
PREDIKSI JUMLAH UMKM BERDASARKAN KATEGORI USAHA DAN LOKASI KABUPATEN/KOTA DI PROVINSI JAWA BARAT MENGGUNAKAN *DECISION TREE*

Oleh:

Rizki Fahrizal¹

Muhammad Nur Falah²

Elkin Rilvani³

Universitas Pelita Bangsa

Alamat: JL. Inspeksi Kalimalang No.9, Cibatu, Cikarang Sel., Kabupaten Bekasi, Jawa Barat (17530).

Korespondensi Penulis: rizkifahrizal777@gmail.com,
muhammadnurfalah49@gmail.com, elkin.rilvani@pelitabangsa.ac.id.

Abstract. West Java Province is one of the regions in Indonesia with the highest number of Micro, Small, and Medium Enterprises (MSMEs). MSMEs play a significant role in driving regional economic growth, creating job opportunities, and improving community welfare. This study aims to predict the number of MSMEs based on business categories and administrative regions (regencies/cities) in West Java. The data used in this research are projections of the number of MSMEs from 2017 to 2024, categorized by types of businesses such as culinary, trade, garment, services, and others. The Decision Tree algorithm was chosen due to its ability to generate predictive models that are easy to interpret and capable of identifying the most influential attributes affecting the fluctuation in MSME numbers. The analysis was conducted using RapidMiner software to build and evaluate the prediction model. The results show varying growth trends of MSMEs across different regions and business categories. These findings are expected to serve as a strategic reference for local governments in formulating more targeted and data-driven MSME development policies based on local potential.

Keywords: MSMEs, West Java, Decision Tree, Data Mining, RapidMiner.

PREDIKSI JUMLAH UMKM BERDASARKAN KATEGORI USAHA DAN LOKASI KABUPATEN/KOTA DI PROVINSI JAWA BARAT MENGGUNAKAN *DECISION TREE*

Abstrak. Provinsi Jawa Barat merupakan salah satu wilayah dengan jumlah Usaha Mikro, Kecil, dan Menengah (UMKM) terbanyak di Indonesia. UMKM memegang peranan penting dalam mendorong pertumbuhan ekonomi daerah, menciptakan lapangan kerja, serta meningkatkan kesejahteraan masyarakat. Penelitian ini bertujuan untuk memprediksi jumlah UMKM berdasarkan kategori usaha dan wilayah administratif kabupaten atau kota di Jawa Barat. Data yang digunakan dalam penelitian ini merupakan data proyeksi jumlah UMKM dari tahun 2017 hingga 2024 yang telah dikelompokkan berdasarkan jenis usaha seperti kuliner, perdagangan, konveksi, jasa, dan lainnya. Metode yang digunakan adalah algoritma *Decision Tree* karena memiliki kemampuan untuk menghasilkan model prediksi yang mudah dipahami dan mampu mengidentifikasi atribut paling berpengaruh terhadap fluktuasi jumlah UMKM. Proses analisis dilakukan dengan menggunakan perangkat lunak RapidMiner untuk membangun dan mengevaluasi model prediksi. Hasil penelitian menunjukkan adanya variasi pertumbuhan UMKM di setiap kabupaten atau kota dan dalam masing-masing kategori usaha. Temuan ini diharapkan menjadi acuan bagi pemerintah daerah dalam menyusun strategi pengembangan UMKM yang lebih tepat sasaran dan berbasis potensi lokal.

Kata Kunci: UMKM, Jawa Barat, *Decision Tree*, Data Mining, RapidMiner.

LATAR BELAKANG

Usaha Mikro, Kecil, dan Menengah (UMKM) memegang peran strategis dalam pembangunan ekonomi di Indonesia, terutama dalam menciptakan lapangan kerja, mendorong pemerataan pendapatan, serta memperkuat struktur ekonomi lokal. Di Provinsi Jawa Barat, sebagai wilayah dengan jumlah penduduk tertinggi di Indonesia, keberadaan UMKM menjadi tulang punggung dalam menggerakkan roda perekonomian masyarakat. Berdasarkan data proyeksi dari Dinas Koperasi dan Usaha Kecil Provinsi Jawa Barat, jumlah UMKM menunjukkan tren pertumbuhan yang fluktuatif dari tahun ke tahun, baik dalam skala mikro, kecil, maupun menengah. Namun, kompleksitas permasalahan yang dihadapi oleh pelaku UMKM, seperti keterbatasan akses terhadap permodalan, pendampingan usaha, dan kebijakan yang kurang tepat sasaran, masih menjadi tantangan utama. Oleh karena itu, diperlukan suatu pendekatan berbasis data untuk membantu proses pengambilan keputusan, terutama dalam merumuskan strategi pengembangan UMKM yang berbasis wilayah dan kategori usaha.

Dalam upaya tersebut, pemanfaatan teknik data mining menjadi sangat relevan, khususnya dalam membangun model prediktif yang dapat mengungkap pola-pola tersembunyi dalam data dan memberikan gambaran proyeksi secara kuantitatif. Salah satu metode yang populer dan mudah diinterpretasikan dalam data mining adalah algoritma Decision Tree, yang mampu memetakan pengaruh berbagai atribut terhadap suatu variabel target dalam bentuk struktur pohon keputusan. Penelitian ini bertujuan untuk menerapkan metode *Decision Tree* dalam memodelkan prediksi jumlah UMKM berdasarkan data proyeksi tahun 2017 hingga 2024 di Provinsi Jawa Barat. Penelitian ini tidak hanya berupaya menjawab bagaimana tren pertumbuhan UMKM selama periode tersebut, tetapi juga mengeksplorasi faktor-faktor dominan yang memengaruhi klasifikasi jumlah UMKM di setiap kabupaten/kota, seperti kategori usaha, pertumbuhan penduduk, atau parameter lainnya jika tersedia.

Hasil dari penelitian ini diharapkan dapat memberikan kontribusi nyata dalam mendukung proses perumusan kebijakan yang lebih tepat sasaran, berbasis data, dan sesuai dengan potensi serta karakteristik ekonomi lokal di wilayah Jawa Barat. Selain itu, penelitian ini juga diharapkan dapat memberikan manfaat bagi akademisi dalam pengembangan studi kasus penerapan data mining di bidang ekonomi regional, serta menjadi acuan bagi pelaku UMKM dan pemerintah daerah dalam memahami sebaran dan potensi pengembangan usaha secara spasial dan temporal. Ruang lingkup penelitian dibatasi pada data resmi jumlah dan kategori UMKM per kabupaten/kota di Provinsi Jawa Barat, yang bersumber dari Dinas Koperasi dan UKM, serta analisis yang dilakukan hanya menggunakan algoritma *Decision Tree* tanpa perbandingan dengan algoritma lain.

KAJIAN TEORITIS

Usaha Mikro, Kecil, dan Menengah (UMKM)

UMKM (Usaha Mikro, Kecil, dan Menengah) merupakan sektor strategis dalam pembangunan ekonomi nasional dan daerah. Menurut Kementerian Koperasi dan UKM (2022), UMKM menyumbang lebih dari 60% terhadap Produk Domestik Bruto (PDB) dan menyerap lebih dari 97% tenaga kerja di Indonesia. Di Jawa Barat, UMKM tersebar hampir di seluruh kabupaten/kota, berperan besar dalam mendorong pertumbuhan ekonomi lokal, terutama pada sektor kuliner, konveksi, perdagangan, dan jasa. Dengan jumlah UMKM yang terus bertambah setiap tahunnya, analisis pola dan tren pertumbuhan

PREDIKSI JUMLAH UMKM BERDASARKAN KATEGORI USAHA DAN LOKASI KABUPATEN/KOTA DI PROVINSI JAWA BARAT MENGGUNAKAN *DECISION TREE*

UMKM menjadi penting agar pemerintah daerah dapat merancang strategi pengembangan yang lebih efektif dan tepat sasaran (Rahayu & Wijaya, 2023).

Data Mining

Data mining adalah proses menemukan pola dan informasi tersembunyi dari kumpulan data besar yang tidak terstruktur maupun terstruktur (Han, Pei, & Kamber, 2022). Dalam konteks pengembangan UMKM, data mining dapat membantu mengidentifikasi faktor-faktor yang memengaruhi pertumbuhan usaha, segmentasi pasar, hingga prediksi jumlah unit usaha berdasarkan lokasi dan kategori. Proses data mining meliputi beberapa tahapan penting: seleksi data, preprocessing, transformasi, pemodelan, dan evaluasi. Teknik ini sangat sesuai untuk menangani data UMKM yang bersifat historis dan numerik. Seiring dengan digitalisasi data oleh pemerintah daerah, penerapan data mining menjadi lebih relevan sebagai dasar pengambilan keputusan berbasis bukti (Wibowo & Handayani, 2021).

Algoritma *Decision Tree*

Decision Tree merupakan salah satu algoritma klasifikasi yang umum digunakan dalam data mining karena sifatnya yang mudah diinterpretasikan dan mampu menangani berbagai tipe data (numerik dan kategorikal). Algoritma ini bekerja dengan membagi data menjadi cabang-cabang berdasarkan nilai atribut, hingga menghasilkan suatu model pohon keputusan yang menggambarkan pola klasifikasi. Menurut Ramadhan et al. (2021), *Decision Tree* cocok diterapkan dalam prediksi keberhasilan atau klasifikasi pertumbuhan UMKM karena model ini dapat menampilkan variabel-variabel yang paling berpengaruh terhadap hasil prediksi. Selain itu, metode ini terbukti memiliki performa akurasi yang baik dan efisien dalam proses komputasi, terutama ketika digunakan pada data statistik ekonomi berskala daerah.

Penggunaan RapidMiner dalam Data Mining

RapidMiner merupakan platform perangkat lunak berbasis GUI (Graphical User Interface) yang mendukung berbagai teknik analisis data, termasuk klasifikasi, clustering, dan prediksi. Dengan kemudahan drag-and-drop, RapidMiner banyak digunakan dalam dunia akademik dan bisnis untuk membangun model prediksi tanpa perlu menulis kode

secara manual. Dalam konteks penelitian ini, RapidMiner digunakan untuk membangun model prediksi pertumbuhan UMKM berdasarkan algoritma Decision Tree, sekaligus melakukan preprocessing data, validasi silang (cross validation), dan visualisasi hasil klasifikasi. Menurut Nugroho & Kurniawan (2022), penggunaan RapidMiner mempercepat proses eksplorasi dan evaluasi model prediktif, serta meminimalkan kesalahan teknis dalam pengolahan data. Aplikasi ini juga menyediakan fitur otomatisasi pengujian performa model, seperti pengukuran akurasi, precision, recall, dan confusion matrix.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining untuk melakukan prediksi jumlah UMKM di Provinsi Jawa Barat berdasarkan kategori usaha dan wilayah kabupaten/kota. Metode ini digunakan karena mampu mengolah data historis dalam jumlah besar dan menemukan pola tersembunyi yang relevan untuk pengambilan keputusan. Teknik *Decision Tree* digunakan sebagai algoritma utama karena kemampuannya dalam membangun model prediktif yang mudah dipahami dan divisualisasikan secara hierarkis.

Menurut Han et al. (2022), data mining merupakan proses penting dalam penemuan pengetahuan dari kumpulan data besar secara sistematis. Sementara itu, algoritma *Decision Tree* dapat menghasilkan model klasifikasi yang kuat dan efektif terutama dalam konteks data yang terstruktur seperti data UMKM provinsi.

Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari portal resmi Open Data Provinsi Jawa Barat (opendata.jabarprov.go.id).

Dataset mencakup:

1. Tahun: 2017–2024
2. Wilayah: Kabupaten/Kota di Provinsi Jawa Barat
3. Jenis UMKM: Mikro, Kecil, dan Menengah
4. Kategori Usaha: Kuliner, jasa, konveksi, fashion, perdagangan, dan lainnya
5. Jumlah UMKM per kategori dan wilayah

PREDIKSI JUMLAH UMKM BERDASARKAN KATEGORI USAHA DAN LOKASI KABUPATEN/KOTA DI PROVINSI JAWA BARAT MENGGUNAKAN *DECISION TREE*

Data tersebut diunduh dalam format .CSV dan diproses menggunakan perangkat lunak RapidMiner Studio, sebuah platform visual data science yang mendukung proses machine learning tanpa perlu pemrograman kompleks (Kotu & Deshpande, 2022).

Tahapan Penelitian

Tahapan penelitian dalam studi ini dilakukan secara sistematis dengan pendekatan data mining, dimulai dari pengumpulan dan praproses data, pembangunan model prediksi menggunakan algoritma Decision Tree, hingga evaluasi performa model. Seluruh proses dilakukan menggunakan aplikasi RapidMiner, yang memudahkan integrasi alur kerja secara visual dan efisien.

1. Praproses Data

Tahap awal yang sangat penting dalam penelitian *data mining* adalah pembersihan dan persiapan data agar sesuai untuk dianalisis.

Adapun langkah-langkah praproses yang dilakukan meliputi:

- a. Pemeriksaan Konsistensi Data: Dataset diperiksa untuk memastikan tidak ada nilai yang kosong (*missing value*) maupun nilai yang tidak konsisten, misalnya nama wilayah yang tidak seragam.
- b. Pembersihan Data: Menghapus duplikasi, baris kosong, serta data yang mengandung karakter atau simbol yang tidak sesuai standar input.
- c. Transformasi Atribut: Menstandarisasi nama kolom dan format penulisan data agar sesuai dengan skema RapidMiner, seperti mengubah teks menjadi kategori atau angka (*label encoding*) bila diperlukan.
- d. Seleksi Atribut: Atribut seperti "wilayah kabupaten/kota", "kategori usaha", dan "tahun" di atur sebagai nominal, sementara "proyeksi jumlah UMKM" di atur sebagai label (target) dan bertipe numerical.
- e. Split Data: Dataset dibagi menjadi 70% untuk pelatihan (training) dan 30% untuk pengujian (*testing*) model, dengan metode random sampling.

Langkah ini memastikan data yang digunakan bersih dan siap diproses tanpa menimbulkan bias pada proses pembelajaran algoritma.

2. Pembangunan Model *Decision Tree*

Setelah data siap, tahap selanjutnya adalah pembangunan model klasifikasi menggunakan algoritma Decision Tree. *Decision Tree* dipilih karena keunggulannya

dalam mengklasifikasikan data dengan menghasilkan model yang dapat divisualisasikan sebagai pohon keputusan, sehingga mudah dipahami bahkan oleh pemangku kebijakan non-teknis.

Proses pembangunan model dilakukan melalui tahapan:

- a. Menghubungkan blok Retrieve dan *Decision Tree* di RapidMiner.
- b. Mengatur parameter tree depth (kedalaman pohon), criterion (seperti least square), dan metode pruning untuk menghindari overfitting.
- c. Menjalankan pelatihan model menggunakan data training dan mengamati hasil struktur pohon keputusan yang terbentuk, termasuk node dan leaf yang menunjukkan hubungan antar atribut.
- d. Mengidentifikasi variabel penting yang paling berpengaruh terhadap jumlah UMKM berdasarkan hasil feature selection dari struktur pohon.

Model ini memungkinkan pengguna untuk memahami aturan-aturan prediktif, seperti “Jika kategori usaha = kuliner dan wilayah = Kota Bandung, maka jumlah UMKM cenderung tinggi”.

3. Evaluasi Kinerja Model

Tahap akhir dari proses penelitian adalah mengevaluasi performa model yang telah dibangun menggunakan data uji. Evaluasi ini bertujuan untuk mengukur sejauh mana model mampu memprediksi jumlah UMKM secara akurat dan andal.

Pengujian dilakukan menggunakan blok Apply Model dan Performance pada RapidMiner. Adapun metrik evaluasi decision tree untuk regresi yang digunakan meliputi:

- a. RMSE (*root_mean_squared_error*): Mengukur rata-rata kesalahan prediksi.
- b. *Absolute error*: Mengukur rata-rata kesalahan absolute.
- c. *Relative error*: Digunakan untuk mengukur kesalahan relatif.

Correlation : Menghitung nilai korelasi.

HASIL DAN PEMBAHASAN

Pengumpulan Data dan Analisis

1. Pengumpulan Data

Sebelum digunakan dan diproses maka data-data tersebut akan disiapkan (*preprocessing*) terlebih dahulu, sehingga nantinya akan memiliki kelas binominal

PREDIKSI JUMLAH UMKM BERDASARKAN KATEGORI USAHA DAN LOKASI KABUPATEN/KOTA DI PROVINSI JAWA BARAT MENGGUNAKAN *DECISION TREE*

dan polynominal sesuai dengan aturan yang telah dibuat dan berdasarkan nilai datanya. Adapun hasil hasil *preprocessing* sesuai dengan jenis kelas datanya dapat dilihat pada tabel berikut :

Tabel 4.1 Data *Preprocessing*

Row No.	proyeksi_ju...	id	kode_provin...	nama_provi...	kode_kabup...	nama_kabup...	kategori_usa...	satuan	tahun
1	927	1	32	JAWA BARAT	3201	KABUPATEN ...	AKSESORIS	UNIT	2017
2	927	2	32	JAWA BARAT	3201	KABUPATEN ...	BATIK	UNIT	2017
3	132	3	32	JAWA BARAT	3201	KABUPATEN ...	BORDIR	UNIT	2017
4	33111	4	32	JAWA BARAT	3201	KABUPATEN ...	CRAFT	UNIT	2017
5	32316	5	32	JAWA BARAT	3201	KABUPATEN ...	FASHION	UNIT	2017
6	19734	6	32	JAWA BARAT	3201	KABUPATEN ...	KONVEKSI	UNIT	2017
7	142508	7	32	JAWA BARAT	3201	KABUPATEN ...	KULINER	UNIT	2017
8	104232	8	32	JAWA BARAT	3201	KABUPATEN ...	MAKANAN	UNIT	2017
9	19337	9	32	JAWA BARAT	3201	KABUPATEN ...	MINUMAN	UNIT	2017
10	45030	10	32	JAWA BARAT	3201	KABUPATEN ...	JASALANNYA	UNIT	2017
11	665	11	32	JAWA BARAT	3202	KABUPATEN ...	AKSESORIS	UNIT	2017
12	665	12	32	JAWA BARAT	3202	KABUPATEN ...	BATIK	UNIT	2017
13	95	13	32	JAWA BARAT	3202	KABUPATEN ...	BORDIR	UNIT	2017
14	23748	14	32	JAWA BARAT	3202	KABUPATEN ...	CRAFT	UNIT	2017
15	23178	15	32	JAWA BARAT	3202	KABUPATEN ...	FASHION	UNIT	2017
16	14154	16	32	JAWA BARAT	3202	KABUPATEN ...	KONVEKSI	UNIT	2017
17	102213	17	32	JAWA BARAT	3202	KABUPATEN ...	KULINER	UNIT	2017
18	74760	18	32	JAWA BARAT	3202	KABUPATEN ...	MAKANAN	UNIT	2017

ExampleSet (1,890 examples, 1 special attribute, 8 regular attributes)

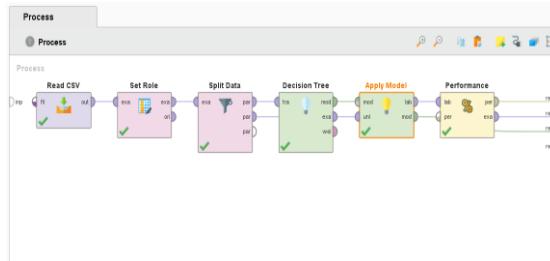
2. Analisis Algoritma *Decission Tree*

Pada penelitian ini dilakukan prediksi dengan menggunakan algoritma *decission tree* pada RapidMiner dengan total 1.890 baris data sampel. Adapun langkah-langkah prediksi data tersebut adalah sebagai berikut :

- Siapkan data kemudian buka aplikasi RapidMiner, import data menggunakan Read CSV kemudian pastikan kolom “nama_kabupaten_kota, kategori_usaha, dan tahun” diatur sebagai nominal. Kemudian pada kolom “proyeksi_jumlah_umkm” di atur sebagai label (target) dan bertipe numerical.
- Setelah data di siapkan buat proses decission tree dengan menambahkan operator set role untuk menentukan “proyeksi_jumlah_umkm” sebagai label, tambahkan operator decission tree, kemudian hubungkan output dari set role ke decission tree. Tambahkan operator split data untuk memisahkan data training (70%) dengan data testing (30%).
- Atur parameter *decission tree*, pada bagian *criterion* pilih *least square*. Kemudian *maximal depth* bisa dibatasi (10) agar pohon tidak *overfitting*.
- Kemudian jalankan dan evaluasi model dengan menambahkan *operator Apply Model* dan *Performance Regression*. Setelah semua operator terhubung jalankan proses dan sistem akan menampilkan hasil struktur pohon dan nilai prediksi.

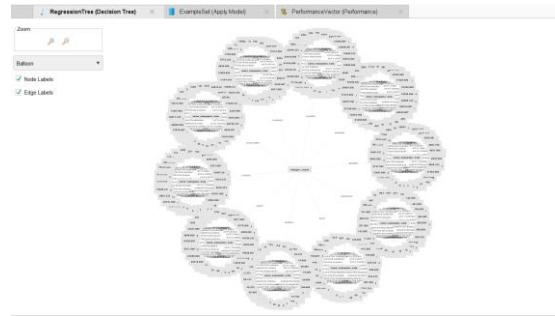
Evaluasi dan Validasi Hasil

Penelitian ini melakukan evaluasi model dengan menggunakan RMSE, Absolute error, Relative error, dan Correlation. Tahapan dalam memperoleh nilai tersebut dapat dilihat pada gambar dibawah ini :



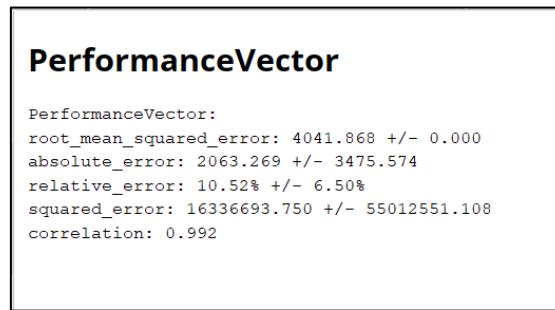
Gambar 4.1 Process Design

Berikut adalah hasil *tree* yang dibuat dengan model *balloon*:



Gambar 4.2 Hasil Regression Tree

Berikut ini merupakan hasil dari *performance vector* yang didapat dari algoritma decision tree yang digunakan pada penelitian kali ini, yaitu :



Gambar 4.3 Hasil Performance Vector

Berdasarkan hasil yang di peroleh dari *performance vector* tersebut dapat definisikan sebagai berikut :

1. RMSE (*root_mean_squared_error*) : 4041.868 yaitu Rata-rata kesalahan prediksi sekitar 4042 unit UMKM.
2. *Absolute Error* : 2063.269 yaitu Rata-rata kesalahan absolut lebih kecil (bagus).

PREDIKSI JUMLAH UMKM BERDASARKAN KATEGORI USAHA DAN LOKASI KABUPATEN/KOTA DI PROVINSI JAWA BARAT MENGGUNAKAN *DECISION TREE*

3. *Relative Error* : 10.52% yaitu Kesalahan relatif sekitar 10,5% dari nilai sebenarnya, ini cukup baik.
4. *Correlation* : 0.992 yaitu Nilai korelasi mendekati 1, menunjukkan model *Decision Tree* yang dibuat ini sangat baik dalam memprediksi.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil pemodelan menggunakan *Decision Tree* Regression di RapidMiner:

1. Model dapat memprediksi jumlah UMKM berdasarkan kategori usaha dan lokasi kabupaten/kota di Jawa Barat dengan baik, dibuktikan dengan nilai korelasi sebesar 0.992 yang menunjukkan hubungan prediksi dengan data aktual sangat kuat.
2. Nilai RMSE sebesar 4041.868 dan Absolute Error 2063.269 menunjukkan tingkat kesalahan prediksi masih dalam batas wajar dibandingkan skala data UMKM.
3. Relative Error 10.52% mengindikasikan model cukup akurat dalam melakukan estimasi jumlah UMKM.

Saran

1. Perlu mencoba algoritma lain seperti *Random Forest* atau *Gradient Boosted Trees* untuk membandingkan kinerja dan memperoleh hasil yang lebih optimal.
2. Lakukan *feature engineering*, misalnya menambahkan variabel tahun sebagai faktor prediktif atau mengelompokkan kategori usaha tertentu agar model lebih memahami pola data.
3. Perlu update data UMKM secara berkala supaya model tetap relevan dan akurat untuk prediksi di masa depan.

DAFTAR REFERENSI

- Han, J., Pei, J., & Kamber, M. (2022). Data Mining: Concepts and Techniques (4th ed.). Morgan Kaufmann.
- Kementerian Koperasi dan UKM. (2022). Laporan Tahunan UMKM Nasional. Jakarta: Deputi Bidang Usaha Mikro.

- Kotu, V., & Deshpande, B. (2022). *Data Science: Concepts and Practice*. Academic Press.
- Nugroho, A., & Kurniawan, B. (2022). Analisis Sentimen Menggunakan RapidMiner: Studi Kasus Review Produk E-Commerce. *Jurnal Teknik Informatika*, 9(1), 45–53.
- Rahayu, R. R., & Wijaya, H. (2023). Pengaruh UMKM Terhadap Ketahanan Ekonomi Daerah: Studi Kasus Provinsi Jawa Barat. *Jurnal Ekonomi Pembangunan Indonesia*, 12(1), 55–64.
- Ramadhan, S., Alamin, Z., Jannah, M., & Akbar, M. (2021). Prediksi Keberhasilan UMKM Menggunakan Decision Tree. *Journix: Jurnal Informatika dan Komputasi*, 3(2), 101–110.
- Wibowo, H., & Handayani, L. (2021). Pemetaan UMKM Berbasis Data Mining: Studi Wilayah Bekasi. *Jurnal Sistem Informasi dan Bisnis*, 9(2), 92–102.