



## Identifikasi Klaster UMKM di Kota Bima menuju Indonesia Emas 2045 dengan Metode *Support Vector Machine*

Khairunnas<sup>1\*</sup>, Alya Masitha<sup>2</sup>, Rafiuddin<sup>3</sup>

<sup>1</sup>Program Studi Ilmu Komputer, Universitas Muhammadiyah Bima, Indonesia

<sup>2</sup>Program Studi Rekayasa Perangkat Lunak, Institut Teknologi Statistika dan Bisnis Muhammadiyah Semarang, Indonesia

<sup>3</sup>Program Studi Ekonomi Syari'ah, Universitas Muhammadiyah Bima, Indonesia

---

### Artikel Info

---

#### Kata Kunci:

Artificial Intelligence;  
Data Mining;  
Machine Learning;  
Transformasi Digital;  
UMKM.

#### Keywords:

Artificial Intelligence;  
Data Mining;  
Machine Learning;  
Digital Transformation;  
MSMEs.

#### Riwayat Artikel:

Submitted: 08 September 2025

Accepted: 27 Oktober 2025

Published: 29 Oktober 2025

**Abstrak:** UMKM merupakan badan usaha yang berperan vital dalam perekonomian daerah, termasuk di Kota Bima dengan jumlah 23.936 unit pada tahun 2022. Mayoritas terdiri dari usaha mikro dengan Kecamatan Raba dan Rasanae Barat sebagai wilayah dengan konsentrasi tertinggi. Penelitian ini bertujuan mengembangkan model klasterisasi UMKM berbasis bidang usaha menggunakan pendekatan machine learning dengan kombinasi algoritma Fuzzy C-Means (FCM) dan Support Vector Machine (SVM). Data penelitian berjumlah 5.176 unit UMKM yang diperoleh dari Dinas Koperasi dan UKM, survei lapangan, serta dokumen administratif. Tahapan penelitian mencakup preprocessing data (penanganan missing values, feature selection, cleaning, normalisasi, dan reduksi dimensi), klasterisasi dengan FCM, klasifikasi menggunakan SVM, serta evaluasi model. Penentuan jumlah klaster optimum dilakukan dengan Silhouette Index dan Davies-Bouldin Index, menghasilkan tiga klaster utama: UMKM tradisional dengan investasi rendah dan digitalisasi minim, UMKM transisi dengan investasi lebih tinggi, serta UMKM digital progresif dengan adopsi teknologi dominan. Hasil evaluasi menunjukkan bahwa SVM mampu mengklasifikasikan klaster dengan akurasi 100% serta precision, recall, dan f1-score sempurna. Fitur Jumlah Investasi, TKI, dan Proxy Digitalisasi terbukti representatif dalam membedakan karakteristik UMKM. Temuan ini menunjukkan bahwa kombinasi FCM-SVM efektif untuk segmentasi UMKM serta dapat menjadi dasar kebijakan pengembangan yang tepat sasaran, mulai dari penguatan usaha tradisional hingga akselerasi transformasi digital.

**Abstract:** MSMEs are business entities that play a significant role in the regional economy, including in Bima City with a total of 23,936 units in 2022. Most of them consist of micro-enterprises with Raba and West Rasanae Districts as the areas with the highest concentration. This study aims to develop a business sector-based MSME clustering model using a machine learning approach with a combination of the FCM and SVM algorithms. The research data amounted to 5,176 MSME units obtained from the Cooperatives and SMEs Office, field surveys, and administrative documents. The research stages include data preprocessing (handling missing values, feature selection, cleaning, normalization, and dimensionality reduction), clustering with FCM, classification using SVM, and model evaluation. Determination of the optimum number of clusters was carried out using the Silhouette Index and Davies-Bouldin Index, resulting in three main clusters: traditional MSMEs with low investment and minimal digitalization, transitional MSMEs with higher investment,

*and progressive digital MSMEs with dominant technology implementation. The evaluation results show that SVM is able to classify clusters with 100% accuracy and perfect precision, recall, and f1-score. The Investment Amount, Indonesian Migrant Workers (TKI), and Digitalization Proxy features are proven to be representative in distinguishing MSME characteristics. These findings indicate that the FCM–SVM combination is effective for MSME segmentation and can serve as the basis for targeted development policies, from strengthening traditional businesses to accelerating digital transformation.*

---

**Corresponding Author:**

Khairunnas

Email: kkhairunnas.khairunnas@gmail.com

---

## PENDAHULUAN

Usaha Mikro, Kecil, dan Menengah (UMKM) merupakan salah satu pilar utama dalam pembangunan ekonomi nasional (Muthiah & Hamidah, 2025). Keberadaannya tidak hanya memberikan kontribusi terhadap pertumbuhan ekonomi, tetapi juga berperan signifikan dalam menciptakan lapangan kerja, meningkatkan pendapatan masyarakat, serta menekan angka kemiskinan (Maharani et al., 2024). UMKM menjadi motor penggerak yang mampu menjaga stabilitas ekonomi, terutama pada saat terjadi krisis, karena sifatnya yang fleksibel dan dekat dengan kebutuhan masyarakat (Boubaker et al., 2025; Vásquez et al., 2021). Peran strategis ini menjadikan UMKM sebagai salah satu faktor penting dalam upaya memperkuat ketahanan ekonomi nasional sekaligus meningkatkan kesejahteraan sosial (Agustiono & Darmawan, 2023; Wibowo et al., 2024). Upaya pemerintah untuk meningkatkan pertumbuhan ekonomi, salah satunya adalah memberdayakan UMKM (Sabale et al., 2024).

UMKM dihadapkan pada tantangan besar, khususnya di era Revolusi Industri 4.0. Perubahan pola bisnis akibat digitalisasi menuntut UMKM untuk beradaptasi dengan cepat terhadap berbagai inovasi, seperti penggunaan platform digital, sistem pembayaran elektronik, hingga pemanfaatan *Internet of Things* (IoT) (Herdijeni et al., 2025; Maharani et al., 2024). Transformasi digital tidak hanya menjadi kebutuhan, tetapi juga keharusan agar UMKM tetap mampu bersaing di pasar yang semakin kompetitif (Patil et al., 2023). Keterbatasan sumber daya manusia, akses terhadap teknologi, serta pemahaman terkait literasi digital masih menjadi kendala utama (Putri et al., 2024). Oleh karena itu, penguatan kapasitas digital UMKM perlu dilakukan melalui pelatihan, pendampingan, dan kebijakan pemerintah yang mendukung (Nararto et al., 2024). UMKM diharapkan dapat memanfaatkan peluang dari perkembangan ekonomi digital sekaligus mengatasi tantangan yang ada, sehingga perannya sebagai penopang ekonomi nasional semakin kuat dan berkelanjutan (Nararto et al., 2024).

Perkembangan UMKM di Bima, berdasarkan laporan dari Dinas Koperasi dan Usaha Mikro Kecil Menengah Kota Bima pada tahun 2022, jumlah UMKM mencapai 5.187 unit, dengan sektor perdagangan mendominasi lebih dari 70% dari total unit yang ada. Meskipun memiliki kontribusi penting terhadap perekonomian daerah, UMKM di Bima masih menghadapi tantangan struktural, terutama dalam hal pemetaan jenis usaha, segmentasi pasar, dan strategi pembinaan yang tepat sasaran. Kurangnya pendekatan berbasis data dalam pengelompokan UMKM menyebabkan intervensi program seringkali tidak efektif dalam memberikan dampak nyata. Klasifikasi usaha UMKM yang ada di Kota Bima ke dalam klaster usaha berdasarkan bidang dan karakteristik usahanya dapat diidentifikasi menggunakan machine learning.

Algoritma machine learning menjadi salah satu pendekatan yang dapat mengatasi masalah tersebut (Gafar & Wang, 2023). Teknologi ini mampu mengidentifikasi pola serta mengelompokkan UMKM berdasarkan bidang dan karakteristik usahanya secara lebih akurat (Chiu et al., 2024). Dalam penelitian ini, algoritma FCM digunakan untuk melakukan klasterisasi dengan mempertimbangkan atribut utama seperti jenis usaha, skala operasional, potensi pertumbuhan, dan tingkat adaptasi digital

(Hafsari & Nurcahyo, 2021; Jonathan et al., 2024). Hasil klasterisasi kemudian dianalisis lebih lanjut menggunakan algoritma SVM untuk melakukan klasifikasi jenis usaha sesuai dengan profil tiap UMKM. Dengan kombinasi kedua algoritma ini, UMKM di Kota Bima dapat dipetakan ke dalam klaster usaha yang representatif, sehingga hasilnya dapat dijadikan dasar dalam merancang strategi pembinaan dan pemberdayaan yang lebih tepat sasaran.

Metode FCM merupakan salah satu teknik *soft clustering* yang lebih fleksibel dibandingkan metode klasterisasi konvensional (Tahyudin et al., 2022). Keunggulan metode ini terletak pada kemampuannya memberikan kemungkinan bagi suatu data untuk menjadi anggota lebih dari satu klaster dengan tingkat keanggotaan tertentu (Beynon et al., 2025). Dalam prosesnya, setiap titik data diberi nilai probabilitas yang menunjukkan kedekatan dengan suatu klaster, sehingga hasil pengelompokan menjadi lebih adaptif dan akurat (Ouyang & Sun, 2024). Karakteristik tersebut dinilai sesuai untuk mengklasterisasi data UMKM, karena pengelompokan dapat dilakukan berdasarkan kemiripan atribut sekaligus tetap mempertahankan fleksibilitas dalam menentukan keanggotaan klaster (Kumar et al., 2021). Metode FCM digunakan untuk mengelompokkan data 5.176 UMKM yang tersebar di Kota Bima. Atribut utama yang dianalisis meliputi jenis usaha, skala operasional, potensi pertumbuhan, dan tingkat adaptasi digital. Melalui pendekatan ini, diharapkan terbentuk klaster UMKM yang representatif sehingga dapat menjadi dasar bagi proses klasifikasi selanjutnya serta mendukung penyusunan strategi pengembangan UMKM yang lebih tepat sasaran.

Algoritma SVM merupakan salah satu metode *machine learning* yang digunakan dalam penelitian ini untuk melakukan klasifikasi jenis usaha UMKM. SVM dikenal efektif dalam memecahkan berbagai permasalahan klasifikasi karena memiliki kemampuan belajar yang baik dari data pelatihan (Mishra et al., 2023). Keunggulan utama algoritma ini terletak pada kemampuannya mencari *hyperplane* optimal yang dapat memisahkan data ke dalam kelas yang berbeda secara maksimal (Bhavani & Santhosh Kumar, 2021; Sheth et al., 2022). SVM tidak hanya mampu menghasilkan model klasifikasi yang akurat, tetapi juga memiliki tingkat generalisasi yang tinggi sehingga dapat diterapkan pada berbagai bidang, termasuk dalam identifikasi dan pengelompokan UMKM (Muthiah & Hamidah, 2021). Dalam konteks penelitian ini, SVM digunakan untuk mengklasifikasikan UMKM berdasarkan hasil klasterisasi sebelumnya dengan mempertimbangkan atribut-atribut utama, seperti jenis usaha, skala operasional, potensi pertumbuhan, dan tingkat adaptasi digital. Melalui kombinasi ini, proses klasifikasi diharapkan mampu menghasilkan pengelompokan UMKM yang lebih terstruktur dan mendukung strategi pengembangan yang tepat sasaran.

Penelitian ini memiliki urgensi yang tinggi karena berfokus pada pembangunan model klasterisasi UMKM yang adaptif dan berbasis data. Selama ini, pengelolaan UMKM seringkali dilakukan tanpa didukung oleh pemetaan yang komprehensif, sehingga intervensi program pembinaan kurang tepat sasaran. Melalui pendekatan berbasis data, penelitian ini diharapkan dapat menghasilkan klaster UMKM yang merepresentasikan kondisi riil di lapangan, baik dari segi jenis usaha, skala operasional, potensi pertumbuhan, maupun tingkat adaptasi digital. Klasterisasi semacam ini akan memberikan gambaran yang lebih jelas mengenai kebutuhan dan tantangan UMKM, sehingga strategi pembinaan dapat dirancang dengan lebih terarah. Penelitian ini relevan dengan arah pembangunan nasional jangka panjang, khususnya dalam mendukung visi Indonesia Emas 2045. Pada visi tersebut, salah satu pilar penting adalah penguatan ekonomi kerakyatan dan digitalisasi usaha mikro. Penelitian ini tidak hanya memberikan kontribusi akademis, tetapi juga memberikan manfaat praktis bagi pemerintah daerah, pelaku usaha, maupun pemangku kepentingan lainnya.

## METODE

Penelitian ini dimulai dengan tahapan *preprocessing data* yang berfungsi untuk memastikan kualitas dataset UMKM Kota Bima sebelum masuk ke tahap analisis. Dataset yang diperoleh umumnya masih memiliki kelemahan seperti data hilang (*missing values*), variabel yang tidak relevan, serta distribusi data yang tidak seragam. Oleh karena itu, langkah pertama adalah melakukan *handling missing data* untuk mengisi atau menghapus data yang hilang agar tidak menimbulkan bias.

Selanjutnya, dilakukan *feature selection* untuk memilih atribut yang paling berpengaruh terhadap analisis UMKM berdasarkan bidang usaha. Proses ini penting agar model tidak terbebani oleh variabel yang kurang relevan. Setelah itu, data juga melalui tahapan *cleaning* untuk menghilangkan duplikasi atau ketidakkonsistenan data. Langkah ini menjadi dasar untuk menghasilkan data yang bersih, valid, dan siap diolah menggunakan metode *machine learning*.

Tahap selanjutnya dalam *preprocessing* adalah normalisasi data dengan metode Min-Max Scaling. Normalisasi diperlukan agar semua fitur memiliki skala nilai yang seragam, sehingga algoritma clustering dan klasifikasi tidak mendominasi hanya karena perbedaan skala antar atribut (Putri et al., 2021). Rumus disajikan pada persamaan (1)

$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Selain itu, dilakukan pula reduksi dimensi menggunakan *Principal Component Analysis (PCA)* untuk menyederhanakan kompleksitas data. PCA bekerja dengan mentransformasikan variabel-variabel asli ke dalam dimensi baru yang lebih ringkas namun tetap mampu mempertahankan informasi utama dalam data. Hal ini sangat membantu mengurangi redundansi fitur sekaligus meningkatkan efisiensi perhitungan pada tahap berikutnya, dan data yang sudah melalui normalisasi dan reduksi dimensi menjadi lebih optimal untuk dianalisis menggunakan algoritma *FCM*.

Pada tahap *clustering*, penelitian ini menggunakan metode *FCM* yang memiliki keunggulan dalam mengakomodasi derajat keanggotaan ganda pada setiap data. Hal ini berbeda dengan *hard clustering* seperti K-Means, di mana setiap data hanya boleh masuk ke satu klaster. Dengan *FCM*, setiap UMKM dapat memiliki nilai keanggotaan pada lebih dari satu klaster, meskipun dengan tingkat dominasi tertentu. Berikut persamaan algoritma *FCM* (Damayanti & Jakfar, 2023), dapat dilihat pada persamaan (2)

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \cdot \|x_i - c_j\|^2 \quad (2)$$

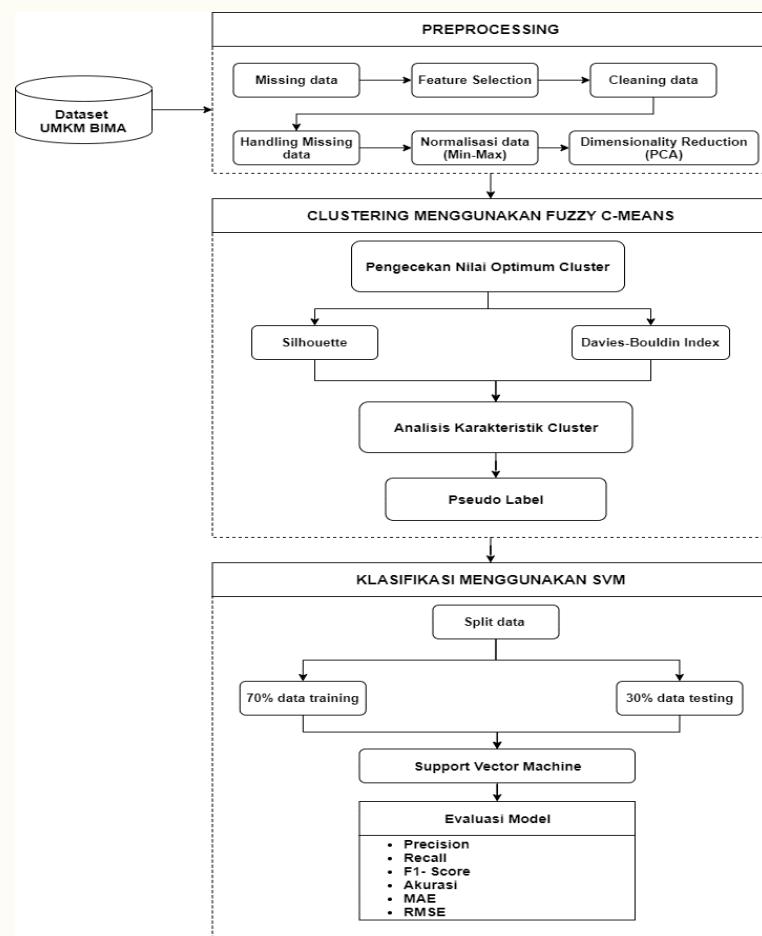
Sebelum menentukan hasil akhir klasterisasi, dilakukan pengecekan jumlah klaster optimum menggunakan dua metrik, yaitu *Silhouette Index* dan *Davies-Bouldin Index*. *Silhouette* mengukur kesesuaian data dalam klaster, sedangkan *Davies-Bouldin* mengukur tingkat pemisahan antar klaster (Ashari et al., 2023). Kedua indeks ini memberikan evaluasi yang lebih objektif dalam menentukan jumlah klaster terbaik yang mewakili struktur data UMKM Kota Bima.

Penelitian berlanjut pada analisis karakteristik masing-masing klaster. Analisis ini bertujuan untuk memahami profil khas dari setiap kelompok UMKM, misalnya kelompok yang dominan di sektor perdagangan, kuliner, jasa, atau sektor lainnya. Proses ini menghasilkan *pseudo-label* yang berfungsi sebagai kategori semu untuk setiap UMKM berdasarkan keanggotaannya pada klaster tertentu. Pemberian *pseudo-label* sangat penting karena hasil klasterisasi tidak menyediakan label asli seperti dalam klasifikasi terawasi (*supervised learning*). Dengan *pseudo-label* ini, penelitian mampu membangun landasan untuk proses klasifikasi berikutnya menggunakan algoritma *SVM*. *pseudo-label* menjadi jembatan penghubung antara tahap klasterisasi tidak terawasi (*unsupervised learning*) dengan klasifikasi terawasi (*supervised learning*).

Tahap klasifikasi dilakukan menggunakan algoritma *SVM*, yang dikenal memiliki performa tinggi dalam membedakan kategori data yang kompleks dan non-linear. Dataset yang sudah dilabeli secara semu kemudian dibagi menjadi dua bagian, yaitu 70% untuk data *training* dan 30% untuk data *testing*. Proses *training* digunakan untuk melatih model *SVM* agar dapat mengenali pola yang membedakan tiap klaster dan data *testing* digunakan untuk mengukur kemampuan model dalam memprediksi kategori klaster pada data yang belum pernah dilihat sebelumnya. Pemilihan rasio 70:30 bertujuan untuk menjaga keseimbangan antara kebutuhan pelatihan model dan evaluasi yang representatif. Proses klasifikasi ini diharapkan dapat menghasilkan model prediktif yang mampu mengenali pola UMKM baru dengan akurasi tinggi.

Tahapan akhir penelitian adalah evaluasi model klasifikasi untuk menilai kinerja *SVM* dalam mengklasifikasikan UMKM Kota Bima berdasarkan hasil *pseudo-label* dari *FCM*. Evaluasi dilakukan dengan menggunakan berbagai metrik, antara lain *precision*, *recall*, *F1-score*, *akurasi*, *Mean Absolute Error*

(*MAE*), dan *Root Mean Square Error (RMSE)*. Penelitian ini tidak hanya berfokus pada pencapaian akurasi tinggi, tetapi juga memastikan model yang dibangun memiliki kemampuan generalisasi yang baik. Hasil dari keseluruhan proses ini diharapkan mampu memberikan kontribusi nyata dalam pemetaan UMKM Kota Bima, sehingga dapat mendukung penyusunan strategi kebijakan pemerintah daerah dalam pengembangan sektor usaha mikro, kecil, dan menengah. Untuk memberikan gambaran yang lebih jelas, alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

Alur penelitian yang ditampilkan pada Gambar 1 menggambarkan keterkaitan antar setiap tahapan mulai dari *preprocessing*, klasterisasi dengan FCM, hingga klasifikasi menggunakan SVM beserta evaluasinya. Rangkaian langkah ini tidak hanya menunjukkan sistematika penelitian, tetapi juga menegaskan bahwa setiap proses memiliki peran penting dalam menghasilkan model yang akurat. Melalui rancangan alur yang komprehensif tersebut, penelitian ini diharapkan mampu memberikan gambaran yang jelas mengenai metodologi yang diterapkan sekaligus menjamin validitas hasil analisis yang diperoleh.

## HASIL DAN PEMBAHASAN

Hasil penelitian ini disajikan secara bertahap sesuai dengan alur penelitian yang telah dirancang. Tahap pertama dimulai dari proses pengumpulan data UMKM yang menjadi objek penelitian, kemudian dilanjutkan dengan tahapan *pre-processing* untuk memastikan kualitas dan kelayakan data. Selanjutnya, dilakukan proses klasterisasi menggunakan metode *FCM* guna mengelompokkan UMKM berdasarkan karakteristik bidang usahanya. Setelah klaster terbentuk, hasil tersebut digunakan sebagai dasar untuk proses klasifikasi menggunakan metode *SVM* sehingga dapat membangun model prediktif yang mampu mengenali pola pada data baru. Bagian akhir menyajikan hasil evaluasi model

dengan menggunakan beberapa metrik pengukuran kinerja guna menilai akurasi serta reliabilitas pendekatan yang digunakan.

### Pengumpulan dataset UMKM

Penelitian ini menggunakan pendekatan kuantitatif berbasis *machine learning* untuk mengidentifikasi klaster UMKM di Kota Bima. Data penelitian diperoleh melalui sumber data sekunder dari Dinas Koperasi, Perindustrian, dan Perdagangan (DISKOPERINDAG) Kota Bima. Data yang terkumpul kemudian diproses melalui tahap *preprocessing* yang meliputi normalisasi, encoding variabel kategorikal, serta penanganan data yang tidak lengkap. Data yang digunakan dalam penelitian ini diperoleh dari seluruh UMKM di Kota Bima yang tercatat sebanyak 5.176 unit pada tahun 2024. Sampel yang digunakan adalah 200 UMKM yang tersebar di 5 kecamatan utama di Kota Bima, dipilih secara purposive berdasarkan representasi bidang usaha, skala, dan lokasi. Jenis data yang digunakan adalah data sekunder dan primer, yang mencakup informasi umum dan spesifik UMKM seperti : jenis usaha, lokasi, jumlah tenaga kerja, kelompok usaha, aset atau modal. Data diperoleh dari dinas terkait (Dinas Koperasi dan UKM), survei lapangan, serta dokumen administratif dan laporan kegiatan UMKM. Survei lapangan dilakukan kepada 200 UMKM dengan kuesioner terstruktur. Data dokumenter diperoleh dari situs BPS, laporan UMKM, serta dokumen kelembagaan pemerintah daerah.

### Preprocessing data UMKM

Tahapan *pre-processing* data dalam penelitian ini diawali dengan pengecekan data yang hilang (*missing values*), dilanjutkan dengan pemilihan fitur yang relevan (*feature selection*), pembersihan data untuk mengatasi inkonsistensi (*data cleaning*), serta penanganan terhadap data yang hilang. Setelah itu dilakukan normalisasi untuk menyamakan skala antar variabel, dan diakhiri dengan proses reduksi dimensi (*dimensionality reduction*) guna menyederhanakan kompleksitas data tanpa mengurangi informasi penting.

### Pengecekan missing value

Pengecekan missing value atau nilai yang hilang merupakan tahapan dalam preprocessing data. Nilai hilang merupakan nilai yang tidak lengkap atau tidak ada dan hal ini dapat mempengaruhi hasil analisis data. Dataset yang didapatkan mengidentifikasi bahwa banyak sekali data yang hilang atau tidak lengkap. Pengecekan *missing values* menggunakan `df_raw.isnull()`, hasil dari pengecekan *missing value* dapat dilihat pada tabel 1.

Tabel 1. Hasil Missing Value

Atribut	Jumlah Missing Value
Nama_Proyek	393
No_Telp	114
Email	85
Judul_KBLI	53
Sektor_Pembina	53
ID_Proyek	0
Nama_Perusahaan	0
NIB	0
Tanggal_Terbit	0
.....	.....
TKI	0

Berdasarkan hasil yang ditunjukkan pada Tabel 1, diketahui bahwa terdapat lima atribut yang memiliki data tidak lengkap, yaitu **Nama\_Proyek**, **No\_Telp**, **Email**, **Judul\_KBLI**, dan **Sektor\_Pembina**. Ketidaklengkapan data pada atribut-atribut tersebut berpotensi menurunkan kualitas analisis karena dapat menimbulkan bias dalam proses pengolahan lebih lanjut. Misalnya, atribut *Nama\_Proyek* berfungsi sebagai identitas administratif yang penting dalam menghubungkan informasi UMKM dengan jenis kegiatan usahanya, sedangkan atribut *No\_Telp* dan *Email* menjadi bagian dari data kontak

yang diperlukan untuk verifikasi maupun komunikasi resmi. Atribut *Judul\_KBLI* merepresentasikan klasifikasi usaha berdasarkan standar yang berlaku secara nasional, sehingga keberadaannya sangat krusial dalam mengelompokkan UMKM sesuai bidang usahanya. Sementara itu, *Sektor\_Pembina* memberikan informasi mengenai instansi pembina atau lembaga yang menaungi UMKM, yang relevan dalam konteks kebijakan pembinaan dan pendampingan usaha. Oleh karena itu, tahap *pre-processing* dilakukan secara cermat untuk menangani atribut yang tidak lengkap ini, baik melalui metode penghapusan data yang tidak relevan, pengisian nilai kosong dengan teknik imputasi yang sesuai, maupun transformasi data untuk memastikan konsistensi. Dengan langkah tersebut, dataset yang dihasilkan diharapkan lebih bersih, terstruktur, dan dapat digunakan secara optimal dalam tahapan analisis selanjutnya.

### Feature Selection

Feature selection merupakan salah satu metode reduksi dimensi yang digunakan untuk menyederhanakan jumlah variabel dengan tetap mempertahankan informasi penting. Reduksi dimensi sendiri merupakan proses pengurangan atribut agar data lebih efisien diolah tanpa kehilangan karakteristik utama yang dibutuhkan dalam analisis. Pada penelitian ini, feature selection dilakukan melalui dua pendekatan, yaitu analisis korelasi dan Principal Component Analysis (PCA), untuk menyaring variabel-variabel yang paling relevan terhadap pembentukan klaster. Hasil proses tersebut menunjukkan adanya sepuluh atribut yang didrop karena dinilai tidak berkontribusi signifikan terhadap analisis, yaitu: No\_Telp, Email, Nama\_Proyek, No, Nama\_User, Alamat, Nama\_Perusahaan, ID\_Proyek, NIB, dan Tanggal\_Terbit. Dengan melakukan eliminasi variabel-variabel tersebut, dataset menjadi lebih terfokus dan dapat meningkatkan kinerja algoritma dalam tahap klasterisasi maupun klasifikasi. Implementasi teknis pada Python dilakukan menggunakan perintah `df_clean = df_clean.drop (...)` untuk menghapus atribut yang tidak digunakan.

### Cleaning data

Pembersihan data (*data cleaning*) dalam penelitian ini dilakukan dengan tujuan untuk memperbaiki serta mengidentifikasi data yang salah, rusak, tidak lengkap, duplikat, maupun tidak konsisten pada dataset mentah. Tahap ini menjadi krusial karena kualitas data yang buruk dapat memengaruhi akurasi analisis, mengurangi validitas pengambilan keputusan, serta menurunkan performa model *machine learning*. Oleh karena itu, dilakukan proses normalisasi format teks, penghapusan spasi ganda, dan perbaikan kesalahan penulisan pada beberapa atribut yang dianggap penting, seperti *Sektor\_Pembina*, *Judul\_KBLI*, *Kecamatan*, *Kelurahan*, *Jenis\_Perusahaan*, *Risiko\_Proyek*, dan *Skala\_Usaha*. Proses ini membantu memastikan konsistensi data sehingga dapat diolah secara optimal pada tahap analisis berikutnya. Tabel 2 merupakan hasil dari cleaning data

Tabel 2. Hasil Cleaning data

Atribut	Sebelum Cleaning data	Sesudah Cleaning data
Sektor_Pembina	"Kementerian Energi dan Sumber DayaMineral"	"Kementerian Energi dan Sumber Daya Mineral"
Judul_KBLI	"Aktivitas KlinikSwasta"	"Aktivitas Klinik Swasta"
Kecamatan	"R asa Nae Barat"	"Rasanae Barat"
Kelurahan	"JatibaruTimur"	"Jatibaru Timur"
Jenis_Perusahaan	"Perseroan Terbatas(PT)"	"Perseroan Terbatas (PT)"
Risiko_Proyek	"Menengah Rendah"	"Menengah Rendah"
Skala_Usaha	"Usaha Mikro"	"Usaha Mikro"

### Handling missing data

Handling missing data merupakan proses penanganan terhadap nilai yang hilang dalam suatu dataset. Nilai yang hilang umumnya direpresentasikan oleh sel kosong, nilai tertentu seperti 'NA', atau kode khusus lainnya. Keberadaan data yang tidak lengkap ini dapat menimbulkan bias dalam analisis, menurunkan kualitas model, dan bahkan menyebabkan algoritma *machine learning* gagal berfungsi

dengan baik. Oleh karena itu, diperlukan strategi khusus untuk mengatasi permasalahan tersebut. Salah satu pendekatan yang umum digunakan adalah metode imputasi, yaitu teknik mengganti nilai yang hilang dengan nilai estimasi. Dengan imputasi, ukuran sampel dataset tetap dapat dipertahankan sehingga menghindari hilangnya informasi akibat penghapusan titik data.

Dalam penelitian ini, dilakukan imputasi terhadap atribut kategorikal dengan menggunakan pendekatan modus. Modus dipilih karena mampu merepresentasikan kategori yang paling sering muncul dalam suatu kolom, sehingga relatif stabil untuk menggantikan nilai yang hilang. Misalnya, pada atribut *Sektor\_Pembina* dan *Judul\_KBLI*, jika ditemukan nilai kosong, maka sistem secara otomatis menggantinya dengan kategori yang paling sering muncul pada kolom tersebut. Implementasi teknis pada Python ditunjukkan pada potongan kode berikut:

```
for col in ["Sektor_Pembina", "Judul_KBLI"]:
    if df_clean[col].isnull().sum() > 0:
        mode_val = df_clean[col].mode()[0]
        df_clean[col] = df_clean[col].fillna(mode_val)
```

Kode di atas melakukan iterasi pada kolom *Sektor\_Pembina* dan *Judul\_KBLI*, kemudian memeriksa apakah terdapat nilai kosong (*missing values*). Jika ditemukan, maka nilai tersebut akan digantikan dengan modus dari kolom yang bersangkutan. Dataset menjadi lebih bersih dan siap digunakan untuk tahap analisis berikutnya, tanpa mengurangi jumlah sampel yang ada.

### Normalisasi Data

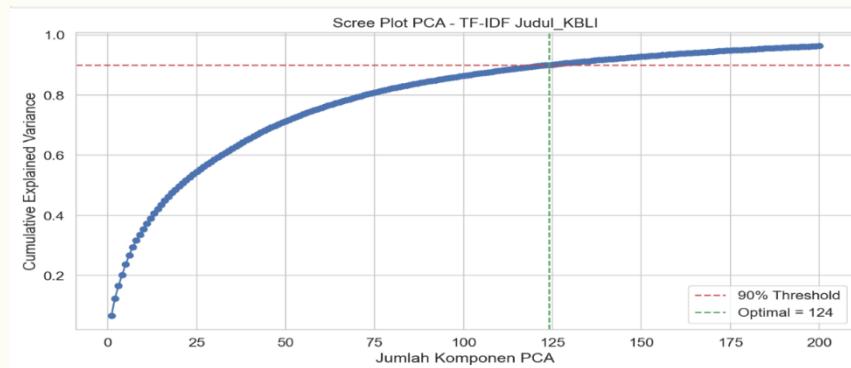
Normalisasi data merupakan tahap penting dalam proses *pre-processing* untuk memastikan bahwa setiap atribut memiliki skala nilai yang seragam. Pada penelitian ini, normalisasi dilakukan menggunakan metode Min-Max Scaling, yaitu teknik yang mengubah rentang nilai data asli ke dalam interval 0 hingga 1. Proses ini dilakukan dengan cara mengurangi setiap nilai dengan nilai minimum pada kolomnya, kemudian dibagi dengan rentang (selisih antara nilai maksimum dan minimum). Data yang semula memiliki rentang nilai berbeda misalnya jumlah karyawan dengan skala ratusan, sedangkan tingkat risiko proyek berskala 1–5—akan ditransformasikan ke dalam skala yang sebanding. Normalisasi ini penting agar algoritma *clustering* seperti FCM maupun algoritma *klasifikasi* seperti SVM tidak bias terhadap atribut yang memiliki rentang nilai lebih besar, sehingga setiap fitur dapat berkontribusi secara proporsional terhadap proses analisis. Implementasi teknis yang dilakukan pada python ditunjukkan pada potongan kode berikut ini:

```
scaler = MinMaxScaler()
df_ready[["Jumlah_Investasi", "TKI", "Proxy_Digitalisasi"]] = scaler.fit_transform(
    df_ready[["Jumlah_Investasi", "TKI", "Proxy_Digitalisasi"]]
)
```

Penerapan Min-Max Scaling, dataset yang semula memiliki perbedaan rentang nilai antar atribut kini berhasil distandardisasi ke dalam skala yang seragam, yaitu 0 hingga 1. Transformasi ini tidak hanya mempermudah proses perhitungan jarak pada algoritma klasterisasi FCM dan klasifikasi SVM, tetapi juga meningkatkan stabilitas model serta mengurangi potensi bias akibat dominasi fitur tertentu. Hasil analisis yang diperoleh dari penelitian ini diharapkan lebih representatif dan memiliki tingkat akurasi yang lebih baik.

### Principal Component Analysis (PCA)

Principal Component Analysis (PCA) pada representasi fitur teks *Judul\_KBLI* yang sebelumnya telah ditransformasikan menggunakan metode TF-IDF. Sumbu horizontal merepresentasikan jumlah komponen PCA yang digunakan, sedangkan sumbu vertikal menunjukkan nilai cumulative explained variance atau proporsi ragam data yang berhasil dijelaskan oleh komponen tersebut. Dari kurva yang terbentuk terlihat bahwa semakin banyak jumlah komponen PCA yang dipertahankan, semakin tinggi pula akumulasi informasi yang dapat dijelaskan oleh model. Hasil PCA dapat dilihat pada Gambar 2



Gambar 2. Hasil PCA

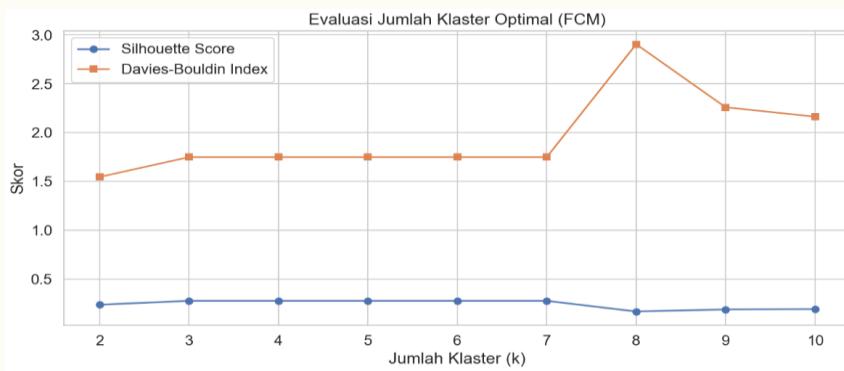
Hasil analisis menunjukkan bahwa titik ambang batas sebesar 90% *cumulative explained variance* dapat dicapai pada sekitar 124 komponen PCA, yang ditandai dengan garis putus-putus hijau pada grafik. Hal ini berarti, dari ratusan fitur yang dihasilkan oleh TF-IDF, cukup dengan mempertahankan 124 komponen utama saja sudah mampu menjelaskan lebih dari 90% variasi informasi dalam data. PCA berhasil mereduksi dimensi data tanpa menghilangkan informasi penting yang dibutuhkan dalam tahap analisis berikutnya. Pemilihan 124 komponen ini bersifat optimal karena mampu menyeimbangkan antara kompleksitas data dan tingkat representasi informasi. Jika terlalu sedikit komponen yang digunakan, maka informasi penting berpotensi hilang sehingga memengaruhi kualitas klasterisasi maupun klasifikasi. Sebaliknya, jika terlalu banyak komponen dipertahankan, maka manfaat reduksi dimensi menjadi kurang signifikan karena dataset masih tetap kompleks dan berisiko menimbulkan beban komputasi yang tinggi.

### Clustering FCM

Penentuan jumlah klaster pada penelitian ini dilakukan dengan menggunakan metode *FCM* yang terlebih dahulu memerlukan evaluasi jumlah klaster optimum. Evaluasi ini dilakukan melalui dua metrik utama, yaitu *Silhouette Index* dan *Davies-Bouldin Index*. *Silhouette Index* digunakan untuk mengukur tingkat kesesuaian data dalam suatu klaster, dengan nilai yang lebih tinggi menunjukkan bahwa objek data berada pada klaster yang tepat dan terpisah secara jelas dari klaster lainnya. Sementara itu, *Davies-Bouldin Index* digunakan untuk menilai tingkat pemisahan antar klaster, di mana nilai yang lebih rendah menandakan bahwa klaster yang terbentuk memiliki pemisahan yang lebih baik dan tidak saling tumpang tindih. Kedua metrik tersebut memberikan dasar evaluasi yang lebih objektif dalam menentukan jumlah klaster terbaik yang dapat merepresentasikan struktur data UMKM di Kota Bima. Dengan memadukan hasil pengukuran dari *Silhouette Index* dan *Davies-Bouldin Index*, proses pemilihan jumlah klaster tidak hanya bergantung pada asumsi subjektif peneliti, melainkan didukung oleh perhitungan kuantitatif yang terukur. Hal ini penting agar hasil klasterisasi dapat mencerminkan distribusi data secara akurat dan mampu digunakan sebagai landasan dalam analisis lebih lanjut, seperti klasifikasi menggunakan *SVM*. Hasil perbandingan dari kedua metrik dapat dilihat pada Tabel 3 dan Gambar 3.

Tabel 3. Hasil *Silhouette Index* dan *Davies-Bouldin Index*

Jumlah Klaster	<i>Silhouette Index</i>	<i>Davies-Bouldin Index</i>
2	0.2399	1.5469
3	0.2795	1.7494
4	0.2795	1.7494
5	0.2795	1.7494
6	0.2795	1.7494
7	0.2795	1.7494
8	0.1702	2.9050
9	0.1915	2.2596
10	0.1949	2.1638



Gambar 3. Hasil Evaluasi Jumlah Klaster Optimal

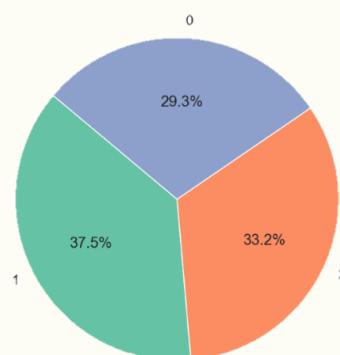
Berdasarkan hasil evaluasi pada Gambar 3, terlihat bahwa nilai *Silhouette Index* tertinggi berada pada kisaran 0.2795 ketika jumlah klaster berada pada  $k = 3$  hingga  $k = 7$ . Hal ini menunjukkan bahwa kualitas pengelompokan relatif stabil dan lebih baik dibandingkan jumlah klaster lainnya. Sebaliknya, ketika jumlah klaster ditingkatkan menjadi  $k = 8$ , nilai *Silhouette Index* turun drastis menjadi 0.1702, yang menandakan penurunan kualitas klasterisasi secara signifikan. Sedangkan, *Davies-Bouldin Index* nilai terendah dicapai pada  $k = 2$  dengan skor 1.5469, menunjukkan bahwa pemisahan antar klaster cukup baik pada kondisi ini. Namun, nilai DBI kemudian meningkat menjadi 1.7494 pada  $k = 3$  hingga  $k = 7$ , menandakan bahwa meskipun kualitas pengelompokan masih cukup stabil, pemisahan antar klaster relatif lebih lemah dibandingkan  $k = 2$ . Ketika jumlah klaster meningkat menjadi  $k = 8$ , DBI naik tajam hingga 2.9050, yang mengindikasikan pemisahan antar klaster memburuk. Pada  $k = 9$  dan  $k = 10$ , nilai DBI memang sedikit menurun (masing-masing 2.2596 dan 2.1638), tetapi masih lebih buruk dibandingkan dengan kondisi  $k \leq 7$ . Secara keseluruhan, hasil ini menunjukkan bahwa jumlah klaster optimal berada pada  $k = 3$  hingga  $k = 4$ . Pada jumlah klaster ini, nilai *Silhouette Index* mencapai titik maksimum (0.2795) dengan stabilitas yang baik, sementara *Davies-Bouldin Index* masih berada pada tingkat yang relatif moderat (1.7494). Oleh karena itu, penelitian ini menggunakan nilai  $k=3$  untuk pengelompokan klaster UMKM Bima.

Tabel 4. Hasil Analisis Karakteristik Klaster

	Klaster 0	Klaster 1	Klaster 2
<b>Label Kategorikal</b>	UMKM Kuliner Tradisional Berbasis Lokal	UMKM Infrastruktur & Energi Skala Menengah	UMKM Perdagangan Modern & Digital
<b>Jumlah anggota</b>	1518	1940	1718
<b>Investasi rata-rata</b>	0.000117	0.001263	0.000509
<b>TKI rata-rata</b>	0.004446	0.011480	0.007599
<b>Digitalisasi rata-rata</b>	0.009442	0.021993	
<b>Sektor Dominan</b>	Kementerian Perindustrian, Badan Pengendalian Lingkungan Hidup, Bank Indonesia	Kementerian Pekerjaan Umum dan Perumahan Rakyat, Kementerian Pariwisata, Kementerian Pertanian	Kementerian Perdagangan, Badan Pengendalian Lingkungan Hidup, Bank Indonesia

<b>Profil Karakteristik</b>	UMKM skala kecil dengan investasi dan tenaga kerja rendah, tingkat digitalisasi sangat minim. Didominasi sektor manufaktur tradisional.	UMKM skala kecil dengan investasi dan tenaga kerja rendah, tingkat digitalisasi sangat minim. Didominasi sektor manufaktur tradisional.	UMKM dengan adopsi digitalisasi yang sangat tinggi, namun kapasitas investasi dan tenaga kerja sedang. Mayoritas bergerak di sektor perdagangan
-----------------------------	---	---	---

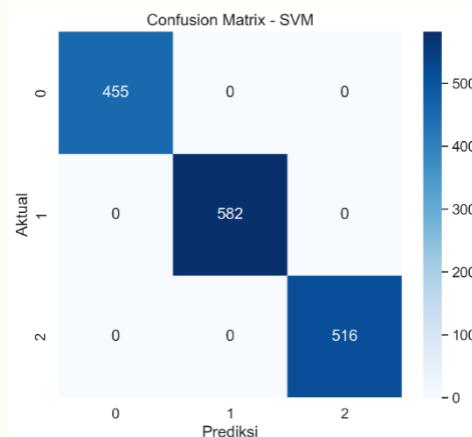
Berdasarkan hasil pengelompokan menggunakan algoritma FCM dengan jumlah klaster optimum tiga klaster ( $k=3$ ), diperoleh ringkasan karakteristik masing-masing klaster yang ditunjukkan pada Tabel 4. Analisis berikut dilakukan dengan membandingkan jumlah anggota, rata-rata investasi, rata-rata tenaga kerja Indonesia (TKI), tingkat digitalisasi, serta sektor dominan pada masing-masing klaster. Klaster 0 beranggotakan 1.518 UMKM yang didominasi oleh Kementerian Perindustrian. Klaster ini memiliki karakteristik nilai investasi relatif kecil dengan rata-rata 0.000117, tenaga kerja rata-rata 0.004446, serta tingkat digitalisasi yang rendah (0.009442). Hal ini menunjukkan bahwa kelompok UMKM pada klaster ini lebih cenderung berada pada sektor manufaktur dengan skala usaha kecil dan keterbatasan pemanfaatan teknologi digital. Klaster 1 terdiri dari 1.940 UMKM, dengan sektor dominan berasal dari Kementerian Pekerjaan Umum dan Perumahan Rakyat (PUPR). Karakteristik klaster ini menampilkan rata-rata investasi tertinggi (0.001263) dan tenaga kerja yang juga relatif lebih besar (0.011480) dibandingkan klaster lainnya. Namun, tingkat digitalisasi masih cukup rendah (0.021993). Hal ini mengindikasikan bahwa klaster ini mewakili UMKM yang bergerak pada sektor konstruksi atau jasa pendukung infrastruktur dengan modal usaha lebih besar dan jumlah tenaga kerja yang lebih tinggi, tetapi pemanfaatan teknologi digital belum optimal. Klaster 2 berisi 1.718 UMKM yang mayoritas berada di bawah pembinaan Kementerian Perdagangan. Karakteristik paling menonjol dari klaster ini adalah tingkat digitalisasi yang jauh lebih tinggi dibandingkan klaster lainnya (0.751261), meskipun rata-rata investasi (0.000509) dan tenaga kerja (0.007599) relatif sedang. Hal ini menunjukkan bahwa UMKM dalam klaster ini cenderung mengadopsi teknologi digital, terutama dalam aktivitas perdagangan dan distribusi barang, meskipun kapasitas modal dan jumlah tenaga kerja tidak sebesar klaster 1. Secara keseluruhan, hasil pengelompokan ini berhasil mengungkap tiga profil utama UMKM di Kota Bima, yaitu: Klaster Industri Tradisional (Klaster 0) berfokus pada sektor manufaktur, skala usaha kecil, dan tingkat digitalisasi rendah, Klaster Infrastruktur dan Konstruksi (Klaster 1) memiliki modal dan tenaga kerja lebih besar, berfokus pada sektor pembangunan, namun digitalisasi masih terbatas dan Klaster Perdagangan Digital (Klaster 2) dicirikan oleh tingkat digitalisasi tinggi, meskipun kapasitas investasi dan tenaga kerja relatif sedang. Temuan ini penting karena memberikan dasar bagi penyusunan kebijakan pembinaan UMKM yang lebih terarah, di mana setiap klaster membutuhkan strategi pengembangan yang berbeda, khususnya dalam peningkatan kapasitas digitalisasi, dukungan investasi dan penguatan sumber daya manusia.

Distribusi UMKM Berdasarkan Klaster (FCM,  $k=3$ )Gambar 4. Hasil Distribusi UMKM berdasarkan Klaster (FCM,  $k=3$ )

Berdasarkan Hasil klasterisasi UMKM Kota Bima pada Gambar 4 menggunakan metode FCM dengan jumlah klaster optimal  $k = 3$  menunjukkan distribusi yang relatif seimbang antar kelompok. Pada Gambar yang ditampilkan, Klaster 0 merepresentasikan sebesar 29,3% dari total UMKM. Kelompok ini dicirikan dengan tingkat investasi yang rendah, jumlah tenaga kerja terbatas, serta adopsi digitalisasi yang masih minim. Karakteristik ini menunjukkan bahwa sebagian UMKM di Kota Bima masih berada pada tahap tradisional, dengan ketergantungan pada pola usaha konvensional tanpa pemanfaatan teknologi secara signifikan. Klaster 1 merupakan kelompok dengan proporsi terbesar, yaitu 37,5% dari keseluruhan UMKM. Klaster ini didominasi oleh sektor infrastruktur dan konstruksi, dengan rata-rata nilai investasi dan jumlah tenaga kerja lebih tinggi dibandingkan dengan klaster lainnya. Distribusi tersebut mengindikasikan bahwa sektor pembangunan fisik menjadi tulang punggung utama aktivitas UMKM di Kota Bima, sekaligus memberikan kontribusi penting terhadap dinamika ekonomi lokal. Keberadaan kelompok ini juga menunjukkan adanya kecenderungan UMKM yang berkembang melalui dukungan pada sektor makro seperti perumahan, jalan, dan fasilitas publik. Klaster 2 memiliki porsi sebesar 33,2% dan dicirikan oleh tingkat digitalisasi yang jauh lebih tinggi meskipun skala investasi dan tenaga kerja berada pada kategori menengah. Kelompok ini banyak terafiliasi dengan sektor perdagangan, sehingga adopsi teknologi digital menjadi salah satu faktor pendorong utama dalam aktivitas bisnisnya. Temuan ini menunjukkan adanya transformasi menuju ekosistem UMKM digital yang semakin kuat, sejalan dengan perkembangan teknologi informasi dan tuntutan pasar yang semakin modern. Distribusi ketiga klaster ini memberikan gambaran bahwa struktur UMKM di Kota Bima tersegmentasi ke dalam tiga tipologi utama, yaitu UMKM Kuliner Tradisional Berbasis Lokal, UMKM Perdagangan Modern dan Digital dan UMKM Perdagangan Modern dan Digital. Kondisi ini mencerminkan adanya diversifikasi pola usaha yang tidak hanya bertumpu pada sektor konvensional, tetapi juga mulai mengarah pada pemanfaatan teknologi digital. Hasil klasterisasi ini dapat menjadi dasar penting bagi pemerintah daerah maupun pemangku kepentingan dalam merumuskan strategi pengembangan dan pembinaan UMKM sesuai dengan karakteristik tiap klaster.

## SVM

Pada penelitian ini, SVM digunakan untuk mengklasifikasikan UMKM Kota Bima berdasarkan hasil pseudo-label yang diperoleh dari algoritma FCM. Proses ini memungkinkan setiap UMKM yang telah dikelompokkan ke dalam klaster melalui FCM untuk dijadikan data latih dalam membangun model klasifikasi. Dengan pendekatan ini, SVM berperan penting dalam menguji sejauh mana pembagian klaster yang dilakukan secara tidak terawasi (*unsupervised*) dapat dikonfirmasi kembali melalui metode klasifikasi terawasi (*supervised*). Strategi ini tidak hanya meningkatkan validitas hasil klasterisasi, tetapi juga memberikan dasar yang lebih kuat dalam memahami perbedaan karakteristik antar UMKM di tiap klaster. Fitur yang digunakan, yaitu "Jumlah\_Investasi", "TKI", dan "Proxy\_Digitalisasi", terbukti cukup representatif dalam membedakan karakteristik antar klaster UMKM. Variabel jumlah investasi menggambarkan kekuatan modal, jumlah tenaga kerja (TKI) mencerminkan kapasitas usaha, sedangkan proxy digitalisasi menunjukkan tingkat pemanfaatan teknologi dalam operasional UMKM. Kombinasi ketiga variabel ini memberikan pemisahan yang jelas antar klaster, sehingga memudahkan SVM untuk menemukan hyperplane optimal yang memisahkan data. Proses pelatihan model dilakukan dengan membagi dataset menjadi 70% data training dan 30% data testing. Pembagian ini memungkinkan model memperoleh cukup informasi dari data pelatihan sekaligus menjaga kemampuan generalisasi terhadap data uji. Selain itu, penggunaan GridSearchCV dengan 5-fold cross validation membantu menemukan parameter terbaik. Berdasarkan hasil pencarian, model memperoleh parameter optimal pada kombinasi nilai C, jenis kernel (linear atau RBF), dan parameter gamma. Parameter ini sangat berpengaruh dalam membentuk margin optimal pada ruang fitur. Hasil evaluasi model SVM dapat dilihat pada Gambar 4.



Gambar 4. Confusion Matrix SVM

Berdasarkan Gambar 5 model SVM yang dibangun menggunakan fitur Jumlah Investasi, TKI, dan Proxy Digitalisasi menunjukkan performa yang sangat baik dengan akurasi mencapai 100%. Nilai precision, recall, dan f1-score pada seluruh klaster juga konsisten sempurna, yang mengindikasikan bahwa model mampu mengklasifikasikan setiap UMKM secara tepat. Hasil ini menegaskan bahwa variabel yang dipilih sangat representatif dalam membedakan karakteristik antar klaster, sekaligus memperkuat validitas hasil pseudo-label yang dihasilkan dari FCM. Pendekatan kombinasi FCM dan SVM ini memberikan kontribusi metodologis yang signifikan. FCM berfungsi sebagai mekanisme untuk menemukan pola alami dalam data UMKM tanpa memerlukan label awal, sedangkan SVM memanfaatkan hasil pseudo-label tersebut untuk membangun model klasifikasi yang terverifikasi.

## KESIMPULAN

Berdasarkan hasil evaluasi, model SVM yang dibangun menggunakan fitur Jumlah Investasi, TKI, dan Proxy Digitalisasi menunjukkan performa yang sangat baik dengan akurasi mencapai 100%. Nilai precision, recall, dan f1-score pada seluruh klaster juga konsisten sempurna, yang mengindikasikan bahwa model mampu mengklasifikasikan setiap UMKM secara tepat tanpa adanya kesalahan prediksi. Hasil ini menegaskan bahwa variabel yang dipilih sangat representatif dalam membedakan karakteristik antar klaster, sekaligus memperkuat validitas hasil pseudo-label yang dihasilkan dari FCM. Selain itu, pendekatan kombinasi FCM dan SVM ini memberikan kontribusi metodologis yang signifikan. FCM berfungsi sebagai mekanisme untuk menemukan pola alami dalam data UMKM tanpa memerlukan label awal, sedangkan SVM memanfaatkan hasil pseudo-label tersebut untuk membangun model klasifikasi yang terverifikasi. Hasil penelitian ini membuktikan bahwa strategi *hybrid unsupervised-supervised learning* efektif dalam menganalisis dan memvalidasi struktur data UMKM.

## UCAPAN TERIMAKASIH

Penulis mengucapkan terimakasih yang sebesar-besarnya kepada semua pihak yang telah membantu dalam penulisan artikel ilmiah ini. Penulis juga mengucapkan terimakasih kepada Direktorat Penelitian dan Pengabdian kepada Masyarakat (DPPM) di bawah Kementerian Pendidikan Tinggi, Sains, dan Teknologi (Kemdiktisaintek) yang telah mendanai penelitian ini melalui skema Penelitian Dosen Pemula (PDP) dengan nomor kontrak 129/C3/DT.05.00/PL/2025; 2166/LL8/AL.04/2025; 024/KEP/8.IV/C/VI/2025.

**DAFTAR PUSTAKA**

- Agustiono, W., & Darmawan, A. K. (2023). E-Marketplace Selection for Indonesian MSMEs: A Sentiment Analysis Approach Combining Lexicon-Based and Support Vector Machine. *Proceeding - IEEE 9th Information Technology International Seminar, ITIS 2023*. <https://doi.org/10.1109/ITIS59651.2023.10420382>
- Ashari, I. F., Nugroho, E. D., Baraku, R., Yanda, I. N., & Liwardana, R. (2023). Analysis Of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, And Rand-Index Evaluation On K-Means Algorithm For Classifying Flood-Affected Areas In Jakarta. *Journal of Applied Informatics and Computing*, 7(1), 89–97. <https://doi.org/10.30871/jaic.v7i1.4947>
- Beynon, M., Pickernell, D., & Jones, P. (2025). Entrepreneurial Activity In The International Trade In Cultural Goods: A Fuzzy Clustering Analysis. *Technological Forecasting and Social Change*, 210(1), 123914. <https://doi.org/10.1016/j.techfore.2024.123914>
- Bhavani, A., & Santhosh Kumar, B. (2021). A Review of State Art of Text Classification Algorithms. *International Conference on Computing Methodologies and Communication*, 1493-1499. <https://doi.org/10.1109/ICCMC51019.2021.9418262>
- Boubaker, S., Le, T. D. Q., Ngo, T., & Manita, R. (2025). Predicting The Performance Of Msmses: A Hybrid DEA-Machine Learning Approach. *Annals of Operations Research*, 350(2), 555–577. <https://doi.org/10.1007/s10479-023-05230-8>
- Chiu, M. C., Huang, Y. J., & Wei, C. J. (2024). Enhancing Smes Digital Transformation Through Machine Learning: A Framework For Adaptive Quality Prediction. *Journal of Industrial Information Integration*, 41(1), 100666. <https://doi.org/10.1016/j.jii.2024.100666>
- Damayanti, D. K. D., & Jakfar, M. (2023). Klasifikasi Status Stunting Balita Menggunakan Algoritma Fuzzy C-Means (Studi Kasus Posyandu Rw 01 Kelurahan Jepara Surabaya). *MATHunesa: Jurnal Ilmiah Matematika*, 11(3), 533–542. <https://doi.org/10.26740/mathunesa.v11n03.p524-533>
- Gafar, G. A., & Wang, G. (2023). Implementation Of Machine Learning For Sharia Financing Scoring In Indonesian MSME Sectors. *Journal Of Social Science Research*, 3(5), 814–830.
- Hafsari, R., & Nurcahyo, G. W. (2021). Penerapan Fuzzy C-Means Pada Teknologi Adopsi Usaha Mikro Kecil dan Menengah. *Jurnal KomtekInfo*, 8(3), 183–188. <https://doi.org/10.35134/komtekinfo.v8i3.165>
- Herdiyeni, Y., Juanda, B., Zhafira, N., Anggraeni, L., & Probokawuryan, M. (2025). Machine-Learning-Based Assessment and Prediction of Digital Transformation Performance for MSMEs in Indonesia. *International Journal of Information Technology and Decision Making*, 1-29. <https://doi.org/10.1142/S0219622025500415>
- Jonathan, J. K., Herwindati, D. E., Ferdinand, K., & Jong, F. (2024). Implementation of Fuzzy C-Means for Clustering MSMEs in Jambi Province. *International Conference on Informatics and Computing*, 1-16. <https://doi.org/10.1109/ICIC64337.2024.10956968>
- Kumar, P., Sumithra, M. G., & Saranya, N. (2021). Particle Swarm Optimization (PSO) With Fuzzy C Means (PSO-FCM)-Based Segmentation And Machine Learning Classifier For Leaf Diseases Prediction. *Concurrency and Computation: Practice and Experience*, 33(3). <https://doi.org/10.1002/cpe.5312>
- Maharani, S., Noerlina, & Nadira. (2024). A Study of MSMEs Business Sustainability Strategies: Transforming Towards Industry. *International Conference on Informatics, Multimedia, Cyber and Information System*, 449–453. <https://doi.org/10.1109/ICIMCIS63449.2024.10957198>
- Mishra, A., Tiwari, M., Mishra, J., & Hung, B. T. (2023). Brain Tumor Segmentation of MR Images Using SVM and Fuzzy Classifier in Machine Learning. *Springer Nature Link*, 81–112. [https://doi.org/10.1007/978-981-99-1949-9\\_5](https://doi.org/10.1007/978-981-99-1949-9_5)

- Muthiah, H., & Hamidah, N. K. (2025). Integrasi Machine Learning Untuk Optimalisasi Prediksi Permintaan Produk Pada UMKM Kuliner. *Jurnal PenKoMi : Kajian Pendidikan Dan Ekonomi*, 8(1), 229–233. <https://doi.org/10.33627/pk.v8i1.2945>
- Nararto, E. K., Budiwati, S. D., & Sari, S. K. (2024). Opinion Classification on MSME Social Media Comments using Support Vector Machine and Random Forest Models. *International Conference of Science and Information Technology in Smart Administration*, 165–170. <https://doi.org/10.1109/ICSINTESA62455.2024.10748163>
- Ouyang, Y., & Sun, Y. (2024). Exploration and Trend Analysis of Corporate Financial Statement Data: A Fuzzy Clustering Algorithm. *ACM International Conference Proceeding Series*, 230–236. <https://doi.org/10.1145/3662739.3672332>
- Patil, K. P., Pramod, D., & Bharathi, S. V. (2023). MSMEs Readiness for Adopting Artificial Intelligence and Machine Learning. *International Conference on Advanced Computing Technologies and Applications*, 1-6. <https://doi.org/10.1109/ICACTA58201.2023.10392396>
- Putri, R. A., Rustam, Z., Pandelaki, J., & Salmi, N. (2021). Kernel Based Comparison between Fuzzy C-Means and Support Vector Machine for Sinusitis Classification. *Journal of Physics: Conference Series*, 1752(1). <https://doi.org/10.1088/1742-6596/1752/1/012031>
- Putri, R. A., Mauritsius, T., Ananda, B., & Jayadi, R. (2024). Optimizing Customer Retention in Urban Laundry Services: A Comparative Analysis of Machine Learning Algorithms in Jakarta's MSME Sector. *International Conference on Information Management and Technology, ICIMTech 2024*, 630–635. <https://doi.org/10.1109/ICIMTech63123.2024.10780922>
- Sabale, D. B., Kaswan, M. S., & Rathi, R. (2024). Development Of Circular Economy Model Using Industry 4.0's Machine Learning Techniques To Improve Performance Metrics Of MSMEs. *International Journal on Interactive Design and Manufacturing*, 1-19. <https://doi.org/10.1007/s12008-024-01782-6>
- Sheth, V., Tripathi, U., & Sharma, A. (2022). A Comparative Analysis of Machine Learning Algorithms for Classification Purpose. *Procedia Computer Science*, 215, 422–431. <https://doi.org/10.1016/j.procs.2022.12.044>
- Tahyudin, I., Firmansyah, G., Ivansyah, A. G., Ma'arifah, W., & Lestari, L. (2022). Comparison of K-Means Algorithms and Fuzzy C-Means Algorithms for Clustering Customers Dataset. *APICS 2022 - 2022 1st International Conference on Smart Technology, Applied Informatics, and Engineering, Proceedings*, 211–216. <https://doi.org/10.1109/APICS56469.2022.9918726>
- Vásquez, J., Aguirre, S., Puertas, E., Bruno, G., Priarone, P. C., & Settineri, L. (2021). A Sustainability Maturity Model For Micro, Small And Medium-Sized Enterprises (Msmes) Based On A Data Analytics Evaluation Approach. *Journal of Cleaner Production*, 311, 127692. <https://doi.org/10.1016/j.jclepro.2021.127692>
- Wibowo, R. A., Budiwati, S. D., & Sari, S. K. (2024). Analyzing Instagram User Sentiment Toward MSME Using Naive Bayes and Logistic Regression. *2024 International Conference on Data Science and Its Applications, ICoDSA 2024*, 561–566. <https://doi.org/10.1109/ICoDSA62899.2024.10651957>