

Decision Tree Classification of Rule-Based MSME Sector Labels for Widowed and Divorced Women Using Socioeconomic Factors

(Klasifikasi Decision Tree terhadap Label Sektor UMKM Berbasis Aturan bagi Perempuan Berstatus Cerai Hidup dan Cerai Mati Berdasarkan Faktor Sosial Ekonomi)

Rahmatsyah Ananta Putra Ginting^{a,1*}, Rakhmat Kurniawan^{a,2}

^a Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara, Medan, 20353, Indonesia

E-mail: ¹rahmatsyahanantaputra@gmail.com; ²rakhmat.kr@uinsu.ac.id

*Corresponding Author: rahmatsyahanantaputra@gmail.com (A. Rahmat).

ARTICLE INFO

Received: 14 July 2026
Revised: 28 July 2026
Accepted: 31 July 2026

How to Cite:

Ginting, R. A. P., & Kurniawan, R. (2026). Decision Tree Classification of Rule-Based MSME Sector Labels for Widowed and Divorced Women Using Socioeconomic Factors. Jurnal IPTEK Bagi Masyarakat, 6(1), 304–312. <https://doi.org/10.55537/j-ibm.v6i1.1820>

ABSTRACT

Widowed and divorced women may rely on micro, small, and medium enterprises to sustain household income, yet preliminary sector screening requires a consistent and transparent mechanism. This study developed a Decision Tree model to classify three rule-based MSME sector labels using demographic and economic characteristics. The dataset comprised 3,659 women aged 18–59 years in Kabanjahe District, Karo Regency, Indonesia. Age, occupation, marital-status category, monthly income, number of dependents, and business-ownership status were used as predictors. The target labels—culinary, fashion and handicrafts, and services and creative products—were constructed from predefined weighting rules. After preprocessing and a stratified 80:20 split, evaluation on 732 test records yielded 94.81% accuracy, 94.86% weighted precision, 94.81% weighted recall, and a 94.75% weighted F1-score. Low-income status had the highest impurity-based feature importance (43.11%), followed by the 18–25 age category (12.05%). The model consistently reproduced the constructed labels and produced interpretable decision rules. However, the results do not establish real-world sector suitability or business success because the labels were generated from the same predictor attributes rather than independent expert judgments or observed business outcomes.

Keywords:

decision tree; rule-based labels; socioeconomic factors; MSME sector screening; widowed and divorced women.

ABSTRAK

Perempuan berstatus cerai hidup atau cerai mati dapat mengandalkan usaha mikro, kecil, dan menengah untuk mempertahankan pendapatan rumah tangga, tetapi penyaringan awal sektor usaha memerlukan mekanisme yang konsisten dan transparan. Penelitian ini mengembangkan model Decision Tree untuk mengklasifikasikan tiga label sektor UMKM berbasis aturan berdasarkan karakteristik demografis dan ekonomi. Dataset terdiri atas 3.659 perempuan berusia 18–59 tahun di Kecamatan Kabanjahe, Kabupaten Karo. Umur, pekerjaan, kategori status perkawinan, penghasilan bulanan, jumlah tanggungan, dan status kepemilikan usaha digunakan sebagai prediktor. Label target—kuliner, fashion dan kerajinan, serta jasa dan produk kreatif—dibentuk melalui aturan pembobotan yang telah ditetapkan. Setelah prapemrosesan dan pembagian data terstratifikasi 80:20, evaluasi terhadap 732 data uji menghasilkan akurasi 94,81%, presisi tertimbang 94,86%, recall tertimbang 94,81%, dan F1-score tertimbang 94,75%. Status penghasilan rendah memiliki kepentingan fitur berbasis impurity tertinggi sebesar 43,11%, diikuti kelompok umur 18–25 tahun sebesar 12,05%. Model mampu mereproduksi label yang dibentuk secara konsisten dan menghasilkan aturan keputusan yang dapat ditelusuri. Namun, hasil penelitian belum membuktikan kesesuaian sektor atau keberhasilan usaha di dunia nyata karena label dibentuk dari atribut prediktor yang sama, bukan dari penilaian ahli independen maupun outcome usaha aktual.

Kata kunci:

decision tree; faktor sosial ekonomi; label berbasis aturan; penyaringan sektor UMKM; perempuan berstatus cerai.

Pendahuluan

Usaha Mikro, Kecil, dan Menengah (UMKM), termasuk industri mikro dan kecil, memiliki peran penting dalam penciptaan lapangan kerja dan penggerak ekonomi lokal. Publikasi statistik resmi juga menunjukkan keragaman karakteristik serta tantangan pengembangan usaha pada tingkat nasional dan Provinsi Sumatera Utara (Badan Pusat Statistik, 2024a; Badan Pusat Statistik Provinsi Sumatera Utara, 2024b). Industri mikro dan kecil juga berperan dalam menciptakan lapangan kerja lokal, tetapi masih menghadapi keterbatasan modal, kualitas sumber daya manusia, manajemen usaha, akses pasar, dan legalitas (Latifah et al., 2022). Kondisi tersebut mendorong pemerintah dan berbagai lembaga untuk menjalankan program pembiayaan, pelatihan, pendampingan, serta bantuan usaha. Namun, keberhasilan program tidak hanya bergantung pada ketersediaan bantuan, tetapi juga pada ketepatan penentuan sasaran dan kesesuaian dukungan dengan karakteristik penerima (Paransa, 2024).

Perempuan memegang peran penting dalam pengembangan UMKM karena kegiatan usaha memberi ruang untuk memperoleh pendapatan sekaligus menjalankan tanggung jawab keluarga (Widia & Octafia, 2022). Meskipun demikian, pelaku UMKM perempuan masih menghadapi keterbatasan akses pasar, pembiayaan, penguasaan teknologi, dan pelatihan teknis (Nofriadi et al., 2024). Mereka juga harus membagi waktu antara pengelolaan usaha dan pekerjaan domestik. Perbedaan umur, pendapatan, jenis usaha, dan kondisi sosial turut memengaruhi kemampuan pelaku UMKM dalam mengakses sumber daya serta memanfaatkan teknologi (Susanti et al., 2023). Oleh karena itu, program pengembangan usaha bagi perempuan tidak dapat menggunakan pendekatan yang seragam, tetapi perlu mempertimbangkan karakteristik dan kebutuhan setiap kelompok penerima (Widia & Octafia, 2023) (Maharani & Setiawati, 2024).

Kebutuhan terhadap dukungan ekonomi menjadi lebih kuat pada perempuan berstatus cerai hidup atau cerai mati, baik akibat cerai mati maupun cerai hidup (Yulfa et al., 2022). Kelompok ini dapat menghadapi tanggung jawab ekonomi rumah tangga secara mandiri, keterbatasan akses terhadap sumber daya produktif, dan rendahnya keterlibatan dalam proses pengambilan keputusan pembangunan (Kasim et al., 2022). Penelitian mengenai perempuan kepala keluarga di Aceh menunjukkan bahwa keterbatasan akses ekonomi dan lemahnya pengakuan dalam kebijakan dapat memperbesar risiko kemiskinan (Alie et al., 2024) (Mahendra & Wibowo, 2023). Program pemberdayaan melalui UMKM dapat membantu perempuan kepala keluarga memperoleh pendapatan dan memperkuat kemandirian ekonomi, tetapi pelaksanaannya perlu didasarkan pada pemetaan kondisi sosial dan ekonomi yang akurat (Ikhsan & Zulkifli, 2022).

Penilaian calon penerima program yang hanya mengandalkan proses manual atau kriteria terbatas dapat menghasilkan keputusan yang tidak konsisten. Model klasifikasi dapat membantu menyusun penyaringan awal yang lebih sistematis, tetapi keputusan akhir tetap memerlukan verifikasi administratif dan asesmen ahli (Pratiwi et al., 2022). Sejumlah penelitian nasional telah menggunakan teknik klasifikasi untuk membantu mengolah atribut ekonomi, pekerjaan, kondisi keluarga, kepemilikan tempat tinggal, umur, dan status sosial (Avianto & Wibowo, 2024). Penerapan algoritma C4.5 pada klasifikasi bantuan sosial menunjukkan bahwa struktur pohon keputusan dapat mengubah data penerima menjadi aturan yang lebih sistematis dan mudah ditelusuri (Avianto & Wibowo, 2024; Bachtiar & Mahradianur, 2023). Pendekatan tersebut telah digunakan dalam klasifikasi bantuan bagi mahasiswa, masyarakat desa, keluarga penerima bantuan langsung tunai, dan calon penerima program sosial pemerintah.

Decision Tree relevan digunakan sebagai alat klasifikasi pendukung keputusan karena menghasilkan pola klasifikasi dalam bentuk cabang dan aturan yang dapat dipahami oleh pengguna nonteknis. Model dapat menunjukkan atribut yang digunakan dalam setiap keputusan sehingga hasilnya lebih mudah diperiksa daripada model yang bekerja sebagai kotak hitam (Bachtiar & Mahradianur, 2023). Namun, kualitas klasifikasi tetap bergantung pada relevansi atribut, kejelasan kelas target, kualitas data, dan kesesuaian model dengan kondisi lokal. Penggunaan faktor demografis dan ekonomi juga memiliki dasar empiris karena umur, pendapatan, jenis usaha, dan karakteristik sosial dapat memengaruhi akses serta keberlanjutan kegiatan UMKM. Penelitian berbasis pohon keputusan pada UMKM di wilayah perbatasan juga menunjukkan bahwa aturan klasifikasi dapat digunakan sebagai pendukung evaluasi usaha secara terstruktur (Hartana & Mandasari, 2024).

Suseno et al. (2023) membandingkan algoritma K-Nearest Neighbor dan C4.5 dalam menganalisis rekomendasi pemberian bantuan kepada pelaku UMKM di Kabupaten Pekalongan. Penelitian tersebut menggunakan 312 data awal yang berkurang menjadi 279 data setelah melalui tahap prapemrosesan. Hasil pengujian menunjukkan bahwa K-Nearest Neighbor memperoleh akurasi sebesar 94,94%, sedangkan C4.5 menghasilkan akurasi sebesar 86,08%, sehingga K-Nearest Neighbor menunjukkan performa lebih tinggi pada dataset yang digunakan (Suseno et al., 2023). Penelitian Aldy Novriandy dan Winarsih (2023) membandingkan algoritma Naive Bayes dan C4.5 dalam mengklasifikasikan kelayakan penerima bantuan UMKM. Hasil penelitian menunjukkan bahwa C4.5 mencapai akurasi sebesar 90%, sedangkan Naive Bayes memperoleh akurasi sebesar 70%. Kedua penelitian tersebut menunjukkan bahwa algoritma klasifikasi dapat menghasilkan performa yang baik pada dataset masing-masing. Namun, penelitian tersebut masih berfokus pada klasifikasi kelayakan penerima bantuan dan belum membahas klasifikasi label sektor UMKM bagi perempuan berstatus cerai hidup atau cerai mati berdasarkan gabungan faktor demografis dan ekonomi (Novriandy & Winarsih, 2023).

Berdasarkan kajian terhadap penelitian Suseno et al. (2023) serta Novriandy dan Winarsih (2023), penelitian sebelumnya lebih banyak memusatkan klasifikasi pada penentuan kelayakan penerima bantuan UMKM

dalam bentuk kelas biner. Pendekatan tersebut belum menjawab kebutuhan klasifikasi multikelas untuk memetakan karakteristik demografis dan ekonomi ke dalam beberapa alternatif sektor usaha yang dapat ditelusuri melalui aturan keputusan. Masalah ilmiah yang diangkat dalam penelitian ini terletak pada kebutuhan membangun struktur klasifikasi yang mampu mengubah kombinasi umur, pekerjaan, status perkawinan, penghasilan, jumlah tanggungan, dan kepemilikan usaha menjadi tiga label rekomendasi secara konsisten dan transparan. Oleh karena itu, kontribusi penelitian ini bukan hanya terletak pada perbedaan objek dan lokasi, tetapi pada pengembangan model Decision Tree untuk merepresentasikan aturan label sektor kuliner, fashion dan kerajinan, serta jasa dan produk kreatif bagi perempuan berstatus cerai hidup atau cerai mati di Kecamatan Kabanjahe. Model tersebut diposisikan sebagai bukti konsep klasifikasi berbasis aturan yang masih memerlukan validasi ahli dan pengujian berdasarkan keberhasilan usaha aktual.

Penelitian ini bertujuan membangun model klasifikasi label sektor UMKM bagi perempuan berstatus cerai hidup atau cerai mati berdasarkan faktor demografis dan ekonomi menggunakan Decision Tree. Penelitian diarahkan untuk mengidentifikasi pola klasifikasi, menentukan atribut yang paling berkontribusi, menghasilkan aturan keputusan yang transparan, dan mengevaluasi kemampuan model dalam mengklasifikasikan sektor usaha. Kontribusi penelitian terletak pada pengembangan pendekatan berbasis data yang berfokus pada kelompok perempuan rentan dan kondisi lokal Kecamatan Kabanjahe. Hasilnya diharapkan dapat menjadi pendukung asesmen awal bagi pemerintah daerah atau lembaga pemberdayaan, bukan sebagai pengganti penilaian ahli dan verifikasi kebutuhan calon penerima.

Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi dan dilaksanakan di Kecamatan Kabanjahe, Kabupaten Karo, pada September 2025 sampai Mei 2026. Data dasar kependudukan diperoleh dari Dinas Kependudukan dan Pencatatan Sipil Kabupaten Karo pada 10 Maret 2025 untuk kepentingan penelitian skripsi atau tugas akhir. Data tersebut merupakan rekapitulasi tahun 2024 dan menjadi data terbaru yang tersedia pada saat pengambilan karena pembaruan data berikutnya dijadwalkan pada pertengahan tahun 2025. Cakupan data yang diizinkan meliputi jumlah perceraian, kategori perceraian berupa cerai hidup dan cerai mati, jenis kelamin, serta wilayah administrasi. Penelitian memfokuskan analisis pada perempuan berstatus cerai hidup atau cerai mati yang tercatat di Kecamatan Kabanjahe, Kabupaten Karo. Atribut nama hanya digunakan sebagai identitas awal dan dihapus sebelum pemodelan. Pemeriksaan data menunjukkan tidak terdapat nilai kosong maupun duplikasi.

Umur dikelompokkan menjadi 18–25, 26–35, 36–45, 46–55, dan 56–59 tahun. Penghasilan dikelompokkan menjadi rendah, sedang, dan tinggi, sedangkan jumlah tanggungan dikelompokkan menjadi tidak ada, sedikit, sedang, dan banyak. Pekerjaan, kategori status perkawinan, dan status usaha diperlakukan sebagai atribut nominal. Label target terdiri atas Kuliner, Fashion/Kerajinan, serta Jasa/Produk Kreatif. Setiap catatan memperoleh skor pada ketiga kelas berdasarkan bobot umur, pekerjaan, kategori status perkawinan, penghasilan, jumlah tanggungan, dan status usaha. Kelas dengan skor tertinggi ditetapkan sebagai label target.

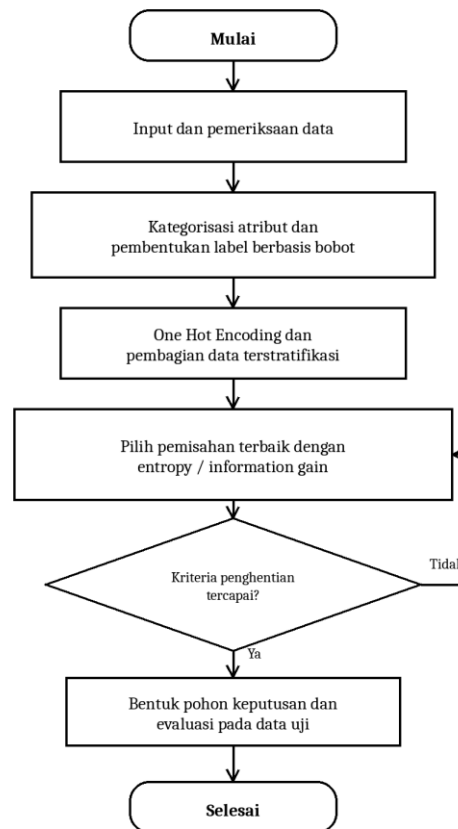
Label target pada penelitian ini merupakan label sintesis berbasis aturan yang dibentuk menggunakan atribut umur, pekerjaan, kategori status perkawinan, penghasilan, jumlah tanggungan, dan status usaha. Atribut yang sama kemudian digunakan sebagai prediktor dalam pengembangan model Decision Tree. Oleh karena itu, penelitian ini diposisikan sebagai proof-of-concept untuk mengevaluasi kemampuan model dalam mempelajari dan mereproduksi aturan pembentukan label secara konsisten. Nilai akurasi, presisi, recall, dan F1-score tidak ditafsirkan sebagai bukti bahwa sektor usaha yang dihasilkan merupakan sektor yang paling sesuai bagi setiap individu. Validitas kesesuaian sektor usaha di dunia nyata masih memerlukan ground truth independen yang diperoleh melalui penilaian panel ahli, asesmen kebutuhan individu, atau data keberhasilan usaha aktual.

Prapemrosesan meliputi pemeriksaan data, kategorisasi atribut numerik, penghapusan nama, dan One Hot Encoding yang menghasilkan 23 fitur. Dataset dibagi secara terstratifikasi menjadi 2.927 data latih atau 80% dan 732 data uji atau 20% agar proporsi setiap kelas tetap terjaga. Model menggunakan Decision Tree dengan kriteria entropy, bukan C4.5, karena implementasinya tidak menunjukkan penerapan gain ratio dan pruning C4.5 secara langsung. Parameter model ditetapkan untuk membatasi kompleksitas pohon dan mencegah terbentuknya cabang dengan jumlah data terlalu kecil.

Tabel 1. Konfigurasi Pengembangan Model Decision Tree

Komponen	Konfigurasi
Jumlah data	3.659 data
Atribut prediktor	Umur, pekerjaan, kategori status perkawinan, penghasilan, tanggungan, dan status usaha
Pengodean	One Hot Encoding
Jumlah fitur	23 fitur
Pembagian data	Stratified 80% data latih dan 20% data uji
Data latih	2.927 data
Data uji	732 data
Model	Decision Tree
Criterion	entropy

Komponen	Konfigurasi
Max depth	6
Min samples split	10
Min samples leaf	5
Random state	42
Evaluasi	Confusion matrix, accuracy, precision, recall, dan F1-score



Gambar 1. Flowchart pembentukan pohon Keputusan

Alur penelitian pada Gambar 1 dimulai dari input dan pemeriksaan data, kategorisasi atribut, pembentukan label berbasis bobot, One Hot Encoding, pembagian data, pelatihan model, pemeriksaan kriteria penghentian, dan evaluasi. Pemisahan node dilakukan selama menghasilkan penurunan impurity dan memenuhi batas parameter. Proses dihentikan ketika node telah homogen, kedalaman maksimum tercapai, atau jumlah sampel tidak memenuhi batas pemisahan. Karena itu, bagian keputusan pada flowchart diubah dari “Semua atribut sudah masuk?” menjadi “Kriteria penghentian telah tercapai?”. Kinerja model dievaluasi menggunakan confusion matrix, accuracy, precision, recall, dan F1-score, sedangkan feature importance digunakan untuk menjelaskan kontribusi fitur terhadap pemisahan node dan tidak ditafsirkan sebagai hubungan sebab-akibat.

Hasil

Karakteristik Data dan Hasil Prapemrosesan

Dataset penelitian terdiri atas 3.659 catatan perempuan berstatus cerai hidup atau cerai mati di Kecamatan Kabanjahe, Kabupaten Karo. Data administratif awal memuat nama, umur, pekerjaan, kategori status perkawinan, penghasilan bulanan, jumlah tanggungan, dan status usaha. Nama tercantum sebagai penanda administratif pada data sumber, tetapi tidak ditetapkan sebagai variabel penelitian. Sebelum pembentukan label dan pemodelan, atribut nama dihapus dari dataset kerja sehingga proses kategorisasi, pengodean, pembagian data, pelatihan, pengujian, dan pelaporan hanya menggunakan enam atribut demografis dan ekonomi tanpa identitas langsung individu.

Tabel 2. Karakteristik Responden Penelitian

Variabel	Kategori	Jumlah	Persentase
Umur	18–25 tahun	761	20,80%
Umur	26–35 tahun	876	23,94%
Umur	36–45 tahun	865	23,64%
Umur	46–55 tahun	831	22,71%

Variabel	Kategori	Jumlah	Persentase
Umur	56–59 tahun	326	8,91%
Pekerjaan	Wiraswasta	2.924	79,91%
Pekerjaan	Mengurus Rumah Tangga	287	7,84%
Pekerjaan	Pelajar/Mahasiswa	138	3,77%
Pekerjaan	Karyawan Swasta	81	2,21%
Pekerjaan	Guru Honorer	80	2,19%
Pekerjaan	Perawat	80	2,19%
Pekerjaan	Petani	69	1,89%
Kategori status perkawinan	Cerai Mati	3.043	83,16%
Kategori status perkawinan	Cerai Hidup	616	16,84%
Status usaha	Ada	3.079	84,15%
Status usaha	Tidak Ada	580	15,85%
Penghasilan	Rendah	1.794	49,03%
Penghasilan	Sedang	1.534	41,92%
Penghasilan	Tinggi	331	9,05%
Jumlah tanggungan	Tidak Ada	533	14,57%
Jumlah tanggungan	Sedikit	2.117	57,86%
Jumlah tanggungan	Sedang	908	24,82%
Jumlah tanggungan	Banyak	101	2,76%

Berdasarkan [Tabel 2](#), data menunjukkan variasi umur, pekerjaan, kategori status perkawinan, penghasilan, jumlah tanggungan, dan status usaha. Pemeriksaan awal menunjukkan rentang umur 18 sampai 59 tahun serta tidak ditemukan nilai kosong maupun data duplikat. Umur, penghasilan bulanan, dan jumlah tanggungan kemudian dikategorikan, sedangkan pekerjaan, kategori status perkawinan, dan status usaha dipertahankan sebagai atribut kategorikal. Seluruh atribut kategorikal ditransformasikan menggunakan One Hot Encoding sehingga menghasilkan 23 fitur numerik tanpa mengubah jumlah data. Dataset hasil prapemrosesan tidak lagi memuat nama dan selanjutnya digunakan dalam pembentukan label, pelatihan, serta evaluasi model Decision Tree.

Untuk melindungi privasi subjek, hasil penelitian disajikan secara agregat tanpa mencantumkan nama atau informasi yang memungkinkan identifikasi langsung. Dataset bernama tidak digunakan sebagai masukan model, tidak disertakan dalam publikasi, dan tidak boleh dipindahkan atau dibagikan kepada pihak yang tidak berwenang. Akses terhadap data dibatasi untuk kepentingan penelitian dan penggunaannya harus didasarkan pada izin instansi pemilik data. Nomor izin penggunaan data serta status persetujuan etik, waiver, atau exemption perlu dicantumkan berdasarkan dokumen resmi penelitian agar aspek legalitas, kerahasiaan, dan perlindungan kelompok rentan dapat diverifikasi.

Distribusi Kelas dan Pembagian Data

Dataset terdiri atas tiga kelas label, yaitu Jasa/Produk Kreatif sebanyak 1.460 data atau 39,90%, Kuliner sebanyak 1.442 data atau 39,41%, serta Fashion/Kerajinan sebanyak 757 data atau 20,69%. Dataset dibagi secara terstratifikasi dengan rasio 80:20 dan `random_state=42`, sehingga diperoleh 2.927 data latih dan 732 data uji. Stratifikasi digunakan untuk mempertahankan proporsi setiap kelas pada kedua kelompok data.

Tabel 3. Distribusi Kelas pada Data Latih dan Data Uji

Kelas label	Data Latih	Data Uji	Total
Fashion/Kerajinan	606	151	757
Jasa/Produk Kreatif	1.168	292	1.460
Kuliner	1.153	289	1.442
Total	2.927	732	3.659

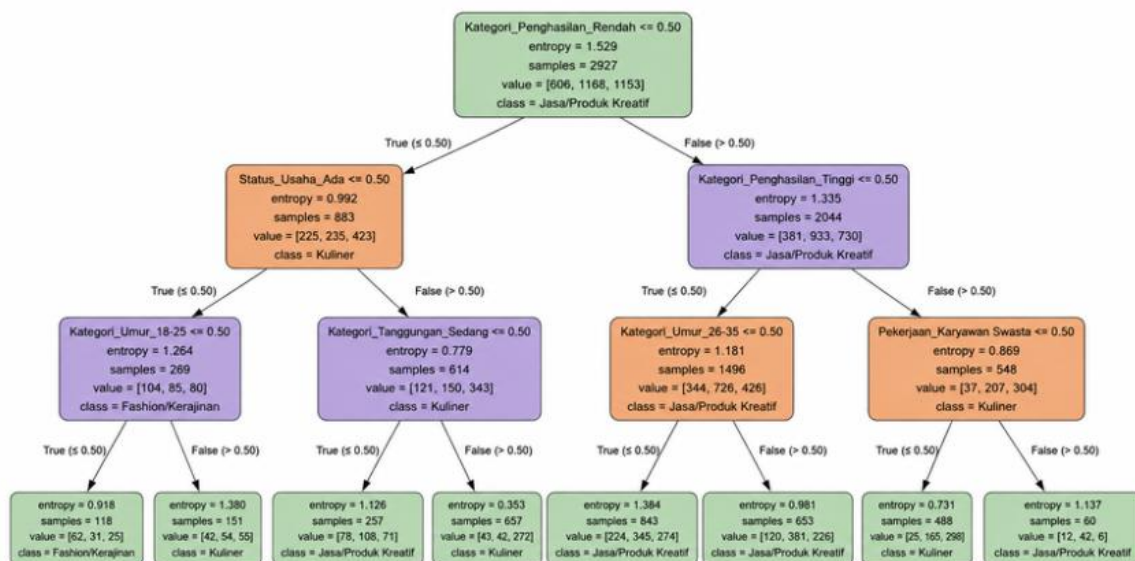
[Tabel 3](#) menunjukkan bahwa seluruh kelas tetap terwakili secara proporsional pada data latih dan data uji. Label Jasa/Produk Kreatif merupakan kelas dengan jumlah terbesar, yaitu 1.460 data atau 39,90%, diikuti Kuliner sebanyak 1.442 data atau 39,41% dan Fashion/Kerajinan sebanyak 757 data atau 20,69%. Proporsi kelas mayoritas sebesar 39,90% digunakan sebagai majority-class baseline, yaitu performa acuan apabila seluruh data diprediksi sebagai kelas terbanyak. Pembagian data dilakukan secara terstratifikasi dengan rasio 80:20 dan `random_state=42`, sehingga data uji tidak digunakan dalam proses pelatihan model.

Evaluasi dalam penelitian ini masih menggunakan satu kali pembagian data latih dan data uji. Oleh karena itu, hasil evaluasi hanya menggambarkan kinerja model pada pembagian data tersebut dan belum menunjukkan kestabilan performa pada berbagai variasi sampel. Penelitian ini belum menerapkan Repeated Stratified K-Fold Cross-Validation dan belum melakukan validasi eksternal menggunakan data dari wilayah atau periode yang berbeda. Dengan demikian, hasil model tidak ditafsirkan sebagai bukti kemampuan generalisasi di luar dataset penelitian.

Hasil Pembentukan Model

Model Decision Tree dibangun menggunakan 2.927 data latih dan 23 fitur numerik hasil One Hot Encoding. Pemisahan node menggunakan kriteria entropy, sedangkan kompleksitas pohon dibatasi melalui $\text{max_depth}=6$, $\text{min_samples_split}=10$, dan $\text{min_samples_leaf}=5$, dengan $\text{random_state}=42$ untuk menjaga keterulangan proses. Ketiga hiperparameter tersebut ditetapkan sebagai konfigurasi awal untuk membatasi kedalaman pohon, mencegah pemisahan node dengan jumlah data yang terlalu kecil, dan mempertahankan keterbacaan aturan. Namun, pemilihannya belum didasarkan pada prosedur pencarian hiperparameter yang sistematis sehingga konfigurasi tersebut tidak dapat dinyatakan sebagai kombinasi yang optimal. Penentuan parameter pada pengembangan berikutnya perlu dilakukan melalui grid search atau randomized search di dalam repeated stratified cross-validation tanpa melibatkan data uji.

Hasil pelatihan menghasilkan pohon dengan kedalaman enam dan 41 simpul daun. Akurasi pada data latih sebesar 94,47%, sedangkan akurasi pada data uji sebesar 94,81%, dengan perbedaan 0,34 poin persentase. Kedekatan kedua nilai tersebut hanya menggambarkan performa pada satu pembagian data dan tidak dapat digunakan sebagai bukti bahwa model terbebas dari overfitting. Nilai akurasi data uji yang sedikit lebih tinggi daripada data latih dapat terjadi akibat variasi sampel pada proses pembagian hold-out. Oleh karena itu, kestabilan performa dan kemampuan generalisasi model harus dinilai melalui validasi silang berulang, interval kepercayaan, dan pengujian pada data eksternal atau prospektif.

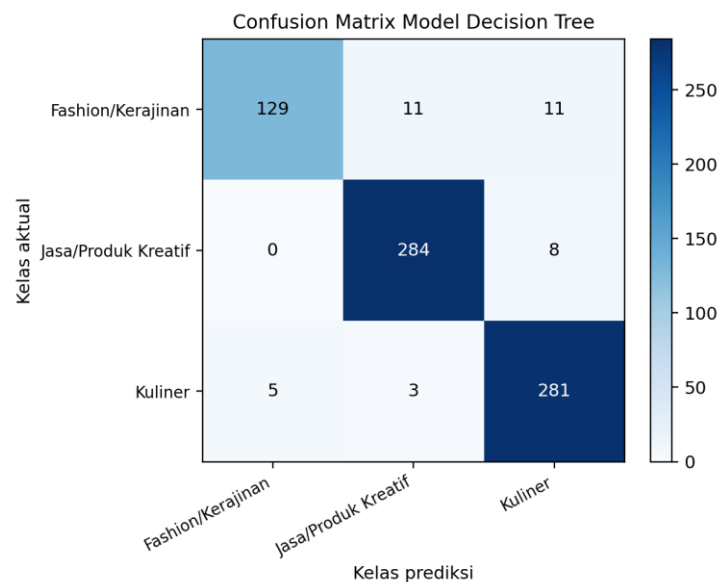


Gambar 2. Struktur Ringkas Pohon Keputusan

Gambar 2 menunjukkan bahwa fitur Kategori_Penghasilan_Rendah dipilih sebagai simpul akar karena menghasilkan penurunan impurity terbesar pada dataset, hasil pengodean, dan konfigurasi model yang digunakan. Nilai fitur kurang dari atau sama dengan 0,50 menunjukkan bahwa catatan tidak termasuk kategori penghasilan rendah, sedangkan nilai lebih dari 0,50 menunjukkan kondisi sebaliknya. Pada tingkat berikutnya, model menggunakan status usaha, kategori penghasilan lain, kelompok umur, jumlah tanggungan, dan pekerjaan untuk membentuk cabang klasifikasi. Posisi fitur penghasilan rendah sebagai akar tidak menunjukkan hubungan sebab-akibat dan tidak membuktikan bahwa tingkat penghasilan secara substantif menentukan sektor usaha terbaik. Hasil tersebut hanya menjelaskan pola pemisahan yang digunakan model untuk mereproduksi label berbasis aturan pada dataset penelitian.

Evaluasi Model

Evaluasi model dilakukan menggunakan 732 data uji untuk mengukur kemampuan Decision Tree dalam mengklasifikasikan label sektor UMKM ke dalam tiga kelas, yaitu Fashion/Kerajinan, Jasa/Produk Kreatif, dan Kuliner. Kinerja prediksi dianalisis melalui confusion matrix yang memperlihatkan hubungan antara kelas aktual dan kelas hasil prediksi. Nilai pada diagonal utama menunjukkan prediksi yang benar, sedangkan nilai di luar diagonal menunjukkan kesalahan klasifikasi antarkelas.



Gambar 3. Confusion Matrix Model Decision Tree

Gambar 3 menunjukkan bahwa model berhasil mengklasifikasikan dengan benar 129 dari 151 data aktual Fashion/Kerajinan, 284 dari 292 data aktual Jasa/Produk Kreatif, dan 281 dari 289 data aktual Kuliner. Pada kelas Fashion/Kerajinan, sebanyak 11 data diprediksi sebagai Jasa/Produk Kreatif dan 11 data diprediksi sebagai Kuliner. Pada kelas Jasa/Produk Kreatif, tidak terdapat data yang diprediksi sebagai Fashion/Kerajinan, sedangkan delapan data diprediksi sebagai Kuliner. Pada kelas Kuliner, lima data diprediksi sebagai Fashion/Kerajinan dan tiga data diprediksi sebagai Jasa/Produk Kreatif. Secara keseluruhan, model menghasilkan 694 prediksi benar dan 38 prediksi salah dari 732 data uji. Kesalahan klasifikasi terbanyak ditemukan pada kelas Fashion/Kerajinan, yaitu 22 data, sedangkan kelas Jasa/Produk Kreatif dan Kuliner masing-masing memiliki delapan kesalahan klasifikasi.

Analisis per kelas menunjukkan bahwa 129 dari 151 data Fashion/Kerajinan diklasifikasikan dengan benar, sedangkan 22 data lainnya masing-masing diprediksi sebagai Jasa/Produk Kreatif dan Kuliner sebanyak 11 data. Kelas ini menghasilkan recall sebesar 85,43% dengan interval kepercayaan 95% sebesar 78,93%–90,18%, yang berarti model tidak mengenali sekitar 14,57% label aktual Fashion/Kerajinan. Sementara itu, kelas Jasa/Produk Kreatif menghasilkan 284 prediksi benar dari 292 data dengan recall 97,26% dan interval kepercayaan 95% sebesar 94,69%–98,61%. Kelas Kuliner menghasilkan 281 prediksi benar dari 289 data dengan recall 97,23% dan interval kepercayaan 95% sebesar 94,63%–98,59%. Hasil tersebut menunjukkan bahwa kesalahan model lebih terkonsentrasi pada kelas Fashion/Kerajinan dibandingkan dua kelas lainnya.

Tabel 4. Kinerja Klasifikasi dan Ketidakpastian pada Data Uji

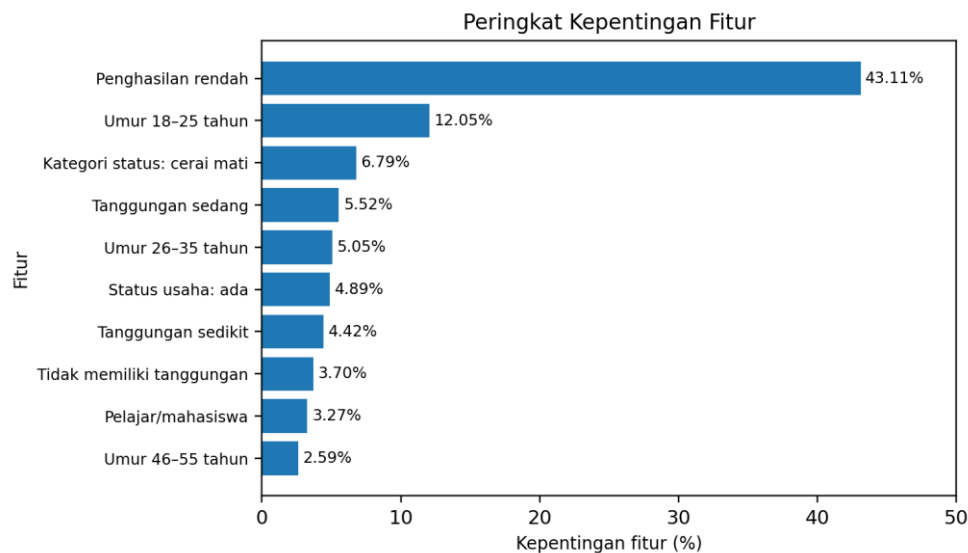
Kelas label	Precision	Recall dan 95% CI	F1-score	Support	False negative
Fashion/Kerajinan	0,96	0,85 (0,79–0,90)	0,91	151	22
Jasa/Produk Kreatif	0,95	0,97 (0,95–0,99)	0,96	292	8
Kuliner	0,94	0,97 (0,95–0,99)	0,95	289	8
Accuracy		0,95 (0,93–0,96)		732	38
Macro average	0,95	0,93	0,94	732	
Weighted average	0,95	0,95	0,95	732	

Kesalahan pada kelas Fashion/Kerajinan perlu diperhatikan karena label sektor fashion dan kerajinan dapat memiliki kebutuhan keterampilan, bahan, peralatan, modal, dan akses pasar yang berbeda dari sektor kuliner maupun jasa dan produk kreatif. Kesalahan klasifikasi berpotensi mengarahkan asesmen awal ke bentuk dukungan yang tidak sesuai dengan label berbasis aturan yang seharusnya diterima. Namun, karena label target dibentuk melalui pembobotan sintetis, kesalahan tersebut hanya menunjukkan ketidaksesuaian prediksi model terhadap aturan pembentukan label dan belum membuktikan ketidaksesuaian sektor usaha secara aktual. Oleh karena itu, hasil model harus digunakan sebagai pendukung penyaringan awal dan tetap memerlukan verifikasi ahli serta asesmen kebutuhan individu sebelum hasil klasifikasi digunakan.

Kepentingan Fitur

Kepentingan fitur dihitung berdasarkan penurunan impurity pada setiap pemisahan pohon. Karena fitur berasal dari One Hot Encoding, nilai ini dapat bias terhadap kategori yang memiliki lebih banyak peluang pemisahan dan tidak stabil pada fitur dummy yang berkorelasi. Oleh karena itu, hasilnya hanya digunakan untuk

menjelaskan struktur internal model.



Gambar 4. Peringkat Kepentingan Fitur

Gambar 4 menunjukkan bahwa kategori penghasilan rendah memiliki nilai tertinggi sebesar 43,11%, diikuti umur 18–25 tahun sebesar 12,05%. Nilai tersebut hanya menunjukkan bahwa fitur tersebut paling banyak menurunkan impurity pada dataset dan konfigurasi model yang digunakan, bukan bahwa penghasilan rendah menyebabkan atau menentukan sektor usaha terbaik.

Untuk memperkuat interpretasi, hasil tersebut perlu divalidasi menggunakan permutation importance atau SHAP pada setiap fold validasi silang dan dilaporkan dalam bentuk rata-rata serta variasi antar-fold. Sebelum analisis tambahan tersebut dilakukan, peringkat fitur harus diperlakukan sebagai hasil eksploratif dan tidak digunakan untuk menarik kesimpulan kausal.

Aturan Keputusan

Struktur pohon keputusan menghasilkan aturan klasifikasi dari setiap jalur yang menghubungkan simpul akar dengan simpul daun. Setiap jalur memuat kombinasi kategori penghasilan, umur, jumlah tanggungan, pekerjaan, dan status usaha yang mengarah pada salah satu kelas label sektor UMKM. Model memiliki 41 simpul daun sehingga aturan lengkap cukup panjang untuk ditampilkan dalam artikel. Oleh karena itu, Tabel 5 menyajikan tujuh pola keputusan utama dalam bentuk kategori asli agar lebih mudah dibaca. Pada hasil pengodean, nilai lebih dari 0,50 menunjukkan bahwa responden termasuk dalam suatu kategori, sedangkan nilai kurang dari atau sama dengan 0,50 menunjukkan kondisi sebaliknya.

Tabel 5. Ringkasan Aturan Keputusan Model

No.	Kondisi Responden	Label Model
1	Penghasilan rendah, berumur 18–35 tahun, serta tidak memiliki tanggungan atau memiliki tanggungan sedikit	Fashion/Kerajinan
2	Penghasilan rendah serta memiliki tanggungan sedang atau banyak	Kuliner
3	Penghasilan sedang atau tinggi serta telah memiliki usaha	Jasa/Produk Kreatif
4	Berumur 18–35 tahun serta tidak memiliki tanggungan atau memiliki tanggungan sedikit	Fashion/Kerajinan
5	Mengurus rumah tangga dan memiliki penghasilan rendah	Kuliner
6	Bekerja sebagai perawat, karyawan swasta, atau guru honorer serta memiliki penghasilan sedang atau tinggi	Jasa/Produk Kreatif
7	Berumur 46–59 tahun serta memiliki tanggungan sedang atau banyak	Kuliner

Nilai coverage dihitung dari proporsi jumlah kasus pada suatu simpul daun terhadap seluruh data latih, sedangkan purity menunjukkan proporsi kelas mayoritas pada simpul tersebut. Tingkat kesalahan diperoleh dari 1–purity. Apabila hanya sebagian aturan ditampilkan pada naskah utama, pemilihannya harus menggunakan kriteria yang jelas, misalnya aturan dengan coverage terbesar, sedangkan seluruh 41 aturan disajikan pada lampiran. Prioritas tidak ditentukan secara manual karena setiap prediksi mengikuti urutan pemisahan model dari akar sampai daun. Interpretasi aturan hanya menjelaskan mekanisme model dalam mereproduksi label berbasis pembobotan dan tidak menunjukkan bahwa kombinasi atribut tersebut secara substantif menentukan sektor usaha terbaik.

Diskusi

Model Decision Tree menghasilkan akurasi 94,81%, presisi tertimbang 94,86%, recall tertimbang 94,81%,

dan F1-score tertimbang 94,75% pada 732 data uji. Namun, performa antar-kelas tidak seragam. Label Fashion/Kerajinan memperoleh recall 0,85 dengan 22 dari 151 data diprediksi sebagai kelas lain, sedangkan kelas Jasa/Produk Kreatif dan Kuliner masing-masing memperoleh recall 0,97. Temuan ini menunjukkan bahwa akurasi keseluruhan perlu dibaca bersama metrik per kelas agar kelemahan pengenalan pada Fashion/Kerajinan tidak tertutupi. Kedekatan akurasi data latih dan data uji juga tidak membuktikan bahwa model terbebas dari overfitting karena hasil tersebut hanya berasal dari satu pembagian data dan dapat dipengaruhi oleh variasi sampel.

Sebagai konteks, Suseno et al. (2023) membandingkan K-Nearest Neighbor dan C4.5 pada klasifikasi rekomendasi bantuan UMKM di Kabupaten Pekalongan. Dari 279 data hasil prapemrosesan, C4.5 menghasilkan akurasi 86,08%, sedangkan K-Nearest Neighbor mencapai 94,94% (Suseno et al., 2023). Novriandy dan Winarsih (2023) melaporkan akurasi C4.5 sebesar 90% dan Naive Bayes sebesar 70% pada klasifikasi kelayakan bantuan UMKM. Hasil kedua penelitian tersebut tidak dapat dibandingkan secara langsung dengan penelitian ini untuk menyatakan keunggulan model karena terdapat perbedaan jumlah data, atribut, definisi kelas, komposisi data, algoritma, dan skema pengujian. Perbandingan tersebut hanya menunjukkan bahwa pohon keputusan telah digunakan untuk membentuk klasifikasi bantuan UMKM secara terstruktur pada konteks penelitian yang berbeda (Novriandy & Winarsih, 2023).

Pada struktur model, kategori penghasilan rendah memperoleh kepentingan fitur berbasis impurity sebesar 43,11%, diikuti umur 18–25 tahun sebesar 12,05%. Nilai tersebut hanya menunjukkan bahwa kedua fitur menghasilkan penurunan impurity yang lebih besar pada dataset, pengodean, dan konfigurasi pohon yang digunakan. Hasil ini tidak menunjukkan hubungan sebab-akibat dan tidak membuktikan bahwa penghasilan atau umur secara substantif menentukan sektor usaha terbaik. Interpretasi perlu semakin dibatasi karena label target dibentuk melalui pembobotan atribut yang sama dengan prediktor, sehingga dominasi fitur tertentu dapat mencerminkan struktur aturan pembentukan label.

Keterbatasan utama penelitian ini terletak pada penggunaan label berbasis pembobotan yang belum berasal dari penilaian pakar independen atau keberhasilan usaha aktual. Dengan demikian, nilai evaluasi hanya menunjukkan kemampuan Decision Tree dalam mereproduksi label yang telah dibentuk, bukan validitas label sektor UMKM di dunia nyata. Model ini lebih tepat diposisikan sebagai bukti konsep klasifikasi berbasis aturan dan belum dapat digunakan sebagai dasar keputusan individual. Penelitian lanjutan perlu membentuk label melalui konsensus ahli atau outcome usaha, menerapkan penalaan hiperparameter dan validasi silang terstratifikasi berulang, melaporkan interval kepercayaan dan baseline, menguji sensitivitas pembobotan, serta melakukan validasi eksternal atau prospektif pada wilayah dan periode berbeda.

Kesimpulan

Penelitian ini menghasilkan proof-of-concept model Decision Tree untuk mereproduksi tiga label sintesis sektor UMKM, yaitu Kuliner, Fashion/Kerajinan, serta Jasa/Produk Kreatif, berdasarkan umur, pekerjaan, kategori status perkawinan, penghasilan, jumlah tanggungan, dan status usaha. Pengujian terhadap 732 data uji menghasilkan akurasi 94,81%, presisi tertimbang 94,86%, recall tertimbang 94,81%, dan F1-score tertimbang 94,75%. Hasil tersebut menunjukkan konsistensi model dalam mempelajari aturan pembobotan yang digunakan untuk membentuk label, tetapi belum membuktikan kesesuaian rekomendasi terhadap kebutuhan individu, keberhasilan usaha aktual, ataupun manfaat bagi pemerintah dan lembaga pemberdayaan. Oleh karena itu, model belum dapat digunakan sebagai dasar pengambilan keputusan operasional. Penelitian selanjutnya perlu membentuk ground truth melalui penilaian ahli atau outcome usaha aktual, melakukan penalaan dan validasi silang berulang, menguji kestabilan model, serta melaksanakan validasi eksternal, uji pengguna, dan evaluasi lapangan sebelum manfaat praktisnya dapat dinyatakan.

Ucapan Terima Kasih

Penulis menyampaikan terima kasih kepada Universitas Islam Negeri Sumatera Utara, dosen pembimbing, serta instansi terkait di Kabupaten Karo yang telah memberikan arahan, dukungan, dan akses data sehingga penelitian ini dapat diselesaikan dengan baik.

Pernyataan Konflik Kepentingan

Seluruh penulis menyatakan bahwa tidak terdapat konflik kepentingan yang berkaitan dengan pelaksanaan penelitian, akses dan penggunaan data, pengembangan model, interpretasi hasil, maupun proses publikasi artikel ini.

Daftar Pustaka

- Alie, A., Elanda, Y., & Wahyudi, R. (2024). Perangkap Kemiskinan Kepala Rumah Tangga Perempuan di Kota Surabaya. *Community: Pengawas Dinamika Sosial*, 10(1), 112–128. <https://doi.org/10.35308/jcpds.v10i1.9145>
- Avianto, D., & Wibowo, A. P. (2024). Pembentukan Pohon Keputusan untuk Penerima Bantuan Beras Miskin Menggunakan Algoritma Decision Tree C4.5. *NERO: Networking Engineering Research Operation*, 9(1), 59–68. <https://doi.org/10.21107/nero.v9i1.28020>
- Bachtari, L., & Mahradianur. (2023). Analisis Data Mining Menggunakan Metode Algoritma C4.5 Menentukan Penerima Bantuan Langsung Tunai. *Jurnal Informatika*, 10(1), 28–36. <https://doi.org/10.31294/inf.v10i1.15115>
- Badan Pusat Statistik. (2024a). *Profil Industri Mikro dan Kecil*. 2023. <https://www.bps.go.id/id/publication/2024/09/18/52d85cbe9de005b6f5d69f95/profil-industri-mikro-dan-kecil-2023.html>
- Badan Pusat Statistik Provinsi Sumatera Utara. (2024b). *Profil Industri Mikro dan Kecil Provinsi Sumatera Utara*. 2023.

<https://sumut.bps.go.id/id/publication/2024/12/31/3958146f1f61cb8ba1550b1d/profil-industri-mikro-dan-kecil-provinsi-sumatera-utara-2023.html>

- Hartana, & Mandasari, R. D. (2024). Implementasi Algoritma Decision Tree untuk Klasifikasi Kelayakan Distribusi Dana Bantuan Sosial. *Jurnal Responsif: Riset Sains dan Informatika*, 6(2), 240–246. <https://doi.org/10.51977/jti.v6i2.1730>
- Ikhsan, & Zulkifli. (2022). Pengaruh Pendapatan Perempuan terhadap Kemiskinan dan Ketimpangan Pendapatan: Bukti Data Panel di Aceh. *Jurnal EMT KITA*, 6(1), 184–190. <https://doi.org/10.35870/emt.v6i1.581>
- Kasim, S. S., Juhaepa, Supiyah, R., & Roslan, D. S. (2022). Etos Kerja dan Kemandirian Ekonomi Perempuan Kepala Rumah Tangga (Janda) pada Keluarga Miskin di Perkotaan. *Proceedings of the 1st Conference on Social, Politics, and Culture (IACS-CSPC 2022)*, 20–27. <https://ojs.literacyinstitute.org/index.php/iacseries/article/view/547>
- Latifah, L., Susilowati, N., & Anisykurillah, I. (2022). Penggunaan Media Sosial Networking dan Pelatihan dalam Meningkatkan Kinerja UMKM pada Entrepreneur Muda di Indonesia. *Jurnal Ekonomi dan Pendidikan*, 19(1), 14–24. <https://doi.org/10.21831/jep.v19i1.45323>
- Maharani, A. J., & Setiawati, S. D. (2024). Penggunaan Storytelling dalam Pemasaran Digital Pelaku UKM Perempuan. *Syntax Literate: Jurnal Ilmiah Indonesia*, 9(1), 265–273. <https://doi.org/10.36418/syntax-literate.v9i1.14836>
- Mahendra, W., & Wibowo, M. G. (2023). The Effect of Income Inequality, Women's Empowerment, Unemployment and Population Density on Poverty in Aceh Province. *JIIET: Jurnal Ilmu Ekonomi Terapan*, 8(1), 11–23. <https://doi.org/10.20473/jiet.v8i1.42803>
- Nofriadi, N., Elfiswandi, E., Rafki, R., & Lusiana, L. (2024). Analisis Hambatan dan Peluang Akses Pembiayaan bagi UMKM Perempuan: Studi Kasus Kota Padang. *Jurnal Akuntansi, Manajemen, dan Perencanaan Kebijakan*, 1(3), 1–10. <https://doi.org/10.47134/jampk.v1i3.203>
- Novriandy, A., & Winarsih. (2023). Implementasi Algoritma Naive Bayes dan Algoritma C4.5 dalam Klasifikasi Kelayakan Bantuan UMKM. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 4(1), 208–217. <https://doi.org/10.30865/klik.v4i1.1099>
- Paransa, R. P. J. (2024). Optimalisasi Pertumbuhan UMKM di Indonesia melalui Pemasaran Digital. *Efisiensi: Kajian Ilmu Administrasi*, 21(2), 164–181. <https://doi.org/10.21831/efisiensi.v21i2.77042>
- Pratiwi, N. W. O., Utami, N. W., & Putra, I. G. J. E. (2022). Klasifikasi Penentuan Penerima Bantuan Sosial Tunai (BST) Menggunakan Algoritma C4.5 di Desa Keramas, Gianyar, Bali. *Jurnal Informatika Teknologi dan Sains*, 4(3), 101–107. <https://doi.org/10.51401/jinteks.v4i3.1667>
- Susanti, E., Mulyanti, R. Y., & Wati, L. N. (2023). MSMEs Performance and Competitive Advantage: Evidence from Women's MSMEs in Indonesia. *Cogent Business & Management*, 10(2), 1–26. <https://doi.org/10.1080/23311975.2023.2239423>
- Suseno, A. T., Amin, M. Al, & Mahardika, F. (2023). Analisis Pemberian Bantuan UMKM Menggunakan Algoritma K-NN dan C4.5. *Journal of Computer System and Informatics (JoSYC)*, 5(1), 248–256. <https://doi.org/10.47065/josyc.v5i1.4312>
- Widia, E., & Octafia, S. M. (2022). Eksistensi UMKM Perempuan di Masa Krisis: Kajian Tantangan dan Peluang di Koto Tengah, Kota Padang. *JABE: Journal of Applied Business and Economic*, 9(2), 111–126. <https://doi.org/10.30998/jabe.v9i2.14622>
- Widia, E., & Octafia, S. M. (2023). The Role of Social Capital for the Development of Women's Enterprises: A Case Study of MSMEs in Padang City. *Optimum: Jurnal Ekonomi dan Pembangunan*, 13(2), 219–230. <https://doi.org/10.12928/optimum.v13i2.8693>
- Yulfa, R., Puspitawati, H., & Muflikhati, I. (2022). Tekanan Ekonomi, Coping Ekonomi, Dukungan Sosial, dan Kesejahteraan Perempuan Kepala Keluarga. *Jurnal Ilmu Keluarga dan Konsumen*, 15(1), 14–26. <https://doi.org/10.24156/jikk.2022.15.1.14>