JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10138
Hal 1509-1518
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Penerapan Algoritma Random Forest untuk Klasifikasi Risiko Usaha UMKM Berdasarkan Faktor Keuangan, Operasional, dan Pemasaran

M. Imam Santoso[1,*], Asha Sembiring[2], Alfan Ramadhan[2]

[1] Fakultas Ekonomi dan Bisnis, Program Studi Kewirausahaan, Universitas Satya Terra Bhinneka, Medan, Indonesia
[2] Fakultas Teknologi dan Ilmu Komputer, Program Studi Informatika, Universitas Satya Terra Bhinneka, Medan, Indonesia
Email: [1,*]mhdimamsantoso@satyaterrabhinneka.ac.id, [2]ashasembiring@satyaterrabhinneka.ac.id,
[3]alfanramadhans@satyaterrabhinneka.ac.id
Email Penulis Korespondensi: mhdimamsantoso@satyaterrabhinneka.ac.id[*]
Submitted **09-09-2026**; Accepted **30-10-2026**; Published **31-10-2026**

Abstrak

Usaha Mikro, Kecil, dan Menengah (UMKM) menghadapi risiko usaha yang berkaitan dengan faktor keuangan, operasional, dan pemasaran. Penelitian ini bertujuan mengklasifikasikan risiko usaha UMKM menjadi kategori risiko rendah dan tinggi menggunakan algoritma Random Forest. Data yang digunakan berupa dataset sintetis UMKM dari Kaggle dengan jumlah 150.000 data berdasarkan pembagian data yang dilaporkan. Tahapan prapemrosesan meliputi pemeriksaan nilai yang hilang, transformasi data, serta pembentukan target Risk dari atribut Class. Kategori Growth dan Elite dipetakan sebagai risiko rendah, sedangkan Struggling dan Critical dipetakan sebagai risiko tinggi. Dataset dibagi secara terstratifikasi menjadi 120.000 data latih (80%) dan 30.000 data uji (20%). Model menggunakan 100 pohon keputusan dan dievaluasi melalui akurasi, precision, recall, F1-score, serta confusion matrix. Pengujian awal menghasilkan akurasi sebesar 100%, dengan precision, recall, dan F1-score masing-masing sebesar 1,00 pada kedua kelas. Sebanyak 19.174 data risiko rendah dan 10.826 data risiko tinggi diklasifikasikan dengan benar. Namun, hasil tersebut belum dapat menunjukkan kemampuan generalisasi model karena atribut Class yang membentuk target masih digunakan sebagai prediktor, sehingga terjadi kebocoran informasi target. Oleh karena itu, klaim keberhasilan model belum dapat ditetapkan sebelum pelatihan dan pengujian ulang tanpa atribut Class. Penggunaan data sintetis juga membatasi penerapan temuan, sehingga diperlukan validasi pada data UMKM riil.

Kata Kunci: UMKM; risiko usaha; Random Forest; machine learning; klasifikasi.

Abstract

Micro, Small, and Medium-sized Enterprises (MSMEs) face business risks associated with financial, operational, and marketing factors. This study aims to classify MSME business risk into low- and high-risk categories using the Random Forest algorithm. The study uses a synthetic MSME dataset from Kaggle comprising 150,000 records, as inferred from the reported data split. Preprocessing includes checking for missing values, transforming the data, and constructing the target variable Risk from the Class attribute. Growth and Elite are mapped to low risk, whereas Struggling and Critical are mapped to high risk. The dataset is divided using stratified sampling into 120,000 training records (80%) and 30,000 test records (20%). The model uses 100 decision trees and is evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. The initial evaluation yields an accuracy of 100%, with precision, recall, and F1-score of 1.00 for both classes. All 19,174 low-risk records and 10,826 high-risk records are correctly classified. However, these results do not establish the model's generalization ability because the Class attribute, which defines the target, remains among the predictors, resulting in target leakage. Therefore, claims of model effectiveness cannot be established until the model is retrained and reevaluated without the Class attribute. The use of synthetic data also limits the applicability of the findings, requiring further validation using real-world MSME data.

Keywords: MSMEs; business risk; Random Forest; machine learning; classification.

1. PENDAHULUAN

Usaha Mikro, Kecil, dan Menengah (UMKM) memiliki peran yang sangat penting dalam perekonomian, khususnya di Indonesia. UMKM berkontribusi besar dalam penyerapan tenaga kerja, peningkatan pendapatan masyarakat, serta pemerataan ekonomi di berbagai daerah. Keberadaan UMKM juga menjadi salah satu pilar utama dalam menjaga stabilitas ekonomi nasional, terutama dalam menghadapi berbagai tantangan global dan krisis ekonomi. [1]

Meskipun memiliki peran yang strategis, UMKM tidak terlepas dari berbagai permasalahan yang dapat memengaruhi keberlangsungan usahanya. Banyak pelaku UMKM menghadapi keterbatasan modal, rendahnya efisiensi operasional, serta kurang optimalnya strategi pemasaran. Kondisi ini menyebabkan tingkat keberhasilan usaha menjadi tidak stabil dan berpotensi menimbulkan risiko yang cukup tinggi [2].

Risiko usaha pada UMKM dapat berasal dari berbagai faktor, di antaranya faktor keuangan, operasional, dan pemasaran. Faktor keuangan mencakup kemampuan dalam mengelola arus kas, menghasilkan laba, serta mengendalikan tingkat hutang. Faktor operasional berkaitan dengan efisiensi proses bisnis, kapasitas produksi, dan pengalaman usaha. Sementara itu, faktor pemasaran meliputi kemampuan dalam menarik pelanggan, mempertahankan loyalitas konsumen, serta pemanfaatan teknologi digital dalam promosi [3] [4].

Pengelolaan risiko usaha yang tepat menjadi hal yang sangat penting bagi pelaku UMKM agar dapat mempertahankan dan mengembangkan bisnisnya. Namun, dalam praktiknya, penilaian risiko seringkali masih dilakukan secara subjektif dan belum berbasis data yang kuat. Hal ini menyebabkan pengambilan keputusan menjadi kurang optimal dan berpotensi menimbulkan kerugian [5] [6].

Copyright © 2026 The **M. Imam Santoso**, Page 1509
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10138
Hal 1509-1518
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Seiring dengan perkembangan teknologi, pendekatan berbasis data melalui machine learning menjadi salah satu solusi yang dapat digunakan untuk membantu dalam pengambilan keputusan. Machine learning mampu mengolah data dalam jumlah besar untuk menemukan pola tertentu dan menghasilkan prediksi yang lebih akurat. Salah satu metode yang banyak digunakan dalam klasifikasi adalah algoritma Random Forest, yang memiliki keunggulan dalam menangani data dengan banyak variabel serta mampu mengurangi overfitting [7].

Penelitian terdahulu telah menerapkan algoritma Random Forest untuk memprediksi kategori penjualan pada usaha rumah makan [8]. Fokus tersebut berbeda dari klasifikasi risiko usaha karena kategori penjualan belum secara langsung menggambarkan kondisi usaha berdasarkan gabungan faktor keuangan, operasional, dan pemasaran. Selain itu, pemanfaatan kecerdasan buatan dalam pengambilan keputusan bisnis UMKM telah menjadi perhatian penelitian sebelumnya [9]. Dalam konteks tersebut, penelitian ini berfokus pada klasifikasi risiko usaha dengan menggabungkan indikator margin laba bersih, rasio pengeluaran terhadap pemasukan, aktivitas transaksi, penilaian pelanggan, tingkat pembelian ulang, dan adopsi teknologi digital. Pembeda penelitian ini terletak pada kombinasi atribut lintas aspek usaha dan pembentukan target klasifikasi biner pada dataset sintetis UMKM. Kategori Growth dan Elite dipetakan sebagai risiko rendah, sedangkan Struggling dan Critical dipetakan sebagai risiko tinggi. Kontribusi penelitian diarahkan pada penerapan dan evaluasi Random Forest untuk definisi risiko tersebut, bukan pada pengembangan algoritma baru. Definisi risiko yang digunakan terbatas pada pemetaan kategori dalam dataset dan belum merupakan ukuran risiko usaha yang tervalidasi pada UMKM riil.

Berdasarkan ruang lingkup tersebut, penelitian ini bertujuan mengevaluasi kemampuan Random Forest dalam mengklasifikasikan risiko usaha UMKM serta mengidentifikasi atribut yang berkontribusi terhadap prediksi model. Kontribusi yang dituju mencakup penyajian pemetaan target yang eksplisit, evaluasi menggunakan akurasi, precision, recall, F1-score, dan confusion matrix, serta analisis kepentingan fitur keuangan, operasional, dan pemasaran. Evaluasi yang valid mensyaratkan pemisahan atribut pembentuk target dari fitur prediktor untuk menghindari kebocoran informasi. Hasil penelitian diharapkan menjadi dasar pengembangan kajian klasifikasi risiko UMKM, dengan tetap memerlukan pengujian pada data riil sebelum digunakan untuk mendukung pengambilan keputusan usaha.

2. METODOLOGI PENELITIAN

2.1 Jenis dan Sumber Data

Penelitian ini menggunakan data sekunder berupa dataset sintetis berjudul *UMKM Dataset* yang dibuat dan diunggah oleh zkyfauzi pada platform Kaggle (https://www.kaggle.com/datasets/zkyfauzi/umkm-dataset). Dataset ini berisi informasi mengenai variabel-variabel yang merepresentasikan kondisi UMKM dari sisi keuangan, operasional, dan pemasaran. Karena dataset dibangkitkan secara sintetis dan bukan berasal dari pencatatan UMKM yang sebenarnya, hasil penelitian ini dibatasi pada evaluasi kinerja model terhadap dataset tersebut dan belum dapat digeneralisasi sebagai bukti penerapan pada UMKM nyata. Atribut utama yang digunakan dalam penelitian ini berada pada tabel dibawah ini

Tabel 1. Atribut

No	Atrbut	Keterangan
1	ID	Identitas unik tiap data
2	Monthly_Revenue	Pendapatan bulanan
3	Net_Profit_Margin (%)	Persentasi keuntungan bersih
4	Burn_Rate_Ratio	Rasio pengeluaran terhadap pemasukan
5	Transaction_Count	Jumlah transaksi
6	Avg_Historical_Rating	Rata-rata rating pelanggan
7	Review_Text	Teks ulasan pelanggan
8	Busines_Tenure_Months	Lama usaha (dalam bulan)
9	Repeat_Order_Rate (%)	Persentase pelanggan yang repeat order
10	Digital_Adoption_Score	Tingkat adopsi teknologi digital
11	Peak_Hour_Latency	Waktu respon saat jam sibuk
12	Location_Competitivness	Tingkat persaingan lokasi usaha
13	Sentiment_Score	Skor sentiment pelanggan
14	Class	Label kondisi usaha

Tujuan utama dari penelitian ini adalah untuk mengklasifikasi risiko usaha UMKM berdasarkan atribut-atribut diatas, kedalam dua kategori risiko, yaitu risiko rendah dan risiko tinggi. Klasifikasi ini digunakan sebagai veariabel target dalam penerapan algoritma random forest, yang diharapakan mampu mengindentifikasi pola dan hubungan antar semua

Copyright © 2026 The **M. Imam Santoso**, Page 1510
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10138
Hal 1509-1518
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

atribut dalam menentukan risiko usaha UMKM secara akurat dan pasti serta memberikan pengetahuan yang bermanfaat bagi pelaku UMKM yang ingin membuka usaha.

2.2 Tahapan Penelitian

a. *Data Collection*
Teknik pengumpulan data pada penelitian ini menggunakan data sekunder yang diperoleh dari platform Kaggle berupa dataset sintetis UMKM [10]. Dataset tersebut berisi berbagai variabel yang merepresentasikan kondisi usaha, meliputi faktor keuangan, operasional, dan pemasaran, seperti pendapatan, laba, arus kas, jumlah pelanggan, efektivitas pemasaran, dan efisiensi operasional [11]. Data dipilih karena telah tersusun secara terstruktur dalam format tabular sehingga siap digunakan untuk proses analisis machine learning. Penggunaan data sekunder ini memungkinkan peneliti memperoleh data yang relevan dan memadai tanpa harus melakukan survei atau wawancara langsung kepada pelaku UMKM [12]. Selanjutnya, data tersebut digunakan sebagai dasar dalam proses preprocessing, pelatihan model, dan evaluasi algoritma Random Forest untuk mengklasifikasikan tingkat risiko usaha UMKM.

b. Preprocessing
Preprocessing data dilakukan untuk mempersiapkan dataset agar konsisten dan siap digunakan dalam pembentukan model machine learning[13]. Tahap awal meliputi pemeriksaan nilai yang hilang (missing value) serta penghapusan kolom ID dan Review_Text yang tidak digunakan sebagai fitur prediktor dalam penelitian ini. Selanjutnya, variabel target Risk dibentuk berdasarkan label pada kolom Class. Kategori Growth dipetakan menjadi Risk = 0 yang diinterpretasikan sebagai risiko rendah, sedangkan kategori Struggling dipetakan menjadi Risk = 1 yang diinterpretasikan sebagai risiko tinggi[14]. Dengan demikian, Risk merupakan pengodean ulang label Class, bukan ukuran risiko yang diperoleh melalui pengukuran independen.

Setelah variabel Risk terbentuk, kolom Class dikeluarkan dari kumpulan fitur input karena merupakan sumber pembentukan target. Langkah ini dilakukan untuk mencegah kebocoran informasi target (target leakage), yang dapat menyebabkan hasil evaluasi model terlihat lebih tinggi daripada kemampuan prediksinya. Dataset kemudian dipisahkan menjadi fitur prediktor (X) dan variabel target (y). Fitur prediktor mencakup atribut yang digunakan dalam pemodelan, tanpa menyertakan ID, Review_Text, Class, maupun Risk, sedangkan variabel target hanya berupa Risk. Pemetaan Growth = 0 dan Struggling = 1 digunakan secara konsisten pada tahap pelatihan, pengujian, dan interpretasi hasil. Melalui tahapan ini, dataset disiapkan untuk pemodelan klasifikasi label risiko pada data sintetis UMKM [15].

c. Random Forest
Pada tahap ini, peneliti memakai algoritma Random Forest sebagai metode utama untuk melakukan klasifikasi risiko usaha UMKM. Algoritma Random Forest merupakan metode klasifikasi berbasis ensemble learning yang bekerja dengan membangun sejumlah pohon keputusan (decision tree) dari subset data yang berbeda secara acak.
Secara matematis Random Forest dirumuskan sebagai berikut: [16]

$$y = mode\ \{T1(x), T2(x), \dots, Tb(x)\} \quad (1)$$

Keterangan :

Y	: Hasil prediksi akhir dari algoritma Random Forest.
Mode	: Nilai yang paling sering muncul (suara terbanyak) dari seluruh hasil prediksi pohon keputusan
T1(x)	: Hasil prediksi dari pohon keputusan pertama.
T2(x)	: Hasil prediksi dari pohon keputusan kedua.
Tb(x)	: Hasil prediksi dari pohon keputusan ke-B.
x	: Data input atau data UMKM yang akan diklasifikasikan.
B	: Jumlah total pohon keputusan (decision tree) yang dibangun dalam Random Forest.
Tb (x)	: Prediksi yang dihasilkan oleh pohon keputusan ke-b, dengan b=1,2,3,…,B.

d. Training
Tahap pelatihan diawali dengan pembagian dataset menjadi data latih dan data uji menggunakan metode train-test split dengan proporsi 80:20. Pembagian dilakukan dengan stratifikasi berdasarkan variabel Risk untuk mempertahankan proporsi kelas pada kedua kelompok data.

Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengevaluasi kemampuan klasifikasi terhadap data yang tidak digunakan selama pelatihan [17]. Variabel target Risk dibentuk melalui pemetaan label Class, yaitu Growth dan Elite menjadi Risk = 0 (risiko rendah), sedangkan Struggling dan Critical menjadi Risk = 1 (risiko tinggi). Pemetaan tersebut merupakan pengelompokan label kondisi usaha pada dataset sintetis menjadi dua kategori risiko. Atribut prediktor numerik digunakan sesuai nilai numeriknya, sedangkan ID dan Review_Text tidak digunakan sebagai prediktor. Model Random Forest dibangun menggunakan 100 pohon keputusan (n_estimators = 100). Pembentukan pohon dilakukan melalui bootstrap sampling dan pemilihan sebagian fitur secara acak pada setiap pemisahan simpul untuk menghasilkan keragaman antarpohon [18]. Konfigurasi 100 pohon merupakan konfigurasi yang digunakan dalam eksperimen ini dan belum dapat dinyatakan sebagai konfigurasi optimal karena naskah belum menyajikan perbandingan jumlah pohon maupun hasil optimasi hiperparameter. Evaluasi yang dilaporkan menggunakan pembagian data latih dan data uji, sedangkan prosedur serta hasil validasi silang belum didokumentasikan. Nilai

Copyright © 2026 The **M. Imam Santoso**, Page 1511
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10138
Hal 1509-1518
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

random_state pada pembagian data dan pembentukan model juga belum tercatat dalam naskah, sehingga reproduksibilitas eksperimen masih memerlukan verifikasi terhadap kode pelaksanaan.

Pemeriksaan hasil menunjukkan adanya ketidaksesuaian antara uraian preprocessing dan keluaran model. Kolom Class dinyatakan telah dihapus pada tahap preprocessing, tetapi masih tercantum dalam hasil feature importance. Karena Class merupakan sumber pembentukan Risk, penyertaannya sebagai prediktor menyebabkan kebocoran informasi target. Oleh karena itu, hasil evaluasi yang dilaporkan belum dapat digunakan untuk menyimpulkan kemampuan generalisasi model. Pelatihan dan pengujian ulang perlu dilakukan dengan memastikan Class dan Risk tidak terdapat dalam fitur prediktor, sementara Risk hanya digunakan sebagai target. Hasil pengujian ulang tersebut menjadi dasar pelaporan accuracy, precision, recall, F1-score, confusion matrix, dan feature importance yang telah diperbarui.

e. Evaluasi

Dalam penelitian ini, proses evaluasi dilakukan menggunakan beberapa metrik kinerja, yaitu accuracy, precision, recall, F1-score, dan confusion matrix. Nilai accuracy menunjukkan persentase jumlah prediksi yang benar dibandingkan dengan keseluruhan data pengujian. Semakin tinggi nilai akurasi, maka semakin baik kemampuan model dalam mengklasifikasikan tingkat risiko usaha UMKM. Precision digunakan untuk mengukur tingkat ketepatan model dalam mengidentifikasi data yang termasuk ke dalam suatu kategori risiko tertentu, sedangkan recall digunakan untuk mengukur kemampuan model dalam menemukan seluruh data yang benar-benar termasuk dalam kategori tersebut. Sementara itu, F1-score merupakan rata-rata harmonis antara precision dan recalls yang memberikan gambaran performa model secara lebih seimbang, terutama ketika distribusi jumlah data pada masing-masing kelas tidak sama. [19]

Selain menggunakan metrik evaluasi tersebut, penelitian ini juga memanfaatkan confusion matrix untuk menganalisis hasil klasifikasi secara lebih rinci. Confusion matrix menampilkan jumlah prediksi yang benar maupun salah pada setiap kategori risiko sehingga dapat diketahui tingkat keberhasilan model dalam membedakan UMKM yang memiliki risiko rendah dan risiko tinggi. [20] Melalui hasil evaluasi ini, dapat diketahui tingkat efektivitas algoritma Random Forest dalam melakukan klasifikasi risiko usaha serta kemampuannya dalam mendukung pengambilan keputusan yang berbasis data. Secara keseluruhan, tahap evaluasi berfungsi untuk memastikan bahwa model yang telah dibangun memiliki kemampuan generalisasi yang baik terhadap data baru. Hasil evaluasi yang diperoleh menjadi dasar dalam menentukan apakah model Random Forest layak digunakan sebagai alat bantu untuk mengidentifikasi dan memprediksi tingkat risiko usaha UMKM secara akurat dan konsisten.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengumpulan dan Analisis Data (*Data Collection*)

Pada penelitian ini, proses pengumpulan data dilakukan menggunakan data sekunder yang diperoleh dari platform Kaggle dalam bentuk dataset sintetis UMKM. Dataset tersebut dirancang untuk merepresentasikan kondisi usaha mikro, kecil, dan menengah berdasarkan berbagai faktor yang memengaruhi tingkat risiko usaha, seperti aspek keuangan, operasional, dan pemasaran. Penggunaan data sekunder dipilih karena data telah tersedia dalam format terstruktur sehingga memudahkan proses analisis dan pengembangan model machine learning.

Dataset yang digunakan dalam penelitian ini diperoleh dari platform Kaggle dan terdiri atas sejumlah atribut yang menggambarkan kondisi usaha UMKM. Variabel yang digunakan meliputi *Monthly_Revenue, Net_Profit_Margin, Burn_Rate_Ratio, Transaction_Count, Avg_Historical_Rating, Review_Text, Review_Volatility, Business_Tenure_Months, Repeat_Order_Rate, Digital_Adoption_Score, Peak_Hour_Latency, Location_Competitiveness, dan Sentiment_Score*. Variabel-variabel tersebut digunakan sebagai atribut prediktor dalam proses klasifikasi. Sementara itu, atribut Class digunakan sebagai variabel target yang menunjukkan kondisi usaha, yaitu Growth dan Struggling. Untuk keperluan pemodelan machine learning, atribut target ditransformasikan menjadi variabel Risk, dengan nilai 0 merepresentasikan risiko rendah dan nilai 1 merepresentasikan risiko tinggi. Data yang telah dikumpulkan selanjutnya digunakan pada tahap preprocessing, pelatihan model, dan evaluasi menggunakan algoritma Random Forest.

Data Awal:

	ID	Monthly_Revenue	Net_Profit_Margin (%)	Burn_Rate_Ratio	\
0	1	6680716	22.72	0.811	
1	2	5819101	4.46	0.968	
2	3	5236404	-10.12	1.047	
3	4	8043552	0.04	0.969	
4	5	6071256	4.22	0.954	

	Transaction_Count	Avg_Historical_Rating	\
0	161	4.75	
1	104	4.21	
2	102	3.51	
3	99	4.33	
4	115	4.34	

	Review_Text	Review_Volatility	\
0	Transaksi digital lancar, proses checkout tida...	0.313	
1	Harga dan kualitas seimbang, pengalaman biasa ...	0.632	
2	Pelayanan standar, masih bisa ditingkatkan.	0.470	
3	Transaksi digital lancar, proses checkout tida...	0.206	
4	Selalu repeat order karena kualitasnya terjaga...	0.232	

	Business_Tenure_Months	Repeat_Order_Rate (%)	Digital_Adoption_Score	\
0	105	19.40	4.24	
1	95	14.87	1.27	
2	17	21.00	3.37	
3	109	30.62	5.41	
4	74	20.87	2.67	

	Peak_Hour_Latency	Location_Competitiveness	Sentiment_Score	Class
0	Low	9	-0.25	Growth
1	Med	10	0.00	Growth
2	Med	8	0.00	Struggling
3	Low	13	-0.25	Growth
4	Low	7	0.25	Growth

Copyright © 2026 The **M. Imam Santoso**, Page 1512
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10138
Hal 1509-1518
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Gambar 1. *Load Data*

Berdasarkan data awal, dataset memiliki atribut yang relevan untuk menggambarkan kondisi usaha UMKM dan dapat digunakan untuk mengklasifikasikan tingkat risiko usaha berdasarkan kategori Growth dan Struggling.

3.2 Hasil Data *Preprocessing*

Tahapan preprocessing data dilakukan untuk memastikan kualitas dan kesiapan data sebelum digunakan dalam proses pemodelan menggunakan algoritma Random Forest. Proses ini mencakup pembersihan data (data cleaning), penyesuaian tipe data, transformasi variabel kategorikal melalui teknik encoding, serta pembentukan variabel target yang akan digunakan dalam klasifikasi. Selain itu, dilakukan pemeriksaan terhadap keberadaan nilai yang hilang (missing values) pada seluruh atribut. Hasil pemeriksaan menunjukkan bahwa tidak terdapat nilai yang hilang pada dataset, sehingga tahapan penanganan missing values tidak diperlukan. Hal ini memperlihatkan bahwa dataset *Syntetic UMKM Data* sudah memiliki kualitas data yang baik dan siap untuk di analisis.

```
Tipe Data Setelah Encoding:
Monthly_Revenue              int64
Net_Profit_Margin (%)      float64
Burn_Rate_Ratio            float64
Transaction_Count            int64
Avg_Historical_Rating      float64
Review_Volatility          float64
Business_Tenure_Months       int64
Repeat_Order_Rate (%)      float64
Digital_Adoption_Score     float64
Peak_Hour_Latency            int64
Location_Competitiveness     int64
Sentiment_Score            float64
Class                        int64
Risk                         int64
dtype: object
```

Gambar 2. Hasil Data *Preprocessing*

Langkah berikutnya adalah melakukan transformasi data agar seluruh atribut yang digunakan dalam proses klasifikasi dapat diproses oleh model machine learning. Pada dataset UMKM, atribut yang digunakan meliputi Monthly_Revenue, Net_Profit_Margin, Burn_Rate_Ratio, Transaction_Count, Avg_Historical_Rating, Review_Text, Review_Volatility, Business_Tenure_Months, Repeat_Order_Rate, Digital_Adoption_Score, Peak_Hour_Latency, Location_Competitiveness, dan Sentiment_Score. Atribut-atribut numerik seperti Monthly_Revenue, Net_Profit_Margin, Burn_Rate_Ratio, Transaction_Count, Avg_Historical_Rating, Business_Tenure_Months, Repeat_Order_Rate, Digital_Adoption_Score, Peak_Hour_Latency, Location_Competitiveness, dan Sentiment_Score dapat langsung digunakan dalam proses pemodelan. Sementara itu, atribut Review_Text yang berupa teks tidak digunakan secara langsung sebagai prediktor dalam proses klasifikasi. Atribut Class digunakan sebagai dasar pembentukan variabel target Risk.

Selanjutnya dilakukan pembentukan label target (Risk) untuk mengklasifikasikan kondisi usaha ke dalam dua kategori risiko. Berdasarkan pemetaan yang digunakan dalam penelitian ini, nilai Growth dan Elite dikategorikan sebagai Risiko Rendah dengan nilai Risk = 0, sedangkan Struggling dan Critical dikategorikan sebagai Risiko Tinggi dengan nilai Risk = 1. Dengan demikian, proses klasifikasi yang dilakukan termasuk dalam klasifikasi biner (binary classification), yaitu membedakan data UMKM ke dalam kategori risiko rendah dan risiko tinggi.

Variabel target Risk dibentuk dengan mengelompokkan empat kategori pada atribut Class menjadi dua kelas risiko. Kategori Growth dan Elite dipetakan menjadi Risk = 0 (risiko rendah), sedangkan kategori Struggling dan Critical dipetakan menjadi Risk = 1 (risiko tinggi). Dengan demikian, seluruh kategori sumber tercakup dalam pemetaan dan menghasilkan target klasifikasi biner. Pengelompokan risiko ini merupakan definisi operasional berdasarkan label kondisi usaha dalam dataset sintetis, bukan hasil pengukuran risiko secara langsung pada UMKM nyata. Rincian pemetaan disajikan pada tabel berikut.

Tabel 2. Rincian Pemetaan

Copyright © 2026 The **M. Imam Santoso**, Page 1513
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10138
Hal 1509-1518
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Nilai	Kategorikal	Keterangan
Growth	0	Risiko Rendah
Elite	0	Risiko Rendah
Struggling	1	Risiko Tinggi
Critical	1	Risiko Tinggi

Setelah proses transformasi dan pembentukan label target selesai, atribut yang tidak digunakan, seperti ID dan Review_Text, dihilangkan dari dataset. Selanjutnya, dataset dipisahkan menjadi atribut prediktor (X) dan variabel target (y), dengan Risk sebagai variabel target. Data kemudian dibagi menjadi dua bagian menggunakan metode train-test split, yaitu sebesar 80% untuk training set dan 20% untuk testing set. Pembagian data dilakukan dengan menggunakan stratify terhadap variabel Risk agar proporsi kategori risiko rendah dan risiko tinggi tetap terjaga pada data pelatihan maupun data pengujian. Data training digunakan untuk membangun model klasifikasi, sedangkan data testing digunakan untuk mengukur kemampuan model dalam melakukan prediksi terhadap data yang belum pernah digunakan selama proses pelatihan.

3.3 Hasil Training Model *Random Forest*

Setelah proses data preprocessing selesai, tahap selanjutnya adalah pelatihan model menggunakan algoritma Random Forest. Dataset yang telah dibagi menjadi 80% data latih (training set) dan 20% data uji (testing set) digunakan untuk membangun serta mengevaluasi model klasifikasi risiko UMKM ke dalam dua kategori, yaitu Risiko Rendah dan Risiko Tinggi. Model kemudian dievaluasi menggunakan metrik precision, recall, dan F1-score, serta akurasi untuk mengetahui kemampuan model dalam melakukan klasifikasi pada masing-masing kategori.

	precision	recall	f1-score	support
Risiko Rendah	1.00	1.00	1.00	19174
Risiko Tinggi	1.00	1.00	1.00	10826
accuracy			1.00	30000
macro avg	1.00	1.00	1.00	30000
weighted avg	1.00	1.00	1.00	30000

Gambar 3. Hasil Training Model *Random Forest*

Berdasarkan hasil evaluasi performa model Random Forest yang ditunjukkan pada tabel di atas, diperoleh nilai akurasi keseluruhan sebesar 1,00 atau 100% dengan jumlah data pengujian sebanyak 30.000 data. Hasil tersebut menunjukkan bahwa model mampu mengklasifikasikan seluruh data pengujian ke dalam kategori risiko yang sesuai berdasarkan hasil prediksi model. Pada kelas Risiko Rendah, model memperoleh nilai precision sebesar 1,00, recall sebesar 1,00, dan F1-score sebesar 1,00 dari total 19.174 data. Nilai tersebut menunjukkan bahwa model mampu mengenali data yang termasuk dalam kategori Risiko Rendah dengan sangat baik. Tidak terdapat kesalahan klasifikasi yang tercermin dari nilai precision dan recalls yang mencapai 1,00. Sementara itu, pada kelas Risiko Tinggi, model juga memperoleh nilai precision sebesar 1,00, recall sebesar 1,00, dan F1-score sebesar 1,00 dengan jumlah 10.826 data. Hasil tersebut menunjukkan bahwa model mampu mengidentifikasi seluruh data yang termasuk dalam kategori Risiko Tinggi secara tepat. Dengan demikian, berdasarkan hasil pengujian, tidak terdapat data pada kelas Risiko Tinggi yang salah diklasifikasikan oleh model. Nilai macro average yang terdiri dari precision, recall, dan F1-score masing-masing sebesar 1,00 menunjukkan bahwa model memiliki performa yang sama baiknya pada kedua kategori risiko tanpa adanya perbedaan performa antar kelas. Sementara itu, nilai weighted average yang juga mencapai 1,00 menunjukkan bahwa model mempertahankan performa yang sangat baik dengan mempertimbangkan jumlah data pada masing-masing kelas.

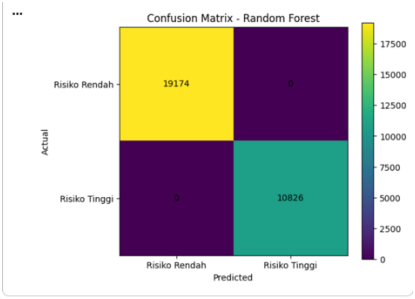
Hasil evaluasi awal menunjukkan akurasi sebesar 100% pada 30.000 data uji. Namun, hasil tersebut belum dapat dijadikan dasar untuk menyimpulkan kemampuan generalisasi model karena atribut Class masih tercantum dalam hasil feature importance, sementara atribut tersebut digunakan untuk membentuk variabel target Risk. Penyertaan Class sebagai prediktor memberikan informasi mengenai target kepada model dan menyebabkan target leakage. Oleh karena itu, nilai akurasi, precision, recall, F1-score, serta confusion matrix yang diperoleh perlu diverifikasi melalui pelatihan dan pengujian ulang setelah Class dikeluarkan dari fitur prediktor. Pemeriksaan juga perlu mencakup atribut lain yang secara langsung merepresentasikan target serta pemisahan data latih dan data uji. Sebelum verifikasi tersebut selesai, hasil ini diperlakukan sebagai hasil eksperimen awal yang mengandung kebocoran informasi target, bukan sebagai bukti keberhasilan klasifikasi berdasarkan faktor keuangan, operasional, dan pemasaran.

Copyright © 2026 The **M. Imam Santoso**, Page 1514
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10138
Hal 1509-1518
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

3.4 Hasil *Evaluation*

Setelah proses pelatihan model menggunakan algoritma Random Forest selesai dilakukan, tahap berikutnya adalah melakukan evaluasi untuk mengukur sejauh mana model mampu melakukan klasifikasi risiko UMKM dengan benar. Evaluasi dilakukan menggunakan Confusion Matrix, serta metrik performa berupa Accuracy, Precision, Recall, dan F1-score.

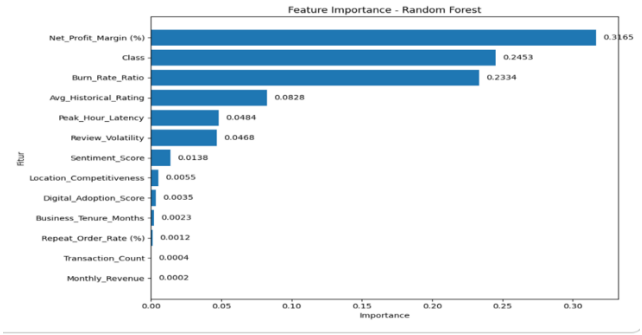


Gambar 4. *Hasil Evaluation*

Berdasarkan Confusion Matrix hasil klasifikasi menggunakan algoritma Random Forest, terlihat bahwa model memiliki performa yang sangat baik dalam membedakan kategori Risiko Rendah dan Risiko Tinggi. Dari total 30.000 data pengujian, sebanyak 19.174 data Risiko Rendah berhasil diklasifikasikan dengan benar sebagai Risiko Rendah, sedangkan 10.826 data Risiko Tinggi berhasil diklasifikasikan dengan benar sebagai Risiko Tinggi.

Pada kelas Risiko Rendah, model menghasilkan 19.174 prediksi yang benar dan tidak terdapat data yang salah diklasifikasikan sebagai Risiko Tinggi. Hal ini ditunjukkan dengan nilai pada diagonal sebesar 19.174, sedangkan nilai kesalahan klasifikasi sebesar 0. Kondisi tersebut menunjukkan bahwa model mampu mengenali seluruh data yang termasuk dalam kategori Risiko Rendah dengan tepat. Sementara itu, pada kelas Risiko Tinggi, model juga berhasil mengklasifikasikan seluruh 10.826 data dengan benar sebagai Risiko Tinggi. Tidak terdapat data Risiko Tinggi yang diprediksi sebagai Risiko Rendah, sehingga nilai kesalahan klasifikasi pada kelas tersebut juga sebesar 0. Hasil ini menunjukkan bahwa model mampu membedakan data Risiko Tinggi dari Risiko Rendah secara sempurna pada data pengujian. Secara keseluruhan, Confusion Matrix menunjukkan bahwa terdapat 19.174 True Negative (TN) dan 10.826 True Positive (TP), sedangkan False Positive (FP) dan False Negative (FN) masing-masing bernilai 0. Dengan demikian, seluruh 30.000 data pengujian berhasil diklasifikasikan dengan benar oleh model Random Forest.

Hasil tersebut sejalan dengan nilai evaluasi model yang memperoleh accuracy sebesar 1,00 atau 100%, serta nilai precision, recall, dan F1-score sebesar 1,00 pada kedua kelas. Dengan demikian, berdasarkan data pengujian yang digunakan, algoritma Random Forest menunjukkan kemampuan yang sangat baik dalam mengklasifikasikan UMKM ke dalam kategori Risiko Rendah dan Risiko Tinggi. Namun, hasil klasifikasi yang mencapai 100% perlu diperhatikan dan dianalisis lebih lanjut untuk memastikan tidak terdapat data leakage atau atribut yang secara langsung maupun tidak langsung merepresentasikan variabel target.



Gambar 5. Hasil Pengujian

Gambar di atas menunjukkan tingkat kepentingan (feature importance) masing-masing atribut yang digunakan dalam proses klasifikasi risiko UMKM menggunakan algoritma Random Forest. Nilai importance menunjukkan seberapa besar kontribusi relatif suatu atribut dalam membantu model menentukan klasifikasi antara Risiko Rendah dan Risiko Tinggi. Semakin besar nilai importance, semakin besar kontribusi atribut tersebut terhadap proses pengambilan keputusan model.

Berdasarkan hasil analisis, atribut Net_Profit_Margin (%) merupakan atribut yang memiliki tingkat kepentingan paling tinggi dengan nilai importance sebesar 0,3165. Hasil tersebut menunjukkan bahwa Net Profit Margin menjadi

Copyright © 2026 The **M. Imam Santoso**, Page 1515
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10138
Hal 1509-1518
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

atribut yang paling dominan dalam membantu model membedakan kondisi risiko usaha. Margin laba bersih menggambarkan kemampuan usaha dalam menghasilkan keuntungan setelah memperhitungkan biaya yang dikeluarkan, sehingga atribut ini menjadi salah satu indikator penting dalam menentukan kondisi usaha.

Analisis feature importance pada model awal menunjukkan bahwa atribut Class masih digunakan sebagai prediktor dengan nilai kepentingan sebesar 0,2453. Padahal, Class merupakan sumber pembentukan target Risk melalui pemetaan Growth dan Elite menjadi risiko rendah, serta Struggling dan Critical menjadi risiko tinggi. Dengan demikian, penggunaan Class sebagai fitur input menimbulkan kebocoran informasi target (target leakage), karena model menerima informasi yang secara langsung menentukan kelas yang diprediksi.

Temuan tersebut membatasi interpretasi seluruh hasil feature importance pada model awal. Nilai kepentingan Net_Profit_Margin (%), Burn_Rate_Ratio, dan atribut lainnya belum dapat digunakan untuk menyimpulkan faktor dominan dalam klasifikasi risiko tanpa informasi target. Nilai-nilai tersebut diperoleh dari model yang masih menyertakan Class, sehingga besaran dan urutan kepentingannya dapat berubah setelah model dilatih ulang. Oleh karena itu, perbaikan tidak cukup dilakukan dengan menghapus Class dari grafik atau uraian hasil.

Untuk memperoleh hasil yang dapat diinterpretasikan, Class perlu dikeluarkan dari fitur prediktor, sedangkan Risk hanya digunakan sebagai variabel target. Selanjutnya, pelatihan dan evaluasi model, termasuk perhitungan confusion matrix dan feature importance, perlu diulang. Grafik serta pembahasan kepentingan fitur kemudian harus diperbarui berdasarkan keluaran model tersebut. Interpretasi hasil dibatasi pada kontribusi prediktif atribut dalam dataset sintetis dan tidak menunjukkan hubungan sebab-akibat terhadap risiko usaha UMKM nyata.

Berdasarkan keseluruhan hasil tersebut, dapat diketahui bahwa kontribusi atribut dalam model Random Forest cukup bervariasi. Net_Profit_Margin (%), Class, dan Burn_Rate_Ratio menjadi tiga atribut dengan tingkat kepentingan tertinggi, sedangkan Monthly_Revenue, Transaction_Count, dan Repeat_Order_Rate (%) memberikan kontribusi yang relatif kecil berdasarkan nilai feature importance yang dihasilkan model. Dengan demikian, hasil feature importance menunjukkan bahwa faktor yang berkaitan dengan kinerja dan kondisi keuangan usaha, khususnya Net Profit Margin dan Burn Rate Ratio, memiliki kontribusi yang lebih besar dalam proses klasifikasi risiko UMKM dibandingkan beberapa atribut operasional dan pemasaran lainnya. Hasil ini dapat menjadi dasar untuk mengidentifikasi atribut yang paling dominan dalam membedakan UMKM yang termasuk Risiko Rendah dan Risiko Tinggi.

3.5 Pembahasan

Hasil pelatihan model Machine Learning menggunakan algoritma Random Forest menunjukkan performa yang sangat tinggi dalam mengklasifikasikan tingkat risiko usaha UMKM berdasarkan dataset Syntetic UMKM Data. Setelah dilakukan tahap data preprocessing dan pembagian dataset dengan rasio 80:20, model dilatih menggunakan sejumlah variabel prediktor yang merepresentasikan faktor keuangan, operasional, dan pemasaran. Pendekatan ini sesuai dengan rancangan penelitian yang menggunakan train-test split 80:20 dan algoritma Random Forest dengan 100 pohon keputusan (n_estimators = 100). Model Random Forest menghasilkan accuracy sebesar 1,00 atau 100%, dengan nilai precision, recall, dan F1-score sebesar 1,00 pada kategori Risiko Rendah maupun Risiko Tinggi. Hasil tersebut menunjukkan bahwa model mampu melakukan klasifikasi terhadap data pengujian dengan sangat baik. Dari total 30.000 data testing, sebanyak 19.174 data termasuk kategori Risiko Rendah dan 10.826 data termasuk kategori Risiko Tinggi. Seluruh data pada kedua kategori berhasil diklasifikasikan dengan benar oleh model.

Pada kategori Risiko Rendah, model memperoleh nilai precision sebesar 1,00, recall sebesar 1,00, dan F1-score sebesar 1,00 dari 19.174 data. Nilai tersebut menunjukkan bahwa seluruh data yang sebenarnya termasuk kategori Risiko Rendah berhasil dikenali dengan tepat. Sementara itu, kategori Risiko Tinggi juga memperoleh nilai precision, recall, dan F1-score masing-masing sebesar 1,00 dari 10.826 data. Hasil tersebut menunjukkan bahwa model mampu mengenali seluruh data Risiko Tinggi tanpa kesalahan klasifikasi. Hasil Confusion Matrix memperkuat temuan tersebut. Model menghasilkan 19.174 data Risiko Rendah yang diklasifikasikan benar sebagai Risiko Rendah dan 10.826 data Risiko Tinggi yang diklasifikasikan benar sebagai Risiko Tinggi. Tidak terdapat data Risiko Rendah yang diprediksi sebagai Risiko Tinggi maupun data Risiko Tinggi yang diprediksi sebagai Risiko Rendah. Dengan demikian, nilai False Positive (FP) dan False Negative (FN) masing-masing adalah 0. Hasil evaluasi ini menunjukkan bahwa pada data pengujian yang digunakan, Random Forest mampu membedakan kedua kategori risiko secara sempurna. Penggunaan confusion matrix memang ditujukan untuk melihat jumlah prediksi benar dan salah pada setiap kategori risiko.

Jika dilihat dari nilai macro average dan weighted average, seluruh metrik juga menghasilkan nilai 1,00. Nilai macro average sebesar 1,00 menunjukkan bahwa performa model tidak hanya baik pada kelas mayoritas, tetapi juga sama baiknya pada kedua kategori risiko. Sementara itu, weighted average sebesar 1,00 menunjukkan bahwa dengan mempertimbangkan jumlah data pada masing-masing kelas, model tetap menghasilkan performa klasifikasi yang sempurna. Analisis lebih lanjut melalui feature importance menunjukkan bahwa Net_Profit_Margin (%) merupakan atribut dengan kontribusi paling besar terhadap model, yaitu sebesar 0,3165. Atribut berikutnya adalah Class dengan nilai 0,2453 dan Burn_Rate_Ratio sebesar 0,2334. Selanjutnya, Avg_Historical_Rating memiliki nilai sebesar 0,0828, Peak_Hour_Latency sebesar 0,0484, dan Review_Volatility sebesar 0,0468. Sementara itu, Sentiment_Score, Location_Competitiveness, Digital_Adoption_Score, Business_Tenure_Months, Repeat_Order_Rate (%), Transaction_Count, dan Monthly_Revenue memiliki kontribusi yang lebih rendah terhadap model. Tingginya feature importance pada Net_Profit_Margin (%) menunjukkan bahwa kondisi keuntungan bersih menjadi salah satu faktor yang paling dominan dalam proses klasifikasi risiko UMKM. Semakin baik kemampuan usaha dalam menghasilkan keuntungan bersih, semakin besar informasi yang diberikan atribut tersebut kepada model dalam membedakan kondisi risiko usaha. Burn_Rate_Ratio yang juga memiliki nilai importance tinggi menunjukkan bahwa kondisi penggunaan atau

Copyright © 2026 The **M. Imam Santoso**, Page 1516
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10138
Hal 1509-1518
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

pengeluaran dana usaha turut memberikan kontribusi penting dalam menentukan klasifikasi risiko. Hal ini sejalan dengan rancangan penelitian yang memanfaatkan faktor keuangan, operasional, dan pemasaran untuk mengenali pola tingkat risiko usaha UMKM.

Interpretasi klasifikasi risiko dalam penelitian ini dibatasi pada label yang dibentuk dari dataset sintetis. Kategori risiko rendah dan risiko tinggi merupakan pengelompokan label kondisi usaha, sehingga tidak dapat langsung disamakan dengan tingkat risiko UMKM yang diukur secara empiris. Selain itu, feature importance menggambarkan kontribusi relatif atribut terhadap prediksi model, bukan hubungan sebab-akibat antara atribut tersebut dan risiko usaha. Penerapan pada UMKM nyata memerlukan validasi menggunakan data empiris yang mencerminkan keragaman karakteristik usaha serta kondisi keuangan, operasional, dan pemasaran.

Dengan demikian, penelitian ini menunjukkan potensi penggunaan Random Forest sebagai metode klasifikasi risiko UMKM berbasis data. Penelitian yang Anda susun juga sejalan dengan tujuan artikel, yaitu memanfaatkan faktor keuangan, operasional, dan pemasaran untuk mengidentifikasi tingkat risiko usaha UMKM dan mendukung pengambilan keputusan berbasis data.

4. KESIMPULAN

Penelitian ini menerapkan algoritma Random Forest untuk klasifikasi label risiko usaha UMKM pada dataset sintetis yang memuat atribut keuangan, operasional, dan pemasaran. Target Risk dibentuk dengan mengelompokkan kategori Growth dan Elite sebagai risiko rendah, serta Struggling dan Critical sebagai risiko tinggi. Namun, pemeriksaan hasil menunjukkan bahwa atribut Class, yang menjadi sumber pembentukan target, masih digunakan sebagai prediktor sehingga menimbulkan kebocoran informasi target. Oleh karena itu, hasil evaluasi dan feature importance yang diperoleh belum dapat menjadi dasar untuk menyimpulkan kinerja model maupun atribut yang paling dominan. Pelatihan dan pengujian ulang setelah mengeluarkan Class dari fitur prediktor merupakan perbaikan yang diperlukan dalam penelitian ini sebelum kesimpulan mengenai kemampuan model dapat ditetapkan. Penggunaan dataset sintetis juga membatasi generalisasi hasil terhadap UMKM nyata, sehingga penerapan praktis memerlukan validasi lebih lanjut menggunakan data empiris.

REFERENCES

[1] F. N. Aftitah, J. L. K, K. Hasanah dan N. L. H. F. M, "Pengaruh UMKMTerhadap Pertumbuhan Ekonomi Di Indonesia Pada Tahun 2023," *JKPIM: Jurnal Kajian dan Penalaran Ilmu Manajemen,* vol. 3, pp. 32-43, 2025.

[2] A. Nurafifah, A. Soleha dan I. Misra, "Peran Manajemen Keuangan Dalam Pengelolaan Umkm Untuk Meningkatkan Keberlanjutan Bisnis," *Opportunity Research and Community Service Journal,* vol. 3, pp. 18-41, 2025.

[3] Misna, A. P. Hafiz dan A. Budianto, "Analisis Manajemen Risiko Dalam Meningkatkan Pendapatan Usaha Pada Umkm Kerupuk Amplang Kecamatan Reteh Riau," *Jurnal Makesya,* vol. 5, pp. 18-38, 2025.

[4] N. D. Atmini, U. F. Hidayanti dan Y. N. Gunara, "Analisis Manajemen Risiko Keuangan padaUMKM Lela Florist Kota Semarang," *Jurnal Visi Manajemen,* vol. 10, pp. 215-226, 2024.

[5] Y. A. Rizky, S. I. Lestari, S. N. Putri dan Hendra, "Penilaian Potensi Investasi Usaha Rintisan (Start-up) melalui Pendekatan Aspek Pasar dan Manajemen Risiko," *An-Najah: Journal of Islamic Economics,* vol. 1, pp. 332-341, 2025.

[6] M. Gustin, I. I. Hawa, N. Safitri, E. Elisa dan R. Asytuti, "Analisis Manajemen Risiko dalam Upaya Meningkatkan Keberlangsungan Usaha di Tengah Ketidakpastian Ekonomi," *JKB Jurnal Kewirusahaan dan Bisnis,* vol. 8, no. 2, pp. 83-98, 2026.

[7] T. Susilawati dan A. B. Trisnawan, "Pemanfaatan Machine Learninguntuk Peningkatan Akurasi Sistem Pendukung Keputusan Prediktif," *JurnalUnitek,* vol. 18, pp. 303-312, 2025.

[8] F. N. Farida, A. Faqih dan S. E. Permana, "Penerapan Model Prediksi Penjualan Pada Usaha Rumah Makan Menggunakan Algoritma Random Forest," *JATI (Jurnal Mahasiswa Teknik Informatika),* vol. 9, pp. 5895-5902, 2025.

[9] H. Yuliastuti, F. Melawati, Iwadiah, E. Khairani dan R. Diana, "Pengaruh Penggunaan Artificial Intelligence Terhadap Pengambilan Keputusan Bisnis Pada Umkm," *Jurnal Lentera Bisnis,* vol. 15, pp. 4482-4492, 2026.

[10] B. N. Lestaria dan L. Riskaa, "Klasifikasi Keberhasilan Usaha Mikro, Kecil, dan Menengah (UMKM) di Indonesia Menggunakan Algoritma C4.5 dengan Pendekatan Data Mining," *Jurnal Mahkota Bisnis,* vol. 5, pp. 75-83, 2026.

[11] P. Periyadi, N. Maulida dan I. Muttaqin, "Pengaruh Pendapatan Biaya Serta Efesiensi Beban Operasional Dalam Mengukur Kinerja Keuangan Perusahaan(Studi Kasus Pada Cv. Intiga Harapan Sukses)," *Jurnal Publikasi Ilmu Manajemen,* vol. 4, pp. 63-74, 2025.

[12] S. Aisyah, Data Science Teori, Metode, dan Aplikasi, Yogyakarta: PT. Tren Digital Publishing, 2025.

[13] Z. Briliantaman, W. Adriyano dan A. B. Pratama, "Analisis Akurasi Algoritma Random Forest Regression dalam Memprediksi Harga Mobil Bekas," *Journal of Information Systems and Business Technology,* vol. 02, pp. 1001-1006, 2026.

Copyright © 2026 The **M. Imam Santoso**, Page 1517
This Journal is licensed under a Creative Commons Attribution 4.0 International License

[14] Y. A. Singgalen, "Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT)," *Jurnal Media Informatika Budidarma,* vol. 7, pp. 1551-1562, 2023.

[15] N. Umar, Data Mining Konsep dan Penerapannya, Jambi: Sonpedia, 2025.

[16] P. K. Sari dan R. R. Suryono, "Komparasi Algoritma Support Vector Machine Dan Random Forest Untuk Analisis Sentimenmetaverse," *Jurnal MNEMONIC,* vol. 7, pp. 31-39, 2024.

[17] R. Amelia, "Analisis Pengaruh Pembagian Data terhadap Kinerja Algoritma Naive Bayesdalam Prediksi Penyakit Diabetes pada Wanita," *POLYSCOPIA,* vol. 3, pp. 56-63, 2026.

[18] Febriansyah dan I. Susilawati, "Klasifikasi Tingkat Penghasilan Freelancer Menggunakan Algoritma Random Forest," *Journal Nasional Teknologi Komputer,* vol. 6, pp. 1158-1168, 2026.

[19] A. R. Hanum*, I. A. Zetha, J. N. Fajrina, R. A. Wulandari, S. C. Putri, S. P. Andina dan N. Yudistira, "Analisis Kinerja Algoritma Klasifikasi Teks Bert Dalam Mendeteksi Berita Hoaks," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK),* vol. 11, pp. 537-546, 2024.

[20] M. R. Romahdoni, Mengenal Algorita Random Forest Cara Mudah Menerapkan Machine Learning di Bisnis UMKM, Payakumbuh: Serasi Media Teknologi, 2025.

Copyright © 2026 The **M. Imam Santoso**, Page 1518
This Journal is licensed under a Creative Commons Attribution 4.0 International License