

## KLASIFIKASI ADOPSI BERBASIS KECERDASAN BUATAN PADA UMKM DI INDONESIA MENGGUNAKAN ALGORITMA RANDOM FOREST

M. Ihsan Dirgantara, Fakhri Sepriansyah<sup>✉</sup>, Nurly Izzatul Maula, Farhan Daffazka,  
Ken Ditha Tania, Alsella Meiriza

Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya, Palembang, Indonesia  
Email: [mfna321123@gmail.com](mailto:mfna321123@gmail.com)

DOI: <https://doi.org/10.46880/jmika.Vol10No1.pp35-44>

### ABSTRACT

*Micro, Small, and Medium Enterprises (MSMEs) play a strategic role in the Indonesian economy; however, digital transformation based on artificial intelligence (AI) remains a significant challenge. This study aims to classify AI adoption among MSMEs in Indonesia using the Random Forest algorithm and to identify the factors that influence it. The dataset was obtained from the Zenodo repository, consisting of questionnaire results regarding AI adoption in MSMEs. The research stages included data cleaning, encoding, splitting the data into training (80%) and testing (20%) sets, implementing the Random Forest algorithm, evaluation, and result analysis. The evaluation results show an accuracy of 80.3% with an ROC-AUC of 0.884. The weighted precision, recall, and F1-score values are 81.2%, 80.3%, and 80.4%, respectively. These evaluation results indicate that the Random Forest algorithm performs well on this dataset. Furthermore, the feature importance analysis revealed several influential variables in AI adoption among MSMEs, including strategic decision-making (10.9%), digital leadership (8.3%), and respondent position (7.8%). In conclusion, the implementation of the Random Forest algorithm demonstrates strong performance in classifying AI adoption among MSMEs in Indonesia and highlights key influential variables such as strategic decision-making, digital leadership, and respondent position.*

**Keyword:** Classification, MSMEs, Random Forest.

### ABSTRAK

*Usaha Mikro Kecil dan Menengah (UMKM) memiliki peran strategis dalam perekonomian Indonesia, namun transformasi digital berbasis kecerdasan buatan masih menjadi tantangan. Penelitian ini bertujuan mengklasifikasikan adopsi kecerdasan buatan pada UMKM di Indonesia menggunakan algoritma Random Forest serta mengidentifikasi faktor yang berpengaruh. Dataset didapatkan melalui repository zenodo berupa hasil kuesioner mengenai adopsi kecerdasan buatan pada UMKM. Tahapan penelitian meliputi pembersihan data, encoding, pembagian data latih (80%) dan uji (20%), implementasi algoritma random forest, evaluasi, dan analisis hasil. Hasil evaluasi menunjukkan akurasi 80,3% dengan ROC-AUC 0,884. Nilai precision, recall, dan F1-score weighted masing-masing 81,2%, 80,3%, dan 80,4%. Hasil evaluasi tersebut menunjukkan bahwa algoritma random forest bekerja dengan baik untuk dataset ini. Selanjutnya, pada analisis feature importance, terdapat beberapa variabel yang mempengaruhi dalam adopsi kecerdasan buatan pada UMKM diantaranya strategic decision making (10,9%), digital leadership (8,3%), dan Jabatan Responden (7,8%). Sehingga dapat disimpulkan bahwa implementasi algoritma random forest memiliki performa yang baik dalam klasifikasi adopsi kecerdasan buatan pada UMKM di Indonesia dan menunjukkan beberapa variabel yang berpengaruh seperti strategic decision making, digital leadership, jabatan.*

**Kata Kunci:** Klasifikasi, Random Forest, UMKM.

### PENDAHULUAN

UMKM (Usaha Mikro, Kecil dan Menengah) memiliki kontribusi yang signifikan dalam mendukung perkembangan perekonomian di Indonesia. Berdasarkan data yang dikeluarkan oleh kementerian keuangan pada tanggal 23 agustus 2023 UMKM ini memberikan sumbangan sebesar 60,5% terhadap Produk Domestik Bruto (PDB), mampu menyerap tenaga kerja hingga 96,9%, serta turut berkontribusi

dalam ekspor nonmigas sebesar 15,69% (Kementerian Keuangan Republik Indonesia, 2023). Sehingga keberlangsungan UMKM memiliki peran penting dalam mendorong kondisi ekonomi Indonesia kedepan. Semakin meningkatnya kebutuhan UMKM untuk beradaptasi dengan pasar yang berbasis data, pemahaman yang jelas mengenai jenis teknologi berbasis kecerdasan buatan yang telah diadopsi menjadi sangat penting (Hikmah et al., 2026). Pada



zaman transformasi digital yang terjadi sekarang mendorong UMKM untuk memanfaatkan adopsi teknologi berbasis kecerdasan buatan dalam mendukung proses bisnis agar dapat meningkatkan efektivitas kegiatan operasional, memperluas jangkauan pasar, serta memperkuat kemampuan bersaing (Cikhal Anwar et al., 2024).

Kecerdasan buatan telah berkembang menjadi teknologi yang dapat mengubah proses bisnis UMKM yang sering memiliki keterbatasan sumber daya, menjadi lebih efisien, memperkuat pengambilan keputusan, dan mendorong kinerja bisnis secara keseluruhan (Arroyabe et al., 2024). Penerapan kecerdasan buatan pada UMKM masih terkendala oleh beberapa hal diantaranya adalah keterbatasan infrastruktur teknologi, biaya implementasi, dan talenta digital (Putri, 2025). Sehingga kondisi tersebut menyebabkan adopsi kecerdasan buatan pada UMKM belum merata.

Berdasarkan *dataset* hasil kuesioner yang diperoleh peneliti mengenai adopsi kecerdasan buatan pada UMKM di Indonesia, terdapat beberapa variabel yang digunakan diantaranya variabel adopsi kecerdasan buatan, variabel kepemimpinan *digital*, dan variabel keputusan strategis. Pada *dataset* tersebut sebanyak 56,7% UMKM yang sudah mengadopsi kecerdasan buatan dan 43,3% lainnya belum menerapkan kecerdasan buatan. Pada hasil tersebut menunjukan bahwa faktor yang mempengaruhi adopsi kecerdasan buatan tidak bersifat tunggal (nonlinear). Sehingga dibutuhkan suatu metode yang dapat menemukan pola antar data dengan variabel yang bersifat biner (adopsi/tidak mengadopsi). Metode yang tepat dalam mengidentifikasi variabel biner adalah klasifikasi. Salah satu model yang bisa digunakan untuk pola *dataset* tersebut adalah *random forest*.

Oleh karena itu pada penelitian ini dilakukan klasifikasi adopsi penerapan kecerdasan buatan pada UMKM di Indonesia menggunakan algoritma *random forest*. Penelitian ini bertujuan untuk mengklasifikasikan adopsi teknologi berbasis kecerdasan buatan pada UMKM di Indonesia berdasarkan pada variabel adopsi kecerdasan buatan, kepemimpinan digital, dan keputusan strategis menggunakan algoritma *random forest*. Sehingga dapat mengetahui pola penerapan kecerdasan buatan pada UMKM di Indonesia.

## TINJAUAN PUSTAKA

Penelitian yang dilakukan oleh (Ardhani & Ditha Tania, 2025) mengenai prediksi *customer churn return* menggunakan algoritma *logistic regression*, SVM, *random forest*, *gradient boosting*, dan *Xgboost*

menghasilkan bahwa *random forest* dapat diandalkan dalam menangani data non linear dan akurat dalam prediksi *customer churn* dengan akurasi sebesar 0,96. Selain itu pada penelitian oleh (Alvines et al., 2025) tentang komparasi ridge Regression, *random forest*, dan *gradient boosting* untuk prediksi curah hujan harian di Sumatra Selatan menunjukkan *random forest* unggul (akurasi 85-92%) pada data nonlinear lokal. Studi ini relevan karena membuktikan *random forest* mampu menangkap pola kompleks dari data *timeseries* regional Indonesia). Kemudian dalam penelitian (Fahria et al., 2026) menunjukkan bahwa algoritma *random forest* memiliki kinerja terbaik dalam mengklasifikasikan data perilaku digital dengan tingkat akurasi 0,84, lebih tinggi dibandingkan XGBoost (0,78) dan CatBoost (0,80). Penelitian ini membuktikan bahwa *random forest* unggul dalam menangani data terstruktur serta mampu mengidentifikasi faktor-faktor utama yang mempengaruhi hasil klasifikasi. Temuan ini relevan dengan penelitian yang akan dilakukan, karena *random forest* dapat digunakan untuk mengklasifikasikan adopsi kecerdasan buatan berdasarkan faktor kepemimpinan digital dan keputusan strategis pada UMKM.

Berdasarkan pada penelitian diatas dapat dilihat bahwa algoritma *random forest* unggul dalam menangani *dataset nonlinear*, tetapi untuk penerapannya masih terbatas pada konteks prediksi cuaca, *customer churn*, dan klasifikasi perilaku digital. Data yang digunakan pada penelitian tersebut berfokus pada data *time series* dan perilaku individu. Dalam penelitian ini menggunakan *dataset* berbasis kuesioner yang mengukur variabel adopsi AI, keputusan strategis dan kepemimpinan digital. Model *random forest* yang digunakan dalam penelitian ini menerapkan kombinasi *hyperparameter tuning* berbasis *grid search* dan *feature selection* berbasis *threshold importance* untuk mengklasifikasikan adopsi kecerdasan buatan pada UMKM di Indonesia.

Selain itu perbedaan penelitian ini dengan ketiga penelitian terdahulu terletak pada objek, variabel, dan metode analisis yang digunakan. Penelitian (Ardhani dan Ditha Tania, 2025) berfokus pada faktor transaksional dalam prediksi *customer churn* dengan bantuan SHAP, namun tidak melakukan *hyperparameter tuning* maupun *feature selection*. Penelitian (Alvines et al., 2025) menggunakan *TimeSeriesSplit* dan analisis koefisien regresi pada data klimatologi, tetapi terbatas pada periode satu tahun. Penelitian (Fahria et al., 2026) membandingkan tiga algoritma dengan *parameter default* tanpa optimasi lanjutan pada data perilaku individual. Ketiga penelitian tersebut sama-sama menggunakan analisis

*feature importance*, namun variabel yang dianalisis bersifat teknis dan individual.

Berbeda dengan penelitian-penelitian tersebut, penelitian ini secara khusus mengeksplorasi faktor-faktor manajerial, yaitu kepemimpinan digital dan pengambilan keputusan strategis, sebagai fokus utama dalam analisis *feature importance* yang belum pernah dilakukan sebelumnya. Selain itu, penelitian ini menerapkan *hyperparameter tuning* berbasis *grid search*, *feature selection* berbasis *threshold importance* “5-fold cross-validation”. Oleh karena itu penelitian ini diharapkan memberikan kontribusi praktis dalam memahami hal yang mempengaruhi adopsi teknologi.

### Data Mining

*Data mining* merupakan suatu ilmu yang mencari informasi berupa pola data dalam database dengan cara statistik, kecerdasan buatan dan machine learning (Dwi Arya Putra et al., 2025). Informasi tersebut diperoleh melalui tahapan analisis yang cukup kompleks dengan memanfaatkan berbagai metode, seperti kecerdasan buatan (*artificial intelligence*), teknik statistik, matematika, serta *machine learning* dan metode lainnya (Sudarsono et al., 2021). *Data mining* dibagi menjadi dua kategori utama, yaitu prediktif dan deskriptif, kategori prediktif digunakan untuk memperkirakan nilai suatu atribut berdasarkan atribut lainnya, sedangkan kategori deskriptif bertujuan untuk menemukan pola atau hubungan yang terdapat di dalam (Ananda Mustari et al., 2024).

### Klasifikasi

Klasifikasi adalah suatu proses untuk menemukan variabel yang membedakan sebuah data ke dalam suatu kelas (Qotrunnada & Utomo, 2022). Tujuan dari proses ini adalah untuk mengidentifikasi pola maupun hubungan yang terdapat dalam data sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan atau melakukan prediksi (Rahayu et al., 2024). Klasifikasi termasuk ke dalam metode *supervised learning*, yaitu metode yang memerlukan data sampel yang telah diberi label sebelumnya untuk menemukan pola hubungan pada data. Dengan demikian, hasil yang diperoleh diharapkan dapat meningkatkan keandalan informasi yang dihasilkan dari data yang tersedia (Srirahayu & Setya Pribadie, 2023.).

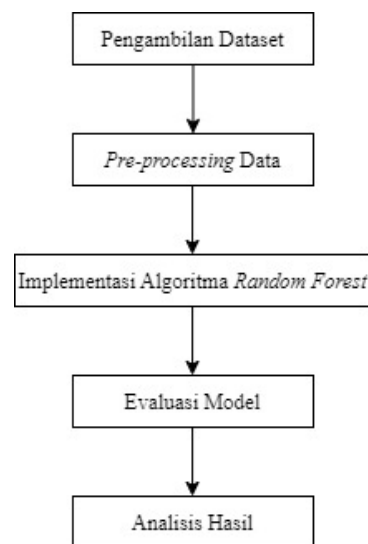
### Random Forest

*Random forest* merupakan salah satu metode dalam *machine learning* yang memanfaatkan kumpulan beberapa *Decision Tree* untuk melakukan proses klasifikasi. Penentuan hasil akhir dari metode ini

dilakukan dengan menggunakan mekanisme *majority voting* dari setiap pohon keputusan yang terbentuk (Larasati et al., 2022). Algoritma ini menerapkan banyak pohon dengan banyak perulangan sehingga mirip seperti hutan, tetapi algoritma ini membuat hasil dari prediksi semua sub pohon memiliki korelasi yang kecil (Yaman et al., 2024). Dalam algoritma random forest terdapat *feature importance* yang berfungsi untuk menghitung keterkaitan suatu variabel terhadap label yang ada (Ramadhan et al., 2026).

### METODE PENELITIAN

Penelitian ini dilaksanakan melalui serangkaian tahapan yang terstruktur untuk memperoleh hasil klasifikasi adopsi kecerdasan buatan pada UMKM. Tahapan tersebut diawali dengan pengambilan *dataset*, *pre-processing* data, kemudian implementasi algoritma *Random Forest*, evaluasi model menggunakan berbagai metrik, serta diakhiri dengan tahap analisis terhadap hasil yang diperoleh. Gambaran umum alur proses penelitian ini dapat dilihat pada gambar 1 berikut ini.



Gambar 1. Tahapan Penelitian

### Pengambilan Dataset

Pada tahap ini dilakukan pengambilan data yang relevan dengan tema penelitian. Peneliti mengambil data adopsi kecerdasan buatan pada UMKM di Indonesia pada platform zenodo (repository dataset online). Data tersebut dapat diakses pada *Digital Object Identifier*: <https://doi.org/10.5281/zenodo.18425107>.

### Pre-processing Data

#### 1. Pembersihan Data

Tahap awal yang dilakukan dalam *preprocessing* data adalah pembersihan data. proses ini dilakukan dengan cara menghapus data yang memiliki *missing value* seperti kolom yang kosong pada suatu baris.

Pembersihan data ditujukan untuk mengurangi pengaruh atribut yang memiliki tingkat relevansi rendah dalam proses klasifikasi (Irma Purnamasari & Ali, 2024).

## 2. Encoding Data

Setelah dilakukan pembersihan tahapan selanjutnya adalah *encoding dataset*. *encoding* merupakan proses Pengubahan Data Non-Numerik Menjadi Numerik Misalnya, mengonversi jenis kelamin (laki-laki atau perempuan) menjadi angka 0 dan 1 (Kuswanto et al., 2025).

## Implementasi Algoritma Random Forest

1. Split data, pada tahap ini *dataset* yang dimiliki dibagi menjadi dua, dimana 80% digunakan untuk data *training*, dan 20% digunakan sebagai *testing*.
2. Pelatihan model, Pelatihan melakukan pelatihan model dengan memanfaatkan data *training* untuk membuat model belajar dari pola-pola yang ada dalam data
3. Pengujian model, pada tahap ini dilakukan pengujian dengan menggunakan data *testing* yang sudah disiapkan.

## Evaluasi Model

1. *Confusion matrix* adalah tabel evaluasi yang digunakan untuk menilai kinerja model klasifikasi. Ini membandingkan hasil prediksi model dengan nilai sebenarnya dalam *dataset* pengujian. *Confusion matrix* memiliki 4 sel utama (Priatmojo et al., 2023).
  - a. *Positive* (TP): Kasus di mana model memprediksi kelas positif dengan benar.
  - b. *True Negative* (TN): Kondisi ketika model berhasil memprediksi kelas negatif secara tepat.
  - c. *False Positive* (FP): Keadaan ketika model memprediksi kelas positif, namun pada kenyataannya termasuk dalam kelas negatif.
  - d. *False Negative* (FN): Situasi ketika model memprediksi kelas negatif, padahal data tersebut sebenarnya termasuk dalam kelas positif.
2. Akurasi, Akurasi merupakan perbandingan antara jumlah prediksi yang benar, baik pada kelas positif maupun negatif, terhadap keseluruhan jumlah data

(Karyadiputra et al., 2022). Berikut rumus yang digunakan dalam perhitungan akurasi.

$$a = \frac{TP + TN}{TP + TN + FP + FN}$$

3. Presisi, merepresentasikan seberapa sering model memprediksi kelas positif dengan benar (Aqil Zidane et al., 2025). berikut rumus yang digunakan dalam menghitung presisi.

$$p = \frac{TP}{TP + FP}$$

4. *Recall*, didefinisikan sebagai persentase data yang relevan dan berhasil ditemukan oleh sistem dari keseluruhan *dataset*. Berikut rumus yang digunakan dalam menghitung *recall*.

$$r = \frac{TP}{TP + FN}$$

5. *F1-score*, merupakan metrik yang merangkum rata-rata nilai metrik *precision* dan metrik *recall* (Fatihaturrahmah & Tania, 2025). Berikut rumus yang digunakan dalam menghitung *F1 score*.

$$F1_{score} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

## Analisis Hasil

Pada tahap ini dilakukan interpretasi terhadap hasil penggunaan model *random forest* dalam klasifikasi. tujuan dari tahap ini adalah memahami hasil klasifikasi adopsi teknologi kecerdasan buatan pada UMKM.

## HASIL DAN PEMBAHASAN

### Pengambilan Dataset

Data yang digunakan dalam penelitian ini adalah adopsi kecerdasan buatan pada UMKM di Indonesia yang diambil dari repository *dataset online* zenodo. Data ini berupa kuesioner yang berisi tentang identitas UMKM, adopsi kecerdasan buatan pada perusahaan, pengaruh kepemimpinan digital, dan keputusan strategis. Data hasil kuesioner yang didapatkan berjumlah 314 data (jawaban responden). Berikut data hasil kuesioner adopsi kecerdasan buatan pada UMKM dapat dilihat pada gambar 2 berikut ini.

1	A	B	C	D	E	
1	Timestamp	Pernyataan Persetujuan: Denoan menosis kuesioner ini saya	1. Apa jabatan Anda di perusahaan?	2. Sektor Industri Mana perusahaan anda bergerak ?	3. Berapa Jumlah Karyawan Teta di Perusahaan Anda ?	4. Sudah i
2	9/29/2025 14:00:21	Saya setuju untuk berpartisipasi	Account executive	Ritel	50 - 249 (Usaha Menengah)	Di atas 10
3	11/25/2025 15:54:33	Saya setuju untuk berpartisipasi	Accounting	Distributor baut	10 - 49 (Usaha Kecil)	3 - 10 Tah
4	9/29/2025 9:24:55	Saya setuju untuk berpartisipasi	Admin	Manufaktur	50 - 249 (Usaha Menengah)	Di atas 10
5	9/29/2025 9:30:06	Saya setuju untuk berpartisipasi	Admin	Manufaktur	50 - 249 (Usaha Menengah)	< 3 tahun
6	11/25/2025 15:17:32	Saya setuju untuk berpartisipasi	admin	Jasa	50 - 249 (Usaha Menengah)	3 - 10 Tah
7	11/25/2025 15:28:50	Saya setuju untuk berpartisipasi	admin	industri kreatif (subsektor musik)	10 - 49 (Usaha Kecil)	Di atas 10
8	11/25/2025 15:43:19	Saya setuju untuk berpartisipasi	admin	Manufaktur	50 - 249 (Usaha Menengah)	3 - 10 Tah
9	11/25/2025 16:10:25	Saya setuju untuk berpartisipasi	admin	Manufaktur	10 - 49 (Usaha Kecil)	Di atas 10
10	11/25/2025 16:27:19	Saya setuju untuk berpartisipasi	admin	Manufaktur	50 - 249 (Usaha Menengah)	< 3 tahun
11	11/25/2025 16:57:19	Saya setuju untuk berpartisipasi	Admin	Jasa	Kurang dari 10 (Usaha Mikro)	3 - 10 Tah
12	11/25/2025 18:54:55	Saya setuju untuk berpartisipasi	Admin	FMCG	50 - 249 (Usaha Menengah)	Di atas 10
13	11/25/2025 18:59:37	Saya setuju untuk berpartisipasi	Admin	Jual beli online	Kurang dari 10 (Usaha Mikro)	3 - 10 Tah
14	11/25/2025 20:04:48	Saya setuju untuk berpartisipasi	Admin	Manufaktur	10 - 49 (Usaha Kecil)	< 3 tahun
15	9/29/2025 9:31:46	Saya setuju untuk berpartisipasi	admin	Ritel	10 - 49 (Usaha Kecil)	< 3 tahun

Gambar 2. Dataset Adopsi Kecerdasan Buatan Pada UMKM di Indonesia Sumber (<https://doi.org/10.5281/zenodo.18425107>)

### Pre-processing

Tahap *pre-processing* data ini menggunakan kuesioner adopsi kecerdasan buatan pada UMKM di Indonesia dilakukan pembersihan dan *encoding* data agar tidak terjadi kesalahan dan untuk memaksimalkan hasil dari implementasi algoritma *random forest*. Pada tahap *pre-processing* data terdapat dua tahapan yang dilakukan diantaranya yaitu:

a. Pembersihan data, langkah ini dilakukan pembersihan untuk data yang memiliki *missing value* seperti kolom yang kosong. Jika terdapat kolom yang kosong pada suatu baris akan dilakukan penghapusan. Berdasarkan pada *dataset* yang digunakan terdapat beberapa baris yang memiliki *missing value* sehingga dilakukan penghapusan. Baris yang memiliki *missing*

value berjumlah 9 buah sehingga *dataset* yang digunakan sekarang berjumlah 305. Serta pada langkah ini dilakukan penghapusan pada kolom yang tidak berkaitan langsung dengan hasil klasifikasi dan penerapan algoritma *random forest*. Beberapa kolom yang dihapus adalah kolom *time stamp*, dan kolom persetujuan mengisi kuesioner.

*Encoding* data, setelah melakukan pembersihan data, langkah selanjutnya adalah *encoding* data. pada tahap ini data yang bersifat kategorikal diubah menjadi numerik. Seperti pada kolom adopsi kecerdasan buatan untuk labelnya berubah menjadi 0 = tidak dan 1 = ya. Berikut hasil *dataset* yang sudah dilakukan *encoding* dapat dilihat pada gambar 3 berikut ini.

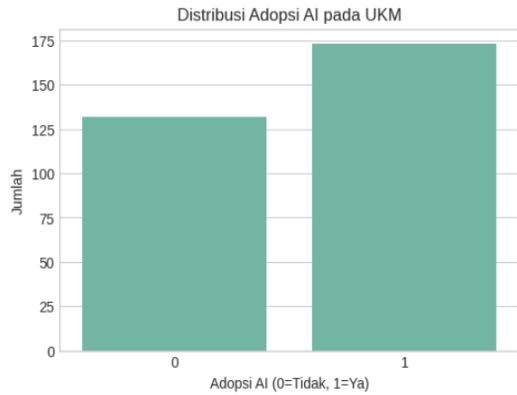
1	A	B	C	D	E
1	1. Apa jabatan Anda di perusahaan?	2. Sektor Industri Mana perusahaan anda bergerak ?	3. Berapa Jumlah Karyawan Teta di Perusahaan Anda ?	4. Sudah Berapa Lama perusahaan anda beroperasi ?	5. Apaka
2	Account executive		1	2	2
3	Accounting		1	1	1
4	Admin		0	1	0
5	Admin		0	2	0
6	admin		0	2	0
7	admin		0	2	1
8	admin		0	1	2
9	Admin		0	2	2
10	Admin		1	0	1
11	admin		1	1	0
12	Admin		2	0	1
13	admin		2	2	1
14	admin		2	1	2
15	Admin		2	2	2

Gambar 3. Hasil *Pre-processing* Data

### Implementasi Algoritma *Random Forest*

Implementasi algoritma *random forest* untuk mengklasifikasikan adopsi teknologi kecerdasan buatan pada UMKM di Indonesia menggunakan

*dataset* yang sudah di *preprocessing*. Distribusi kelas adopsi kecerdasan buatan pada UMKM dapat dilihat pada gambar 4 berikut ini.



**Gambar 4.** Distribusi Kelas Adopsi Kecerdasan Buatan pada UMKM

Berdasarkan hasil pengolahan data, distribusi target penelitian menunjukkan bahwa dari *dataset* yang digunakan, sebanyak 173 UMKM (56,7%) tergolong dalam kelas Adopsi Kecerdasan Buatan, sementara 132 UMKM (43,3%) lainnya tergolong dalam kelas Tidak Adopsi Kecerdasan Buatan. Data kemudian terbagi menjadi 244 data latih dan 61 data uji (rasio 80:20), dengan rasio ketidakseimbangan kelas 0,768 yang menunjukkan data cukup seimbang sehingga tidak perlu dilakukan SMOTE. Kode pembagian data latih dan data uji dapat dilihat pada gambar 5 berikut ini.

```
RANDOM_STATE = 42
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=RANDOM_STATE, stratify=y)
print(f" Data Train: {X_train.shape} ({len(X_train)/len(X)*100:.1f}%)")
print(f" Data Test: {X_test.shape} ({len(X_test)/len(X)*100:.1f}%)")
```

**Gambar 5.** Kode Pembagian Data Latih dan Data Uji

Selanjutnya, pada *hyperparameter tuning* dengan menggunakan *grid search* dengan *3-fold cross-validation* pada 108 kombinasi parameter menghasilkan nilai parameter optimal: *n\_estimators* = 300, *max\_depth* = 10, *min\_samples\_split* = 2, dan *min\_samples\_leaf* = 1, dengan skor *F1-Macro* 0,7824. Kode *hyperparameter tuning* dengan *grid search* dapat dilihat pada gambar 6 berikut ini.

```
param_grid = [
    {'n_estimators': [100, 200, 300],
     'max_depth': [10, 15, 20, None],
     'min_samples_split': [2, 5, 10],
     'min_samples_leaf': [1, 2, 4]}]
grid_search = GridSearchCV(
    RandomForestClassifier(random_state=RANDOM_STATE, class_weight='balanced'),
    param_grid,
    cv=StratifiedKFold(n_splits=3, shuffle=True, random_state=RANDOM_STATE),
    scoring='f1_macro',
    n_jobs=-1,
    verbose=1)
grid_search.fit(X_train_resampled, y_train_resampled)
print(f" Parameter Terbaik: {grid_search.best_params_}")
print(f" Skor Terbaik (F1-Macro): {grid_search.best_score_:.4f}")
```

**Gambar 6.** Kode *Hyperparameter Tuning* dengan *Grid Search*

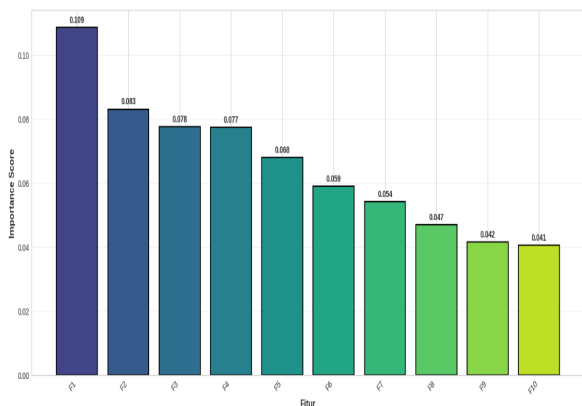
Setelah mendapatkan parameter optimal, dilakukan *feature selection* menggunakan metode *threshold importance* untuk memilih fitur-fitur yang memiliki pengaruh terbesar terhadap model. Eksperimen dilakukan dengan lima *threshold* berbeda (0,01 hingga 0,05), dan hasil terbaik diperoleh pada *threshold* 0,02 yang menghasilkan 21 fitur dengan akurasi 80,3% dari total 22 fitur. Kode *feature selection* dengan berbagai *threshold* dapat dilihat pada gambar 7 berikut ini.

```
thresholds = [0.01, 0.02, 0.03, 0.04, 0.05]
selection_results = []
for th in thresholds:
    selected = feature_importance[feature_importance['Importance'] > th]['Fitur'].tolist()
    if len(selected) == 0:
        continue
    X_selected = X[selected]
    X_train_sel, X_test_sel, y_train_sel, y_test_sel = train_test_split(
        X_selected, y, test_size=0.2, random_state=RANDOM_STATE, stratify=y)
    if ratio < 0.7:
        smote = SMOTE(random_state=RANDOM_STATE)
        X_train_sel_resampled, y_train_sel_resampled = smote.fit_resample(X_train_sel, y_train_sel)
    else:
        X_train_sel_resampled, y_train_sel_resampled = X_train_sel, y_train_sel
    rf_sel = RandomForestClassifier(**grid_search.best_params_)
    rf_sel.fit(X_train_sel_resampled, y_train_sel_resampled)
    y_pred_sel = rf_sel.predict(X_test_sel)
    acc_sel = accuracy_score(y_test_sel, y_pred_sel)
    selection_results.append({
        'threshold': th,
        'n_features': len(selected),
        'accuracy': acc_sel,
        'features': selected
    })
print(f" Threshold {th:.2f}: {len(selected):2d} fitur -> Akurasi: {acc_sel:.1%}")
```

**Gambar 7.** Kode *Feature Selection* dengan Berbagai *Threshold*

Setelah memperoleh 21 fitur terpilih melalui mekanisme *feature selection* dengan *threshold* 0,02, dilakukan *analisis feature importance* untuk mengukur kontribusi setiap fitur terhadap kinerja model *random forest* dalam mengklasifikasikan adopsi kecerdasan buatan pada UMKM di Indonesia.

Hasil analisis *feature importance* menunjukkan terdapat 10 fitur yang terpenting. Tiga faktor teratas diantaranya Strategic Decision Making, Digital Leadership, dan jabatan respon. Faktor lainnya yang turut berkontribusi adalah persepsi terhadap adopsi kecerdasan buatan (A. Mean), integrasi kecerdasan buatan dalam operasional (A12), serta penggunaan awal teknologi kecerdasan buatan (A11). Selain itu, keputusan berbasis data (SD1), sektor industri, keputusan strategis jangka panjang (SD3), dan persepsi efisiensi biaya melalui kecerdasan buatan (A13). Berikut hasil fitur yang mempengaruhi adopsi kecerdasan buatan pada UMKM di Indonesia, dapat dilihat pada gambar 8 berikut ini.



Gambar 8. Sepuluh Feature Importance Tertinggi

Berdasarkan pada gambar 8 fitur yang paling berpengaruh adalah Strategic Decision Making (10,9%), diikuti Digital Leadership (8,3%) dan jabatan responden (7,8%). Faktor lain yang berkontribusi meliputi persepsi adopsi AI, integrasi AI dalam operasional, penggunaan awal AI, keputusan berbasis data, sektor industri, keputusan strategis jangka panjang, dan persepsi efisiensi biaya.

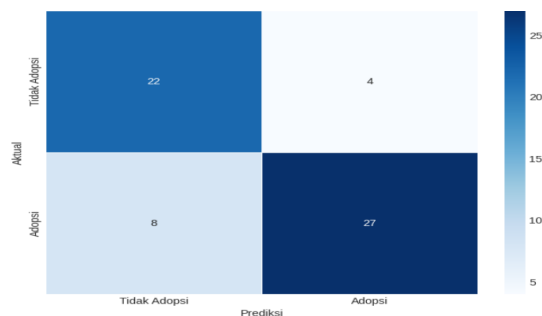
Kapabilitas manajerial UMKM dalam pengambilan keputusan strategis berbasis data menjadi faktor pembeda utama dalam adopsi inovasi. UMKM di Indonesia banyak memiliki keterbatasan modal dan sumber daya SDM dalam implementasi teknologi berbasis kecerdasan buatan. Sehingga hal ini menunjukkan bahwa adopsi kecerdasan buatan pada UMKM bergantung pada kesiapan *governance*. Selain itu, kepemimpinan digital yang visioner juga diperlukan untuk mendorong perubahan budaya organisasi di lingkungan UMKM. Keterbatasan infrastruktur dan biaya implementasi dapat diatasi jika ada kepemimpinan yang kuat mendorong transformasi. Hal ini mengindikasikan bahwa hambatan adopsi kecerdasan buatan pada UMKM Indonesia lebih bersifat human dan managerial daripada teknis. Dengan demikian, program intervensi pemerintah perlu bergeser dari pendekatan yang berfokus pada penyediaan infrastruktur fisik, menuju penguatan kapasitas kepemimpinan digital dan peningkatan literasi data bagi para pemilik serta manajer UMKM.

Setelah model berhasil diimplementasikan

dengan parameter optimal, tahap selanjutnya adalah mengevaluasi kinerja model dalam mengklasifikasikan adopsi kecerdasan buatan pada UMKM. Evaluasi dilakukan menggunakan *Confusion matrix classification report*, kurva ROC, dan *cross-validation* untuk memastikan akurasi, stabilitas, serta mencegah *overfitting*. Pengujian model menggunakan 20% dari total data.

### Evaluasi Model

Evaluasi pertama dilakukan melalui *confusion matrix* yang menjadi dasar perhitungan berbagai metric evaluasi. Hasil *confusion matrix* dari pengujian model *random forest* dapat dilihat pada gambar 9 berikut ini.



Gambar 9. Confusion Matrix Model Random Forest

Pada hasil *Confusion matrix* dari 20% data testing yang digunakan, model berhasil memprediksi dengan benar sebanyak 49 UMKM, terdiri dari 22 UMKM yang benar-benar tidak mengadopsi kecerdasan buatan (*True Negative*) dan 27 UMKM yang benar-benar mengadopsi kecerdasan buatan (*True Positive*). Model mengalami kesalahan prediksi pada 12 UMKM, yang terdiri dari *false positive* (FP) sebanyak 4 UMKM tidak mengadopsi kecerdasan buatan namun diprediksi sebagai adopsi, dan *False Negative* (FN) sebanyak 8 UMKM telah mengadopsi namun diprediksi tidak adopsi.

Setelah mendapatkan hasil *confusion matrix* Langkah selanjutnya dalam tahapan perhitungan metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*. Berikut hasil perhitungan metrix evaluasi berdasarkan pada nilai *confusion matrix* dapat dilihat pada tabel 1 berikut ini.

Tabel 1. Hasil Evaluasi Model Random Forest

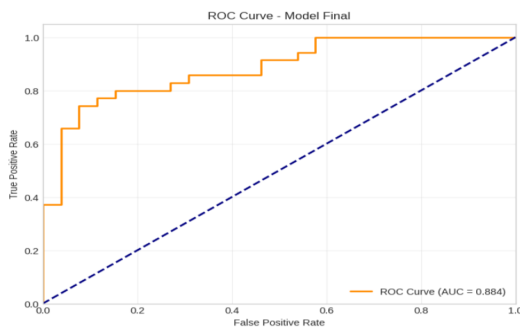
Kelas	Presisi	Recall	F1-Score	Support
Tidak Adopsi (0)	0,73	0,85	0,79	26
Adopsi Kecerdasan Buatan (1)	0,87	0,77	0,82	35
Macro Avg	0,80	0,81	0,80	61
Weighted Avg	0,81	0,80	0,80	61

Metrik	Nilai
Akurasi	80,3%
Presisi ( <i>Weighted</i> )	81,2%
Recall ( <i>Weighted</i> )	80,3%
<i>F1-score (Weighted)</i>	80,4%

Berdasarkan tabel diatas, hasil metrik evaluasi model *random forest* dalam memprediksi kelas Adopsi kecerdasan buatan dengan presisi 87% dan *F1-score* 82%, dibandingkan kelas Tidak Adopsi yang memiliki presisi 73% dan *F1-score* 79%. Hal ini mengindikasikan bahwa prediksi model terhadap UMKM yang mengadopsi kecerdasan buatan memiliki tingkat keakuratan yang tinggi.

Namun, *recall* kelas Adopsi (77%) lebih rendah dibanding kelas Tidak Adopsi (85%), yang berarti model masih cukup banyak melewatkan UMKM yang sebenarnya mengadopsi kecerdasan buatan (8 dari 35 UMKM tidak terdeteksi). Sementara itu, nilai *weighted average* yang seimbang (*precision* 81,2%, *recall* 80,3%, *F1-score* 80,4%) menunjukkan bahwa kinerja model tetap stabil antar kelas.

Hasil *5-fold cross-validation* menghasilkan akurasi rata-rata  $80,3\% \pm 6,6\%$ , mengonfirmasi bahwa model memiliki stabilitas yang baik. Tidak ditemukan perbedaan yang signifikan pada akurasi data uji (80,3%) dan hasil *cross-validation* ( $80,3\% \pm 6,6\%$ ), sehingga model tidak menunjukkan indikasi *overfitting*. Kurva ROC model *random forest* dapat dilihat pada Gambar 10 berikut ini.



Gambar 10. Kurva ROC Model *Random Forest*

Gambar kurva ROC menampilkan nilai AUC sebesar 0,884. Nilai AUC sebesar 0,884 mengindikasikan bahwa model memiliki kemampuan diskriminatif yang sangat baik dalam mengidentifikasi perbedaan kelas adopsi dan tidak adopsi kecerdasan buatan. Kurva ROC yang terletak jauh di atas garis diagonal (AUC 0,5) mengonfirmasi bahwa model yang dikembangkan jauh lebih baik dibandingkan dengan model acak.

Dari hasil penelitian, algoritma *random forest* mampu melakukan klasifikasi terhadap tingkat adopsi kecerdasan buatan pada UMKM dengan tingkat akurasi yang baik 80,3% dan nilai ROC-AUC 0,884 yang mengindikasikan kemampuan diskriminasi yang sangat baik berdasarkan nilai AUC. Evaluasi model menunjukkan bahwa model memberikan nilai *precision* (81,2%), *recall* (80,3%), dan *F1-score* (80,4%) yang memuaskan, menunjukkan efektivitasnya dalam mengklasifikasikan adopsi kecerdasan buatan pada UMKM. Analisis *feature importance* mengidentifikasi tiga faktor dominan yang mempengaruhi adopsi kecerdasan buatan, yaitu *Strategic Decision Making* (10,9%), *Digital Leadership* (8,3%), dan Jabatan Responden (7,8%). Temuan ini dapat menjadi masukan penting bagi program transformasi digital UMKM di Indonesia. Agar program tersebut berjalan efektif, prioritas utama perlu difokuskan pada peningkatan kemampuan pengambilan keputusan berbasis data dan penguatan kapasitas kepemimpinan digital bagi para pemilik dan manajer UMKM.

### KESIMPULAN

Berdasarkan analisis dari studi yang dijalankan, algoritma yang digunakan adalah *random forest* terbukti mampu mengklasifikasikan adopsi kecerdasan buatan pada UMKM di Indonesia dengan performa yang baik untuk membedakan UMKM yang telah mengadopsi dan belum mengadopsi kecerdasan buatan. Kinerja model yang stabil juga terkonfirmasi melalui hasil evaluasi *precision*, *recall*, dan *F1-score* yang relatif seimbang. Hasil penelitian ini menunjukkan bahwa keberhasilan transformasi *digital* UMKM tidak hanya dipengaruhi oleh kesiapan infrastruktur teknologi, melainkan lebih pada kualitas kepemimpinan dan kapasitas manajerial dalam pengambilan keputusan berbasis data. Penelitian ini berkontribusi dalam penerapan *data mining* untuk analisis adopsi teknologi dan diharapkan dapat menjadi acuan bagi pemerintah dalam merancang program percepatan transformasi digital UMKM yang lebih tepat sasaran.

## DAFTAR PUSTAKA

- Alvines, M., Wijayanto, M. R., Archi, M., Danendra, D., Herdiansyah, M. R., Lubis, M. K., Ichsan, M., Rachmad, F., Ditha Tania, K., & Meiriza, A. (2025). Komparasi *Ridge Regression*, *Random Forest*, Dan *Gradient Boosting* Untuk Prediksi Curah Hujan Harian Di Sumatera Selatan Berbasis *Time Series Cross-Validation*. *Jurnal Mahasiswa Teknik Informatika*, 9(4).
- Ananda, M. K., Assiroj, P., Hartati, B., & Samuel, F. (2024). Implementasi Data Mining Pada Instansi Pemerintahan (Systematic Literature Review). *Jurnal Mahasiswa Teknik Informatika*, 8(3).
- Aqil, Z. M., Naufaldihanif, R., Nuraini Kusuma, A., Hanggara, B., Clark Peter Wijaya, A., Ditha Tania, K., & Kurnia Sari, W. (2025). Penilaian Komparatif Metode Klasifikasi Neural Network Dan Random Forest Untuk Knowledge Discovery Pada Penyakit Diabetes. *Jurnal Mahasiswa Teknik Informatika*, 9(3).  
<https://www.kaggle.com/datasets/alextebo>
- Ardhani, D. A., & Ditha Tania, K. (2025). Knowledge Discovery on E-Commerce Customer Churn Using Interpretable Machine Learning: A Comparative Study of SHAP-Based Classifiers. *Journal of Applied Informatics and Computing (JAIC)*, 9(5).  
<http://jurnal.polibatam.ac.id/index.php/JAIC>
- Arroyabe, M. F., Arranz, C. F. A., Fernandez De Arroyabe, I., & Fernandez de Arroyabe, J. C. (2024). *Analyzing AI adoption in European SMEs: A study of digital capabilities, innovation, and external environment*. *Technology in Society*, 79.  
<https://doi.org/10.1016/j.techsoc.2024.102733>
- Cikhal, M., Andini, S. A., Adellia, E. A., & Yulaeli, T. (2024). Peran Transformasi Digital dan Tantangan Inovasi terhadap Keberlanjutan UKM di Indonesia (Studi *Literature Review*): II (Number 2).
- Dwi, A., Putra, V., Vindua, R., Raya Puspitak, J., Pamulang, K., & Tangerang Selatan, K. (2025). Implementasi *Data Mining* Dalam Optimasi Stok Bahan Baku Kue di Pt. Xyz Menggunakan Metode K Nearest Neighbor. *Jurnal Mahasiswa Teknik Informatika*, 9(3).
- Fahria, D. N., Tania, K. D., & Kurnia, R. D. (2026). Analisis Komparatif Algoritma *Random Forest*, *Xgboost*, Dan *Catboost* Untuk Klasifikasi Tingkat Stres Pengguna Media Sosial. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, 11(1), 1843–1853.  
<https://doi.org/10.36341/rabit.v11i1.7449>
- Fatihaturrahmah, A., & Tania, K. D. (2025). Review: A Hybrid Approach of Aspect-Based Sentiment Analysis and Knowledge Extraction for Evaluating Security Perceptions in Digital Payment Applications. *Scientific Journal of Informatics*, 12(4), 635–650.  
<https://doi.org/10.15294/sji.v12i4.31557>
- Hikmah, F., Suakanto, S., Kunci, K., Buatan, K., Mesin, P., Generatif, A. I., & Alami, B. (2026). Peran Artificial Intelligence Dalam Usaha Kecil Dan Menengah. *Systematic Literature Review (SLR)*. 14(1), 746–751.  
<https://doi.org/10.37081/ed.v14i1.7965>
- Irma Purnamasari, A., & Ali, I. (2024). Perbandingan Tingkat Akurasi Algoritma Decision Tree Dan Random Forest Dalam Mengklasifikasi Penerima Bantuan Sosial Bpnt di Desa Slangit. *Jurnal Mahasiswa Teknik Informatika*, 8(1).
- Karyadiputra, E., Teknologi Informasi, F., & Islam Kalimantan Muhammad Arsyad Al Banjari Banjarmasin, U. (2022). Penerapan Data Mining Untuk Klasifikasi Spesies Ikan di Lingkungan Akuatik Air Tawar. *Technologia*, 13(3).
- Kementerian Keuangan Republik Indonesia. (2023, August 5). *Peran UMKM dalam Perekonomian Indonesia*.  
<https://djpb.kemenkeu.go.id/kppn/purwakarta/id/data-publikasi/berita-terbaru/2891-peran-umkm-dalam-perekonomian-indonesia.html>
- Kuswanto, J., Lukmanul, H., Info, A., & Kunci, K. (2025). Penerapan Algoritma *Random Forest* untuk memprediksi Performa Akademik Mahasiswa. 5(1), 262–270.  
<https://doi.org/10.51454/del>
- Larasati, F. A., Ratnawati, D. E., & Hanggara, B. T. (2022). Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest (Vol. 6, Number 9). <http://j-ptiik.ub.ac.id>
- Priatmojo, H., Saputra, F., Prasetyo, M. H., Puspitasari, D., & Nurlaela, D. (2023). Perbandingan Klasifikasi Tingkat Penjualan Buah di Supermarket dengan Pendekatan Algoritma Decision Tree, Naive Bayes dan K-Nearest Neighbor. *Jurnal INSAN (Journal of Information Systems Management Innovation)*, 3(1). <http://jurnal.bsi.ac.id/index.php/jinsan>
- Putri, L. A. (2025, July). *Apa Saja Tantangan Adopsi AI bagi Pelaku Usaha di Indonesia?*  
<https://www.tempo.co/ekonomi/apa-saja-tantangan-adopsi-ai-bagi-pelaku-usaha-di-indonesia--1663161>
- Qotrunnada, F. M., & Utomo, P. H. (2022). Metode Convolutional Neural Network untuk Klasifikasi Wajah Bermasker. In *Prosiding Seminar Nasional Matematika*, 5, 799–807.  
<https://journal.unnes.ac.id/sju/index.php/prisma>
- Rahayu, P. W., Sudipa, I. G. I., Suryani, Ridwan, A., Darmawiguna, I. G. M., Sutoyo, Muh. N., Slamet, I., Harlina, S., & Maysanjaya, I. M. D. (2024). *Buku Ajar Data Mining* (1st ed.). PT. Sonpedia Publishing Indonesia.  
[www.buku.sonpedia.com](http://www.buku.sonpedia.com)
- Ramadhan, F., Herlambang, D., & Dipta, A. P. (2026). Prediksi Status Kesehatan Berdasarkan Gaya Hidup Menggunakan Metode Decision Tree dan Feature Importance. *RIGGS: Journal of Artificial Intelligence and Digital Business*,

- 4(4), 9616–9623.  
<https://doi.org/10.31004/riggs.v4i4.5246>
- Srirahayu, A., & Setya Pribadie, L. (2023). *Review Paper Data Mining Klasifikasi Data Mining. Jurnal Ilmiah Informatika Global, 14(1)*, 7-12.  
<https://doi.org/10.36982/jiig.v13i2.2307>
- Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. *JBASE - Journal of Business and Audit Information Systems, 4(1)*.  
<https://doi.org/10.30813/jbase.v4i1.2729>
- Yaman, N. I., Juwita, A. R., Lestari, S. A. P., & Faisal, S. (2024). Perbandingan Kinerja Algoritma Decision Tree dan Random Forest untuk Klasifikasi Nutrisi pada Makanan Cepat Saji. *Jurnal Algoritma, 21(2)*, 184–196.  
<https://doi.org/10.33364/algoritma/v.21-2.164>