondary data and determining important variables to be processed in data processing.
2. *Data preprocessing*

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 2, July 2025, Page 60-71
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i2.8900

At this stage, *data cleaning*, *encoding* and data standardization (normalization) are carried out so that the data is suitable for the next process, the process includes:

a. Cleaning, the process of removing duplicate data and empty data in important columns and adjusting the text (e.g. *lowercase*, remove extra spaces).
b. Categorical Data Encoding, To be used in mathematical calculations (Euclidean distance), categorical values are converted to numerical form.
c. Data *Standardization* (*Z-Score Normalization*), Since the DBSCAN distance calculation is scale sensitive, the numerical features are standardized to have mean = 0 and standard deviation = 1.

Z-Score Formula

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

With

Z : *Z-Score* value (standardized value)
X : Original value of a variable
μ : Average (mean) of all values in the variable
σ : Standard deviation of the variable

3. Determination of Data Components
The component uses three main dimensions as input variables for the clustering process:
a. x1 - Risk : Describes the level of potential hazard or surveillance liability.
b. x2 - Business Scale : Indicates how large the scope of business activities is.
c. x3 - License Status : Represents the level of compliance with applicable regulations or laws.
4. DBSCAN Parameters
The two main parameters used in DBSCAN are:
a. Epsilon (eps) : The maximum radius between two points to be considered neighbors, tested several values from 0.5 to 1.0.
b. Minimum Samples : The minimum number of points within the eps radius for a point to be considered a *core point*, determined as 4 points.
5. *Distance Calculation*
Using the *3D Euclidean* formula to calculate the distance between points in three-dimensional space:

$$Jarak\ (A, B) = \sqrt{(X_{1A} - X_{1B})^2 + (X_{2A} - X_{2B})^2 + (X_{3A} - X_{3B})^2} \quad (2)$$

If the distance between two points < eps, and the point has ≥ *min_samples* neighbors, then the point is considered *density-reachable.*
6. *Silhouette Coefficient and score*
The Silhouette Score is the average of the *Silhouette Coefficient of* all data points in a clusterization result. This value is used to measure the overall quality of the cluster.
Silhouette Coeficient Formula (per point):

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

By
a(i) : Average distance of point *i* to all other points in the same cluster
b(i) : Average distance of point *i* to points in other nearby clusters
Silhoutte Score Formula:
If there are *n* data points, then :

$$Silhouette\ Score = \frac{1}{n} \sum_{i=1}^{n} s(i) \quad (4)$$

Table 4. Interpretation of *Silhoutte score*

Score Value	Interpretation
+1 (close to 1)	Very well clustered and clearly separated
0	Point is on the border of two clusters, can be ambiguous
< 0	Point may be misclustered (more with another cluster)

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 2, July 2025, Page 60-71
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i2.8900

7. Iterative Process of Clustering
   The iterative steps of DBSCAN are as follows:
   a. Trace each point in the dataset one by one.
   b. If a *core point* is found, the clustering process starts through expansion.
   c. The cluster is formed by adding all *density-reachable* points from the *core*.
   d. If it cannot be expanded, then the process continues to other points.
   Points that do not belong to any cluster are categorized as *outliers* (labeled -1).
8. Output and Interpretation
   The result of DBSCAN is:
   a. *Clusters*: Groups of points that have high spatial proximity and density.
   b. *Outliers*: Points that do not have enough neighbors or do not fit into any cluster.

3. RESULTS AND DISCUSSION

This research was conducted at the Investment Office and One-Stop Integrated Licensing Services of Langkat Regency using business licensing data for 2024. The data used is secondary data which includes information related to Limited Liability Companies (PT) and *Commanditaire Vennootschap* (CV) obtained from the One-Stop Integrated Investment and Licensing Service of Langkat Regency, with an observation period from January to December 2024.The sample used in this study consists of 3747 company data selected from the population of business licenses issued in Langkat Regency This type of research is hypothesis testing research because it uses the DB-SCAN method as a clustering technique. In addition, this research is also descriptive, because it aims to describe the phenomena that occur without emphasizing on analyzing the causes.

Table 5: Sample Data

Index	Nib	Company name	KBLI	Description of project risk	Description of business scale	License status
0	1101240054533	Harapan makmur lestari	46201	Low	Micro business	Already have a license
1	1101240054533	Harapan makmur lestari	46202	Low	Micro business	Already have a license
2	1101240054533	Harapan makmur lestari	46314	Low	Micro business	Already have a license
3	1101240054533	Harapan makmur lestari	46693	Low	Micro business	Already have a license
4	1201240064239	Kaysan Construction	46100	Low	Small business	No license yet
5	1201240064239	Kaysan Construction	46319	Low	Small business	No license yet
6	1201240064239	Kaysan Construction	46412	Low	Small business	No license yet
7	1201240064239	Kaysan Construction	46421	Low	Small business	No license yet
8	1201240064239	Kaysan Construction	46422	Low	Small business	No license yet
9	1201240064239	Kaysan Construction	41011	Medium High	Small business	No license yet

3.1 Model Testing of DBScan Algorithm (*Density-Based Spatial Clustering With Noise*)

1. *Encoding* sample data into ordinal format data can be seen in Table 4 below:

Table 6: *Encoding* Result Data

Index	Nib	Company name	KBLI	Risk_Num	Scale_Num	License_Num
0	1101240054533	Harapan makmur lestari	46201	2	2	1
1	1101240054533	Harapan makmur lestari	46202	2	2	1
2	1101240054533	Harapan makmur lestari	46314	2	2	1
3	1101240054533	Harapan makmur lestari	46693	2	2	1

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 2, July 2025, Page 60-71
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i2.8900

4	1201240064239	Kaysan Construction	46100	2	1	0
5	1201240064239	Kaysan Construction	46319	2	1	0
6	1201240064239	Kaysan Construction	46412	2	1	0
7	1201240064239	Kaysan Construction	46421	2	1	0
8	1201240064239	Kaysan Construction	46422	2	1	0
9	1201240064239	Kaysan Construction	41011	1	1	0

2. Sample data normalized with *Z-Score Normalization* can be seen in the following table 5:

Table 7: Normalization Result Data

Index	Nib	Company name	KBLI	X1_Risk	X2_Scale	X3_License
0	1101240054533	Harapan makmur lestari	46201	0.58723751	0.465926088	0.505996217
1	1101240054533	Harapan makmur lestari	46202	0.58723751	0.465926088	0.505996217
2	1101240054533	Harapan makmur lestari	46314	0.58723751	0.465926088	0.505996217
3	1101240054533	Harapan makmur lestari	46693	0.58723751	0.465926088	0.505996217
4	1201240064239	Kaysan Construction	46100	0.58723751	-1.969626181	-1.97629936
5	1201240064239	Kaysan Construction	46319	0.58723751	-1.969626181	-1.97629936
6	1201240064239	Kaysan Construction	46412	0.58723751	-1.969626181	-1.97629936
7	1201240064239	Kaysan Construction	46421	0.58723751	-1.969626181	-1.97629936
8	1201240064239	Kaysan Construction	46422	0.58723751	-1.969626181	-1.97629936
9	1201240064239	Kaysan Construction	41011	-0.567519147	-1.969626181	-1.97629936

3. Sample data after normalization:
   a. Point A = index 3 → company: Harapan Makmur Lestari**,** low risk
   b. Point B = index 9 → company: KAYSAN KONSTRUKSI, medium high risk

Table 8: Manual Calculation Sample Data

Index	X1 (Risk)	X2 (Scale)	X3 (Permit)
3	0.58723751	0.465926088	0.505996217
9	-0.567519147	-1.969626181	-1.97629936

4. Calculating *Euclidean* Distance
   Formula:

$$d(A,B) = \sqrt{(X1^A - X1^B)^2 + (X2^A - X2^B)^2 + (X3^A - X3^B)^2}$$

Calculate:

$$d(A,B) = \sqrt{(0.58723751 - (-0.567519147))^2 + (0.465926088 - (-1.969626181))^2 + (0.505996217 - (-1.97629936))^2}$$

$$= \sqrt{(1.154756657)^2 + (2.434442269)^2 + (2.4822955772)^2}$$

$$= \sqrt{1.333469 + 5.932917 + 6.161800} = \sqrt{13.428186} = 3.6646$$

5. Comparison Eps = 0.1
   a. Since$d(A,B) = 3.6646 > 1.0$, then point B is not within eps radius of point A
   b. Hence, point B is not density-reachable from A.
   c. If point A is a core point (has ≥ 3 neighbors within 0.1 radius), then A will start a *cluster of* its own.
   d. But since B is not reachable from A**,** DBSCAN will evaluate B separately.
   e. If B does not have ≥ 3 neighbors within 0.1 radius → B will be an outlier (*Cluster* = -1).
   f. If there is another *core point* that bridges A ↔ B, the two can be merged.
   g. Mark points that do not belong to any cluster as noise (*outliers*).
6. Calculate *Silhouette Score* based on Company data:

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 2, July 2025, Page 60-71
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i2.8900

a. Point A (Index 3) = Harapan Makmur Lestari
b. Point B (Index 9) = Kaysan Konstruksi

Table 9: *Sillhouette* Components

Point	*a(i)*	*b(i)*	*s(i)*
A (Harapan Makmur Lestari)	0.000	3.57095	1.0000
B (Kaysan Konstruksi)	1.15476	3.66431	0.68486

$$Silhouette\ Score = \frac{s_A + s_B}{2} = \frac{1.0000 + 0.68486}{2} = 0.8424$$

The dataset shown is the result of preprocessing that has been adjusted for the needs of DBSCAN modeling. The three main categorical variables Project Risk Description, Business Scale Description, and Permit Status are converted to numeric form through label encoding, respectively into Risk_Num, Scale_Num, and Permit_Num. Next, the three numeric features were normalized using Standard *Scaler* to produce a standardly distributed vector of X1_Risk, X2_Scale, and X3_Permit (mean = 0, standard deviation = 1). This process aims to equalize the scale between features so that the distance between data points in the 3-dimensional feature space can be calculated fairly and proportionally. Entity identities such as NIB and Company Name are still included as referential information, but are not involved in the clustering process. To measure how good DBSCAN is at clustering data, an evaluation through the *Silhouette Score* value is used. The results of manual calculations of two companies from different clusters Harapan Makmur Lestari and Kaysan Konstruksi show *Silhouette Coefficient* values of 1.000 and 0.684 respectively. From these two values, an average *Silhouette Score* of 0.842 is obtained, which reflects excellent clustering quality. This figure shows that the data objects are very close to their own cluster members and quite far from other clusters, which is the hallmark of an effective and structured cluster separation.

Table 10: DBSCAN Clustering Result Data

Index	Nib	Company name	KBLI	X1_Risk	X2_Scale	X3_License	Cluster
0	1101240054533	Harapan makmur lestari	46201	0.58723751	0.465926088	0.505996217	0
1	1101240054533	Harapan makmur lestari	46202	0.58723751	0.465926088	0.505996217	0
2	1101240054533	Harapan makmur lestari	46314	0.58723751	0.465926088	0.505996217	0
3	1101240054533	Harapan makmur lestari	46693	0.58723751	0.465926088	0.505996217	0
4	1201240064239	Kaysan Construction	46100	0.58723751	-1.969626181	-1.97629936	1
5	1201240064239	Kaysan Construction	46319	0.58723751	-1.969626181	-1.97629936	1
...	...	...	...	...	...	...	
3745	2511240077199	Jakida Usiba Tama	47845	0.58723751	0.465926088	0.505996217	0
3746	2511240077199	Jakida Usiba Tama	47845	0.58723751	0.465926088	0.505996217	0
3747	2711240012228	Hasya Sehat Sejahtera	47721	1.741994168	0.465926088	0.505996217	4

Based on the results of the implementation of the DBSCAN algorithm with parameters eps = 1 and min_samples = 4 on company data that includes information on project risk, business scale, and license status, three main cluster groups are obtained without any data classified as *noise*. The normalization process was carried out on three numerical variables from the *encoding* results, namely Risk_Num, Scale_Num, and Permit_Num, which were then used as a three-dimensional vector representation (X1_Risk, X2_Scale, X3_Permit). The cluster results show that

1. *Cluster* 0 is dominated by low-risk, micro-scale, and licensed companies, which can be recommended for routine supervision.
2. Cluster 1 consists of low-risk but unlicensed companies, which could potentially be given an administrative warning. Meanwhile,
3. Cluster 2 includes medium-high risk companies without licenses, which require stricter supervisory actions such as temporary or permanent closure.

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 2, July 2025, Page 60-71
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i2.8900

Evaluation of the cluster results shows a *Silhouette Score* value of 0.9984 (close to 1), indicating that the separation between clusters is optimal and the cluster structure is clearly formed. The *score* reinforces that the DBSCAN algorithm is effective in mapping business supervision and control priorities automatically, data-driven, by considering risk parameters, business scale, and license status.

index	NIB	Nama Perusahaan	Uraian Risiko Proyek	Uraian Skala Usaha	Status Izin	KBLI	Kategori Pengawasan	Cluster_DBSCAN
0	1101240054533	HARAPAN MAKMUR LESTARI	rendah	usaha mikro	sudah punya izin	46201	Monitoring rutin	0
1	1101240054533	HARAPAN MAKMUR LESTARI	rendah	usaha mikro	sudah punya izin	46202	Monitoring rutin	0
2	1101240054533	HARAPAN MAKMUR LESTARI	rendah	usaha mikro	sudah punya izin	46314	Monitoring rutin	0
3	1101240054533	HARAPAN MAKMUR LESTARI	rendah	usaha mikro	sudah punya izin	46693	Monitoring rutin	0
4	1201240064239	KAYSAN KONSTRUKSI	rendah	usaha kecil	belum punya izin	46100	Peringatan administratif	1
5	1201240064239	KAYSAN KONSTRUKSI	rendah	usaha kecil	belum punya izin	46319	Peringatan administratif	1
6	1201240064239	KAYSAN KONSTRUKSI	rendah	usaha kecil	belum punya izin	46412	Peringatan administratif	1
7	1201240064239	KAYSAN KONSTRUKSI	rendah	usaha kecil	belum punya izin	46421	Peringatan administratif	1
8	1201240064239	KAYSAN KONSTRUKSI	rendah	usaha kecil	belum punya izin	46422	Peringatan administratif	1
9	1201240064239	KAYSAN KONSTRUKSI	menengah tinggi	usaha kecil	belum punya izin	41011	Pemberhentian sementara kegiatan operasional	2
10	1201240064239	KAYSAN KONSTRUKSI	menengah tinggi	usaha kecil	belum punya izin	41012	Pemberhentian sementara kegiatan operasional	2
11	1201240064239	KAYSAN KONSTRUKSI	menengah tinggi	usaha kecil	belum punya izin	41015	Pemberhentian sementara kegiatan operasional	2
12	1201240064239	KAYSAN KONSTRUKSI	menengah tinggi	usaha kecil	belum punya izin	42101	Pemberhentian sementara kegiatan operasional	2
13	1201240064239	KAYSAN KONSTRUKSI	menengah tinggi	usaha kecil	belum punya izin	42201	Pemberhentian sementara kegiatan operasional	2
14	1201240064239	KAYSAN KONSTRUKSI	menengah tinggi	usaha kecil	belum punya izin	41016	Pemberhentian sementara kegiatan operasional	2
15	1201240064239	KAYSAN KONSTRUKSI	menengah tinggi	usaha kecil	belum punya izin	41019	Pemberhentian sementara kegiatan operasional	2
16	1201240068754	KARYA UDIN BETON	rendah	usaha mikro	sudah punya izin	47528	Monitoring rutin	0
17	1201240068754	KARYA UDIN BETON	rendah	usaha mikro	sudah punya izin	47920	Monitoring rutin	0
18	1201240072591	BAIHAQI KONSTRUKSI	rendah	usaha kecil	belum punya izin	46100	Peringatan administratif	1
19	1201240072591	BAIHAQI KONSTRUKSI	rendah	usaha kecil	belum punya izin	46319	Peringatan administratif	1
20	1201240072591	BAIHAQI KONSTRUKSI	rendah	usaha kecil	belum punya izin	46412	Peringatan administratif	1
21	1201240072591	BAIHAQI KONSTRUKSI	rendah	usaha kecil	belum punya izin	46421	Peringatan administratif	1
22	1201240072591	BAIHAQI KONSTRUKSI	rendah	usaha kecil	belum punya izin	46422	Peringatan administratif	1
23	1201240072591	BAIHAQI KONSTRUKSI	menengah tinggi	usaha kecil	belum punya izin	41011	Pemberhentian sementara kegiatan operasional	2
24	1201240072591	BAIHAQI KONSTRUKSI	menengah tinggi	usaha kecil	belum punya izin	41012	Pemberhentian sementara kegiatan operasional	2

Show 25 per page 1 2 10 100 140 150

Figure 2. DBSCAN results

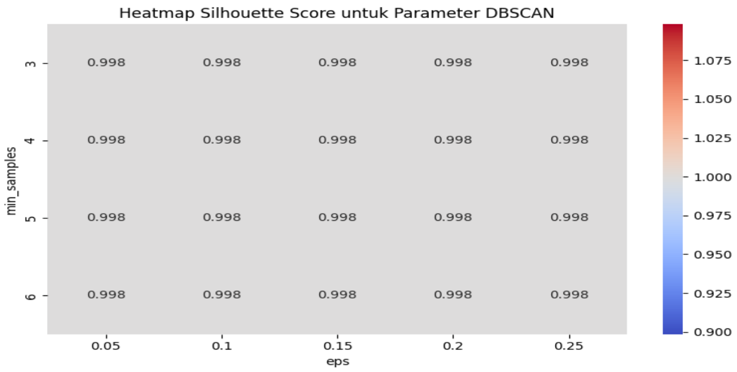

Figure 3. Silhouette Score Heatmap for DBSCAN Parameters

As can be seen in Figure 3, DBSCAN with eps = 1.00 and minimum samples = 4 produces 17 clusters with a *Silhouette Score of* 0.9984 (close to 1), indicating excellent cluster separation. This value indicates that the entities within each cluster are very homogeneous and clearly separated from other clusters. This configuration is effective for grouping businesses based on risk, scale, and license status in support of data-driven supervision prioritization.

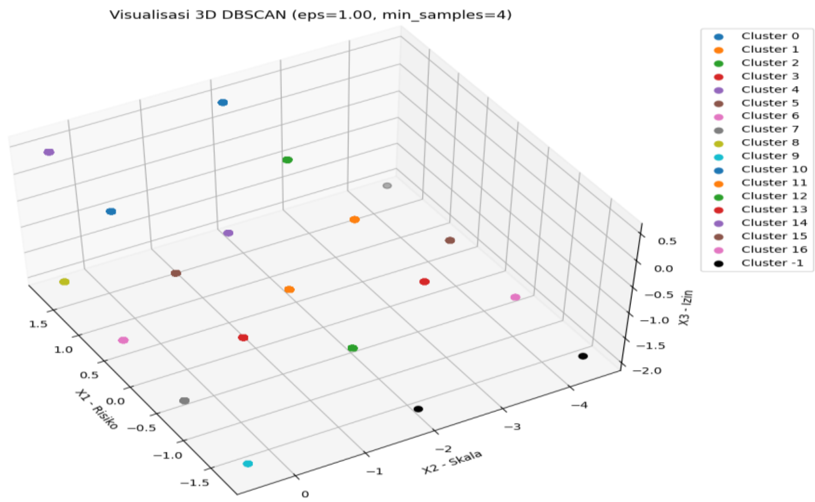

Figure 4. 3D Result of Risk, Scale and License Comparison

The IJICS (International Journal of Informatics and Computer Science)
Vol 9 No 2, July 2025, Page 60-71
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i2.8900

The clustering results using the DBSCAN algorithm show that business entities are successfully grouped based on three main normalized variables, namely business risk (X1_Risk), business scale (X2_Scale), and license status (X3_Izin). Each business is identified through NIB and equipped with information on company name, risk level, scale, and license ownership. Clustering results in the variables Cluster_DBSCAN and Supervision Category, which are constructed based on a combination of risk and permit parameters. For example, HARAPAN MAKMUR LESTARI, a low-risk, micro-scale, and licensed company, falls under the category of Routine Monitoring and Evaluation in cluster 0. In contrast, BAIHAQI KONSTRUKSI with medium-high risk and no license is classified as Temporary Business Closure in cluster 2. This approach proves that DBSCAN is effective in clustering business patterns based on risk and compliance, and supports objective decision-making in business supervision.

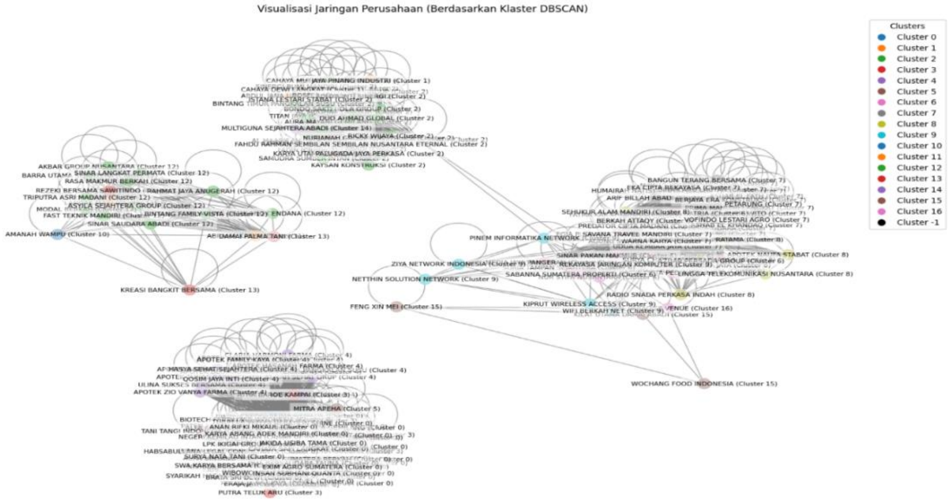

Figure 5. Company Network Visualization Results

4. CONCLUSIONS

The results showed that the DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) algorithm can be effectively applied to map the priority of business supervision and control in Langkat Regency based on project risk parameters, business scale, and license status. Using secondary data of 3,748 business entities that have been normalized and encoded into three numerical features (X1_Risk, X2_Scale, X3_Izin), DBSCAN successfully formed 17 clusters with a *Silhouette Score* value of 0.9984 which means close to 1, indicating optimal cluster separation and strong density within each group. Each cluster was interpreted into supervision categories, such as routine supervision, administrative warning, and recommendation for temporary or permanent business closure. DBSCAN is also able to identify *outliers*, i.e. companies that do not belong to any cluster and potentially have specific supervision characteristics. Visualizations in the form of 3D scatter plots and heatmaps help clarify the distribution of clusters and spatial relationships between business entities. This approach makes an important contribution in supporting the efficiency, transparency, and objectivity of the supervisory decision-making process by DPMPTSP. Thus, this method is worth adopting as a data-driven analysis tool in regional business supervision and control policies.

REFERENCES

[1] S. N. Girfita, W. Siswanti, and E. Altiarika, “Pemanfaatan Teknologi Artificial Intelligence (AI) Dalam Digitalisasi UMKM Di Desa Namang,” *Jurnal Pengabdian kepada Masyarakat Nusantara (JPkMN)*, vol. Vol. 5 No. 4, pp. 5609–5615, Oct. 2024, Accessed: Jul. 01, 2025. [Online]. Available: https://ejournal.sisfokomtek.org/index.php/jpkm/article/download/4579/2899/33161

[2] Z. Sitorus, Ganefri, and Refdinal, “Development of Deeper Learning Cycle-Project Based Learning Model Based on Resource Sharing in Artificial Neural Network Courses,” *International Journal of Recent Technology and Engineering (IJRTE)*, vol. Volume-8, no. Issue-5, pp. 1698–1702, Jan. 2020, Accessed: Jul. 02, 2025. [Online]. Available: https://www.researchgate.net/profile/Zulham-Sitorus-2/publication/354986730_International_Journal_of_Recent_Technology_and_Engineering_IJRTE/links/6156db2d61a8f46670910e34/International-Journal-of-Recent-Technology-and-Engineering-IJRTE.pdf

[3] R. S. Lisudatu, M. Marchelin, and L. K. Wibisono, “Klasifikasi UMKM Berdasarkan Kinerja Keuangan Menggunakan Algoritma K-Means Clustering,” *Jurnal Bisnis Mahasiswa*, vol. 5, no. 2, pp. 891–900, Mar. 2025, doi: 10.60036/JBM.554.

[4] L. N. I. Fitriana and Rohmaniyah, “Penaksiran Laba Pada Akuntansi UMKM Via Online di Kabupaten Sumenep,” *AKUA: Jurnal Akuntansi dan Keuangan*, vol. 1, no. 2, pp. 178–185, Apr. 2022, doi: 10.54259/AKUA.V1I2.613.

Vol 9 No 2, July 2025, Page 60-71
ISSN 2548-8384 (online), ISSN 2548-8449 (print)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/ijics/index
DOI 10.30865/ijics.v9i2.8900

[5] A. I. Zalukhu and M. Iqbal, "Analysis of Product Demand Prediction Using Decision Tree on Sales Data of Ceria Toys Store," *Journal Of Data Science* , vol. 3, no. 01, pp. 10–22, Mar. 2025, doi: 10.58471/JDS.V3I01.6458.

[6] A. I. Zalukhu, D. Sartika, and S. Wahyuni, "Penerapan Algoritma Apriori untuk Optimasi Strategi Penjualan Berdasarkan Analisis Pola Pembelian di Torsa Cafe," *Bulletin of Information Technology (BIT)*, vol. 5, no. 4, pp. 295–304, Dec. 2024, doi: 10.47065/BIT.V5I4.1715.

[7] A. I. Zalukhu, M. Iqbal, and D. Nasution, "ANALISIS DATA MINING DALAM PENGELOLAAN PERSEDIAAN STOK DENGAN ALGORITMA RANDOM FOREST DAN APRIORI (STUDI KASUS: TOKO CERIA BABYSHOP)," *JOURNAL OF SCIENCE AND SOCIAL RESEARCH*, vol. 8, no. 3, pp. 3396–3405, Jun. 2025, doi: 10.54314/JSSR.V8I3.3544.

[8] Z. Sitorus, I. Syahputra, C. I. Angkat, and D. Sartika, "Implementation of K-Means Clustering for Inventory Projection," *International Journal of Science, Technology & Management*, vol. 5, no. 3, pp. 673–678, May 2024, doi: 10.46729/IJSTM.V5I3.856.

[9] Suhardiansyah and M. Iqbal, "Optimization of Kaayana Store Inventory through Transaction Pattern Analysis Using the Apriori Algorithm," *Journal Of Data Science* , vol. 3, no. 01, pp. 1–9, Feb. 2025, doi: 10.58471/JDS.V3I01.6398.

[10] D. Sartika, M. Iqbal, and Z. Sitorus, "Analisis Algoritma Exponensial Smoothing dan K-Means untuk Optimal-isasi Penerimaan Mahasiswa Baru di Universitas Haji Sumatera Utara," *Jurnal Multimedia dan Teknologi Informasi (Jatilima)*, vol. 7, no. 02, pp. 1–10, Apr. 2025, doi: 10.54209/JATILIMA.V7I02.1269.

[11] S. Simamora, M. Iqbal, A. P. U. Siahaan, K. Khairul, and Z. Sitorus, "Comparison of K-Means and Self Organizing Map Algorithms for Ground Acceleration Clustering," *Sinkron : jurnal dan penelitian teknik informatika*, vol. 8, no. 4, pp. 2345–2353, Oct. 2024, doi: 10.33395/SINKRON.V8I4.14120.

[12] A. Fitri, N. Masruriyah, C. Emilia, S. Budi, and A. Dermawan, "Memahami Data Mining dengan Python: Implementasi Praktis," *Eureka Media Aksara*, Mar. 2024, Accessed: Jun. 26, 2025. [Online]. Available: https://repository.penerbiteureka.com/publications/568010/

[13] N. Q. Rahmawati, "IMPLEMENTASI DATA MINING TERHADAP DATA MINAT MAHASISWA MENGGUNAKAN METODE APRIORI," *Jurnal Teknologi Kimia Unimal*, vol. 13, no. 2, pp. 215–228, Nov. 2024, doi: 10.29103/JTKU.V13I2.19563.

[14] T. D. Saribu, M. S. Wahyuni, and S. F. Rezky, "Data Mining Dalam Menentukan Pola Penjualan Barang Menggunakan Algoritma Apriori," *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, vol. 3, no. 2, pp. 183–190, Mar. 2024, doi: 10.53513/JURSI.V3I2.6145.

[15] R. R. (Randi) Putra and C. (Cendra) Wadisman, "Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K Means," *Intecoms*, vol. 1, no. 1, pp. 72–77, Mar. 2018, doi: 10.31539/INTECOMS.V1I1.141.

[16] A. O. S. Ayu and M. Iqbal, "Analisis Data Mining Terhadap Data Faktor Perceraian Di Sumatera Utara Dengan Metode K-Means Clusstering," *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, vol. 4, no. 1, pp. 214–221, Jan. 2025, doi: 10.53513/JURSI.V4I1.10671.

[17] A. Nur *et al.*, "Implementasi Perbandingan Algoritma k-Means dan DB-Scan Pada Beban Listrik Rumah Tangga," *INTEGER: Journal of Information Technology*, vol. 10, no. 1, Mar. 2025, doi: 10.31284/J.INTEGER.2024.V10I1.7479.

[18] M. F. Rizki and F. Sulianta, "Analisis Klasterisasi Data Peserta Asuransi PT Xyz Menggunakan Metode Denisty-Based Spatial Clustering of Applications with Noise (DBSCAN)," *Jurnal Algoritma*, vol. 22, no. 1, pp. 993–1001, May 2025, doi: 10.33364/ALGORITMA/V.22-1.2417.

[19] N. A. Permatasari, Y. H. Chrisnanto, and A. K. Ningsi, "Segmentasi Kasus Data Kematian Covid 19 Di Jawa BaratMenggunakan Algoritma DBSCAN," *International Journal of Engineering, Economic, Social Politic and Government* , vol. Vol. 1, No. 4, 2023, Accessed: Jul. 01, 2025. [Online]. Available: https://ijespgjournal.org/index.php/ijespg/article/view/78/66

[20] Risman, Syaripuddin, and Suyitno, "IMPLEMENTASI METODE DBSCAN PADA PENGELOMPOKAN KABUPATEN/KOTA DI PULAU KALIMANTAN BERDASARKAN INDIKATOR KESEJAHTERAAN RAKYAT," *Prosiding Seminar Nasional Matematika dan Statistika*, vol. 1, pp. 22–28, May 2019, Accessed: Jul. 01, 2025. [Online]. Available: https://jurnal.fmipa.unmul.ac.id/index.php/SNMSA/article/view/522

[21] S. Agustian Hudjimartsu, T. Aurilia Viona, H. Sharfina, and Yani, "Optimasi Parameter Algoritma DBSCAN untuk Mendeteksi Titik Panas Kebakaran Hutan dan Lahan," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 9, no. 3, pp. 355–361, Dec. 2023, doi: 10.26418/JP.V9I3.61714.