JURNAL TIMES
TECHNOLOGY INFORMATICS & COMPUTER SYSTEM

Volume XIII No 2, Desember 2024
*p*ISSN : 2337 – 3601
*e*ISSN : 2549 – 015X

Tersedia *online* di http://ejournal.stmik-time.ac.id

KOMPARASI KINERJA ALGORITMA *RANDOM FOREST* DAN *SUPPORT VECTOR MACHINE* BERBASIS *ADAPTIVE BOOSTING* UNTUK ANALISIS KELAYAKAN KREDIT NASABAH

Andi[1)], Thamrin[2)]
[1]Program Studi Sistem Informasi
[2]Program Studi Magister Manajemen
[1,2]Institut Informasi Teknologi dan Bisnis
Jl. Mahoni No.16, Gaharu, Kec. Medan Tim., Kota Medan, Sumatera Utara 20235, Telp: 061-4530505
email: andi@itnb.ac.id[1)], thamrin@itnb.ac.id[2)]

Abstrak
Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Random Forest* dan *Support Vector Machine* (SVM) yang ditingkatkan dengan *Adaptive Boosting* dalam penentuan kelayakan pinjaman dana. Metode ini dilakukan dengan menggunakan dataset dari Kaggle yang terdiri dari 614 data calon peminjam, dengan atribut-atribut seperti jenis kelamin, status pernikahan, jumlah tanggungan, pendidikan, status usaha, pendapatan, jumlah pinjaman, jangka waktu pinjaman, riwayat kredit, dan lokasi rumah. Hasil penelitian menunjukkan bahwa *Random Forest* berbasis *Adaptive Boosting* mencapai akurasi 92,35%, presisi 90,12%, recall 89,87%, dan *missclassification error* 7,65%, sedangkan SVM berbasis *Adaptive Boosting* mencapai akurasi 85,76%, presisi 82,45%, *recall* 81,93%, dan *missclassification error* 14,24%. Temuan ini mengindikasikan bahwa *Random Forest lebih* unggul dibandingkan dengan SVM dalam memprediksi kelayakan pinjaman dana ketika ditingkatkan dengan *Adaptive Boosting.* Penelitian ini memberikan kontribusi dalam pengembangan teknik penilaian risiko kredit menggunakan pendekatan data *mining* yang dapat diterapkan dalam industri keuangan untuk meningkatkan ketepatan dalam pengambilan keputusan kredit.
Kata Kunci: Komparasi Algoritma, *Data Mining*, *Random Forest*, *Support Vector Machine*, *Adaptive Boosting*

1. Pendahuluan

Pembiayaan kredit atau pemberian pinjaman dana merupakan kegiatan penyaluran dana kepada individu atau bisnis berdasarkan perjanjian antara peminjam dan lembaga keuangan, dengan ketentuan bahwa peminjam harus melunasi pinjamannya dalam jangka waktu yang telah ditentukan. Dalam menentukan kelayakan kredit, lembaga keuangan melakukan berbagai analisis terhadap calon peminjam, termasuk menilai tingkat risiko keterlambatan pembayaran. Proses ini bertujuan untuk meminimalkan risiko gagal bayar serta meningkatkan efisiensi dalam pengelolaan pinjaman [1].

Seiring dengan meningkatnya jumlah peminjam, terutama dari sektor usaha yang membutuhkan modal tambahan, penentuan kelayakan kredit menjadi semakin kompleks. Kesalahan dalam analisis kelayakan peminjam dapat menyebabkan peningkatan risiko kredit macet bagi lembaga keuangan. Oleh karena itu, diperlukan metode berbasis teknologi informasi yang mampu mengotomatisasi dan meningkatkan akurasi dalam proses evaluasi kelayakan kredit. Salah satu pendekatan yang dapat digunakan adalah penerapan algoritma pembelajaran mesin (machine learning) untuk mengolah data nasabah secara lebih akurat dan efisien [2].

Pemanfaatan teknologi informasi dalam penentuan kelayakan pinjaman dana dapat dilakukan melalui teknik data mining. Data Mining adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari data tersebut [3]. Data mining merupakan suatu rangkaian atau tindakan untuk menemukan hubungan yang memiliki arti melalui pola dan kecenderungan dalam himpunan besar data yang tersimpan dengan menggunakan metode ataupun algoritma [4]. Data Mining digunakan untuk menemukan pola atau penyimpangan dalam sejarah data kredit untuk membuat model pengklasifikasian dan mengenali masalah baru [5]. Kelebihan dari pemanfaatan sistem informasi dan teknik data mining adalah dapat melakukan penentuan kelayakan kredit nasabah secara cepat dan akurat. Selain itu, dengan data mining, maka dapat dilakukan prediksi dengan melibatkan parameter-parameter yang lebih banyak tidak hanya berdasarkan satu parameter saja namun juga melibatkan parameter-parameter lain [6].

Penelitian yang membahas mengenai implementasi data mining dalam penentuan kelayakan kredit nasabah sudah pernah dilakukan pada tahun 2023 dengan mengimplementasikan algoritma Support Vector Machine, Artificial Neural Network, dan Naïve Bayes. Hasil penelitian yang telah menguji ketiga metode dengan menggunakan Cross Validation, Confusion Matrix, dan kurva ROC, metode Support Vector Machine (SVM) yang merupakan metode dengan hasil terbaik akurasi 92,0%, kemudian metode kedua adalah Artificial Neural Network

(ANN) sebesar 91,2% dan akurasi terendah adalah metode Naïve Bayes dengan akurasi 81,2% [7]. Selanjutnya, penelitian terbaru di tahun 2024 yang melakukan analisis kredit nasabah dengan algoritma Random Forest. Hasil pengujian algoritma Random Forest dalam analisis kelayakan kredit nasabah menggunakan *Confusion Matrix* pada penelitian ini memperoleh nilai akurasi, sensitifitas, spesitifitas dan AUC tertinggi yaitu sebesar 84,75%, 87,88%, 80,77% dan 84,32% [8].

Berdasarkan uraian penelitian terdahulu, maka pada penelitian ini dilakukan studi komparasi terhadap algoritma data *mining* yang diimplementasikan sebelumnya yaitu algoritma *Random Forest* dan *Support Vector Machine* dalam melakukan penentuan kelayakan kredit nasabah. Penelitian ini memilih kedua algoritma tersebut dikarenakan berdasarkan studi komparasi pada penelitian sebelumnya didapatkan kesimpulan bahwa kedua algoritma data *mining* tersebut memiliki akurasi yang lebih unggul dibandingkan algoritma data *mining* lainnya. Selain itu, untuk memberikan kontribusi pada penelitian ini, algoritma *Random Forest* dan *Support Vector* juga dikombinasikan dengan *Adaptive Boosting* agar dapat meningkatkan kinerja akurasi dari kedua algoritma tersebut.

2. Landasan Teori

Data *Mining*

Data *mining* adalah proses penggalian informasi dan pola yang bermanfaat dari suatu data yang sangat besar. Proses data *mining* terdiri dari pengumpulan data, ekstraksi data, analisa data, dan statistik data [9]. Empat proses dalam data *mining* ini akan menghasilkan model/pengetahuan yang sangat berguna. Data *mining* termasuk ke dalam *Knowledge Discovery* di dalam *database* (KDD) [10].

Knowledge Discovery

Knowledge discovery atau *knowledge discovery in database* (KDD) adalah salah satu metode untuk memperoleh pengetahuan dari basis data yang memiliki tabel-tabel yang saling berhubungan atau berelasi. Hasil dari pengetahuan yang didapatkan dari proses tersebut menjadi dasar sebagai basis pengetahuan *(knowledge base)* dalam pengambilan keputusan [10].

Pohon Keputusan

Pohon keputusan atau sering dikenal *decision tree* adalah salah satu teknik data mining yang terkenal dan salah satu metode paling populer untuk menentukan keputusan suatu kasus. Metode ini adalah cara pengolahan data untuk memprediksi masa mendatang dengan membuat model regresi atau klasifikasi menggunakan bentuk struktur pohon. Model *decision tree* yang menggunakan struktur hierarki atau struktur pohon konsepnya adalah dengan mengubah data dan dijadikan aturan keputusan serta pohon keputusan. Proses ini dilakukan dengan membaginya lebih lanjut menjadi himpunan bagian yang lebih kecil dan mengembangkan secara bertahap pohon keputusan. Pada tahapan tersebut hasil akhirnya yaitu pohon yang memiliki *node* keputusan dan *node* daun. Contoh dari note keputusan adalah cuaca dan memiliki cabang hujan, mendung dan cerah [10]. Terdapat beberapa algoritma yang sering digunakan untuk membangun *decision tree* yaitu [10]:

a. ID3 (*Iterative Dichotomiser* 3)
b. C4.5
c. C5.0
d. *Random Forest*
e. CHAID *(Chi-squared Automatic Interaction Detection)*
f. MARS *(Multivariate Adaptive Regression Splines)*

Algoritma *Random Forest*

Random Forest merupakan sebuah metode *ensemble* yang terdiri atas sekumpulan *decision tree* (pohon keputusan) yang digunakan untuk klasifikasi data ke suatu kelas. Langkah awal untuk menentukan pohon keputusan adalah dengan menghitung nilai *entropy* dan *information gain*. Terdapat dua buah formula dari algoritma *Random Forest* dalam membentuk pohon keputusan antara lain [11]:

a. Entropi.

Pendekatan entropi sebagai penentu ketidakmurnian atribut. Penghitungan nilai entropi dapat dilihat pada persamaan berikut:

$$Entropy(S) = \sum_{i=1}^{n} - p_i log_2\, p_i \quad (1)$$

Keterangan:
S : Himpunan dataset
n : Banyaknya jumlah kelas
p_i : Proporsi dari S_i terhadap S

b. *Information Gain*.

Information gain merupakan nilai perolehan informasi dalam memilah simpul. Perhitungan *gain* sangat berpengaruh terhadap setiap node teratas dan node pemisah. Perhitungan gini masih berlanjut ketika hasil akhir gini masih berupa angka dan berhenti ketika hasil akhir gini adalah nol.

$$Information\ Gain\ (A) = Entropy(S) = \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(Si) \quad (2)$$

Keterangan:

A : Atribut
S : Himpunan dataset
$|S_i|$: Jumlah sampel untuk nilai ke-i
$|S|$: Banyaknya jumlah data

Algoritma *Support Vector Machine*

Model algoritma SVM merupakan salah satu algoritma dari metode klasifikasi, yang bekerja dengan cara mencari suatu garis *(hyperplane)* untuk memisahkan dua kelompok data. *Hyperplane* adalah garis pemisah terbaik antara dua kelas. Untuk mencari hyperplane dapat dilakukan dengan mencari margin *hyperplane* dan mencari titik maksimum. Margin adalah jarak antara data terdekat di antara dua kelas yang berbeda, yang disebut dengan support vektor. Garis *solid* pada gambar 1-b menunjukkan *hyperplane* yang terbaik, karena terletak tepat diantara kedua *class*, sedangkan *support vector* dilambangkan dengan titik merah dan kuning yang berada di dalam lingkaran hitam [12].

Hyperplane klasifikasi linear SVM dinotasikan:

$$f(x) = w,x + b = 0 \quad (3)$$

Dari persamaan di atas di dapatkan pertidaksamaan kelas +1 (negatif)

$$w.x + b \leq +1 \quad (4)$$

Pertidaksamaan kelas -1:

$$w.x + b \geq -1 \quad (5)$$

w adalah bidang normal dan b adalah posisi bidang relatif terhadap koordinat pusat. Dengan mengoptimalkan nilai jarak antara hyperplane dan titik berikutnya, margin terbesar dapat ditemukan, yaitu 1 / ‖w‖. Ini dapat dirumuskan sebagai masalah pemrograman kuadratik (QP) di mana titik minimum persamaan 11. dengan mengingat kendala dari persamaan 12).

$$\mathbf{min}\ \tfrac{1}{2}||\,w\,||^2 = \mathbf{min}\ \ \tfrac{1}{2}(\,w1^2 + w2^2\,) \quad (6)$$

$$y_i\,(w_{xi}^2 + \ b)\ \geq 1,\ \ i\ =\ 1,2,3\ \ldots,\ N \quad (7)$$

Adaptive Boosting

Algoritma *Adaptive Boosting* merupakan algoritma yang digunakan untuk pengambilan keputusan *Adaboost* merupakan *ensemble learning* yang digunakan pada algoritma *boosting*. Algortima ini ditujukan untuk *supervised learning* yang memberikan nilai atribut pada dataset yang digambarkan oleh koleksi atribut dan termasuk salah satu dari serangkaian kelas yang saling berhubungan Langkah-langkah pada algoritma *Adaboost* adalah sebagai berikut [13]:

a. Inisial setiap bobot awal dengan rumus berikut untuk semua i.

$$W_i^1 = \frac{1}{n} \quad (8)$$

b. Untuk setiap m = 1, 2, 3, …, M. Lakukan:

c. Bangun model klasifikasi $G_m(X_i)$ menggunakan bobot $W_i^{(m)}$.

d. Hitung kesalahan klasifikasi dengan rumus:

$$\varepsilon_m = \frac{\sum_{i=1}^{n} w_i^{(m)}, G_m(X_i)\ \neq\ Y_i}{\sum_{i=1}^{n} w_i^{(m)}} \quad (9)$$

e. Menghitung nilai α menggunakan rumus:

$$\alpha_m = \ log\frac{1 - \ \varepsilon_m}{\varepsilon_m} \quad (10)$$

f. Perbaharui koefisien pembobotan kesalahan klasifikasi dengan rumus:

$$w_i^{m+1} = \ w_i^m \exp(-\alpha_m)\,, G_m(X_i) = \ Y_i \quad (11)$$

$$w_i^{m+1} = \ w_i^m \exp(\alpha_m)\,, G_m(X_i)\ \neq\ Y_i \quad (12)$$

g. Penentuan klasifikasi menggunakan rumus:

$$Y = Sign\left(\sum_{m=1}^{M} \alpha_m G_m(x)\right) \quad (13)$$

Confusion Matrix

Confusion Matrix adalah salah satu cara yang sering digunakan pada proses evaluasi model data mining klasifikasi dengan memprediksi kebenaran objek. *Confusion Matrix* digambarkan dengan tabel yang menyatakan

jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan. Berdasarkan tabel *Confusion Matrix* di atas [14][15]:

a. *True Positives* (TP) adalah jumlah *record* data positif yang diklasifikasikan sebagai nilai positif.

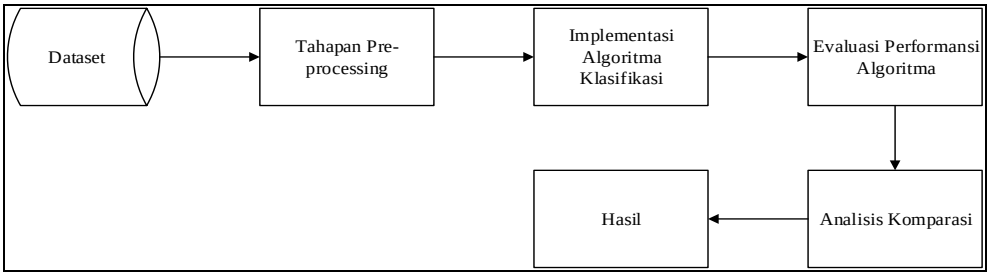
b. *False Positives* (FP) adalah jumlah *record* data negatif yang diklasifikasikan sebagai nilai positif.
c. *False Negatives* (FN) adalah jumlah *record* data positif yang diklasifikasikan sebagai nilai negatif.
d. *True Negatives* (TN) adalah jumlah *record* data negatif yang diklasifikasikan sebagai nilai *negative.*

Nilai yang dihasilkan melalui metode *Confusion Matrix* adalah berupa evaluasi sebagai berikut [14]:

a. *Accuracy*, persentase jumlah *record* data yang diklasifikasikan (prediksi) secara benar oleh algoritma.
Rumus: (TP + TN) / Total data = *Accuracy* (14)
b. *Precision*, persentase rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positf.
Rumus: (TP) / (TP+FP) = *Precission* (15)
c. *Recall*, persentase rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif.
Rumus: (TP) / (TP + FN) = *Recall* (16)
d. *Misclassification (Error) Rate*, persentase jumlah record data yang diklasifikasikan (prediksi secara salah oleh algoritma).
Rumus : (FP + FN) / Total data = *Misclassification Rate* (17)

3. Metode Penelitian

Dalam penelitian ini adapun tahapan metode penelitian yang akan dilakukan pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Metode Penelitian

Dataset

Dataset yang digunakan dalam penelitian ini diambil dari *website* Kaggle yaitu Loan Eligible Dataset dengan total data sebanyak 614 data. Tabel 1 menunjukkan dataset yang digunakan dalam penelitian ini.

Tabel 1. Dataset Penelitian

No.	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status
1	LP001002	Male	No	0	Graduate	No	5849	0		360	1	Urban	Y
2	LP001003	Male	Yes	1	Graduate	No	4583	1508	128	360	1	Rural	N
3	LP001005	Male	Yes	0	Graduate	Yes	3000	0	66	360	1	Urban	Y
4	LP001006	Male	Yes	0	Not Graduate	No	2583	2358	120	360	1	Urban	Y
5	LP001008	Male	No	0	Graduate	No	6000	0	141	360	1	Urban	Y
6	LP001011	Male	Yes	2	Graduate	Yes	5417	4196	267	360	1	Urban	Y
7	LP001013	Male	Yes	0	Not Graduate	No	2333	1516	95	360	1	Urban	Y
8	LP001014	Male	Yes	3+	Graduate	No	3036	2504	158	360	0	Semiurban	N
9	LP001018	Male	Yes	2	Graduate	No	4006	1526	168	360	1	Urban	Y
…	…	…	…	…	…	…	…	…	…	…	…	…	…
614	LP002990	Female	No	0	Graduate	Yes	4583	0	133	360	0	Semiurban	N

Tahapan *Pre-processing*

Tahapan *pre-processing* data dalam penelitian ini terbagi menjadi beberapa cara yaitu:

a. Metode imputasi.
Metode imputasi rata-rata merupakan metode untuk menangani data yang hilang pada dataset [16]. Dalam proses ini, nilai yang hilang pada atribut pada suatu fitur (kolom) diganti dengan nilai *mean, median, modus*, dan imputasi dengan kategori baru. Metode ini membantu menjaga konsistensi dan integritas data pada dataset [17]. Pada penelitian ini, metode imputasi khususnya atribut kategorikal diterapkan dengan imputasi melalui kategori baru, sedangkan untuk kolom dengan data numerik menggunakan *mean, median*, atau *modus.*

b. Metode normalisasi.
Metode normalisasi mengubah bentuk data menjadi dataset yang mudah dipahami dalam bentuk numerikal agar memudahkan proses pengolahan data ke algoritma. Metode normalisasi pada data numerikal yang diterapkan pada penelitian ini adalah *Z-score Normalization*. Sedangkan, metode *encoding* untuk data kategorikal adalah proses mengubah variabel kategori menjadi bentuk yang dapat dipahami oleh algoritma *machine learning.* Salah satu metode yang umum digunakan adalah *One-Hot Encoding* [18]
Dengan dibuatnya 3 buah kunci, maka kelemahan penggunaan kunci pada algoritma Vigenere Cipher dapat diatasi.

Implementasi Algoritma Klasifikasi

Pada tahap ini diimplementasikan kedua algoritma klasifikasi yang akan di analisis pada penelitian ini yaitu algoritma *Random Forest* dan *Support Vector Machine*, berbasis *AdaBoost* dalam dalam penentuan kelayakan kredit nasabah. Implementasi algoritma ditulis dengan menggunakan bahasa pemrograman Python dan *library Flask.*

Evaluasi Performansi Algoritma

Evaluasi performansi algoritma dalam penelitian ini dilakukan dengan menggunakan *Confusion Matrix,* yaitu tabulasi silang data kelas positif dan negatif yang dikelompokkan ke dalam kelas prediksi dan kelas aktual

Analisis Komparasi

Pada tahap ini dilakukan analisis perbandingan dari kedua algoritma klasifikasi yang di implementasikan pada penelitian ini sesudah di implementasikan *AdaBoost*. Hasil analisis komparatif menunjukkan tingkat akurasi, *recall*, presisi, dan kesalahan kesalahan klasifikasi pada masing-masing algoritma sehingga dapat diketahui dari kedua algoritma tersebuh manakah yang lebih unggul.

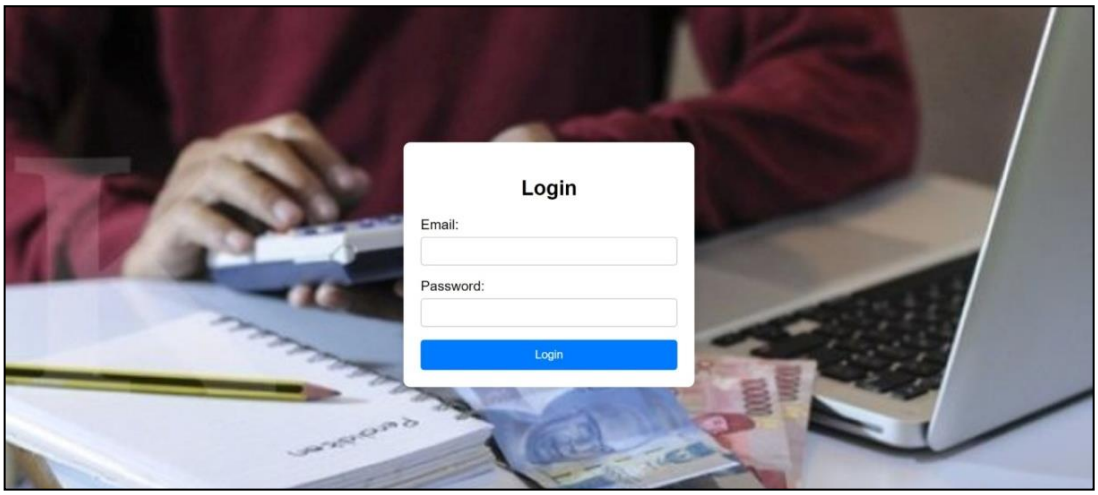
Hasil

Hasil penelitian berupa pembahasan komprehensif atas analisis yang telah dilakukan dan diuraikan secara rinci.

4. Hasil Penelitian

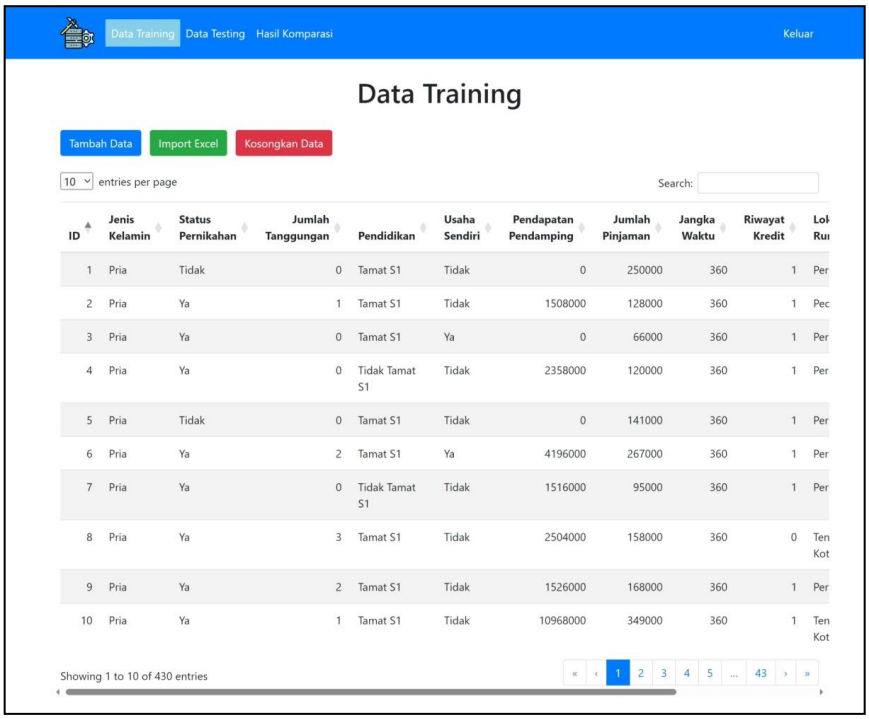
Hasil penelitian berupa dibangunnya sebuah *website* untuk menganalisis komparasi algoritma *Random Forest* dan *Support Vector Machine* berbasis *Adaptive Boosting* untuk penentuan kelayakan kredit nasabah. Berikut ini keseluruhan rancangan tampilan dari sistem penentuan kelayakan kredit nasabah yang digunakan sebagai alat untuk melakukan komparasi algoritma *Random Forest* dan *Support Vector Machine* berbasis *Adaptive Boosting* antara lain:

a. Halaman Awal/*Login*
Tampilan halaman awal/*login* adalah antarmuka pertama yang dilihat pengguna saat mereka mengakses sistem. Rancangan tampilan ini mencakup elemen-elemen seperti *formulir* login dan *tombol* login yang harus dilewati untuk menuju ke *dashboard* sistem.



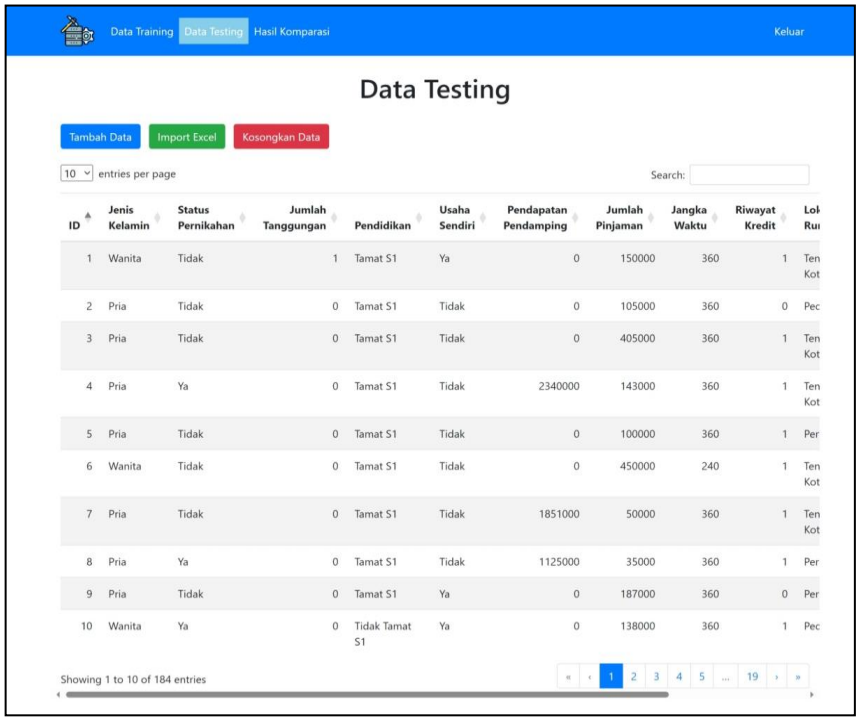
Gambar 2. Tampilan Awal/*Login*

b. Tampilan Halaman Kelola Data *Training.*
Tampilan halaman kelola data *training* berisikan informasi *list* data-data *training* yang tersedia. Selain itu, administrator juga dapat memasukkan, mengubah, dan menghapus data *training* dengan tombol-tombol yang tersedia.



Gambar 2. Tampilan Halaman Kelola Data *Training*

c. Tampilan Halaman Kelola Data *Testing*.
Tampilan halaman kelola data *testing* berisikan informasi *list* data-data *testing* yang tersedia. Selain itu, administrator juga dapat memasukkan, mengubah, dan menghapus data *testing* dengan tombol-tombol yang tersedia.



Gambar 3. Tampilan Halaman Kelola Data *Testing*

Berikutnya dilakukan pembahasan mengenai hasil penelitian yang didapatkan. Berdasarkan hasil analisis komparasi algoritma *Random Forest* dan *Support Vector Machine* berbasis *Adaptive Boosting* dengan total data sebanyak 614 data calon peminjam dengan pembagian data *training* sebanyak 80% dan data *testing* sebanyak 20% didapatkan perbandingan kinerja dari kedua algoritma tersebut yang ditunjukkan pada Tabel 2.

Tabel 2. Hasil Komparasi Algoritma

Algoritma	Akurasi (%)	Presisi (%)	*Recall* (%)	*Missclassification Error* (%)
Random Forest + Adaptive Boosting	92,35	90,12	89,87	7,65
Support Vector Machine + Adaptive Boosting	85,76	82,45	81,93	14,24

Berdasarkan hasil yang tercantum pada Tabel 1, terlihat bahwa algoritma ***Random Forest + Adaptive Boosting*** menunjukkan performa yang lebih unggul dibandingkan dengan ***Support Vector Machine* (SVM) + *Adaptive Boosting***. Secara keseluruhan, *Random Forest* mencapai akurasi sebesar 92,35%, yang secara signifikan lebih tinggi dibandingkan dengan 85,76% yang dicapai oleh SVM. Hal ini mengindikasikan bahwa model *Random Forest* lebih tepat dalam mengklasifikasikan data kelayakan kredit nasabah, sehingga dapat memberikan kepercayaan yang lebih tinggi terhadap keputusan pemberian kredit.

Dalam hal metrik presisi, *Random Forest + Adaptive Boosting* memperoleh nilai 90,12% yang artinya proporsi prediksi positif yang benar sangat tinggi. Artinya, model ini lebih mampu mengidentifikasi calon peminjam yang layak dengan tingkat kesalahan positif *(false positive)* yang lebih rendah. Sebaliknya, SVM + *Adaptive Boosting* memperoleh presisi sebesar 82,45%, menandakan adanya kemungkinan kesalahan dalam mengklasifikasikan calon peminjam yang seharusnya tidak memenuhi syarat. Kinerja presisi yang tinggi sangat penting dalam konteks penilaian risiko kredit, karena kesalahan dalam mengklasifikasikan nasabah dapat berdampak pada keputusan pemberian kredit yang tidak tepat dan potensi kerugian finansial bagi lembaga pemberi pinjaman.

Nilai *recall* yang diperoleh oleh *Random Forest* adalah 89,87%, yang menunjukkan bahwa model ini efektif dalam menangkap sebagian besar data yang benar-benar positif (calon peminjam yang layak). Di sisi lain, SVM menunjukkan recall sebesar 81,93%, yang berarti terdapat lebih banyak kasus layak yang terlewatkan. Dalam konteks analisis kelayakan kredit, recall yang tinggi sangat penting karena mengurangi risiko kegagalan dalam mendeteksi calon peminjam yang seharusnya disetujui kreditnya. Dengan demikian, *Random Forest* tidak hanya lebih akurat secara keseluruhan, tetapi juga lebih andal dalam mengidentifikasi nasabah yang berpotensi memberikan kontribusi positif.

Selain itu, perbandingan nilai missclassification error semakin memperkuat keunggulan Random Forest. Dengan nilai *error* sebesar 7,65% dibandingkan dengan 14,24% pada SVM, dapat disimpulkan bahwa algoritma *Random Forest* dengan *Adaptive Boosting* lebih konsisten dalam memberikan prediksi yang benar dan mengurangi jumlah kesalahan klasifikasi. Tingkat kesalahan yang lebih rendah ini sangat penting dalam penerapan nyata, di mana setiap kesalahan klasifikasi dapat berdampak pada risiko finansial dan operasional yang signifikan.

Secara ilmiah, keunggulan performa *Random Forest + Adaptive Boosting* dapat dijelaskan melalui mekanisme *ensemble learning* yang diusungnya. Metode ini mengkombinasikan beberapa pohon keputusan untuk mengurangi varians dan menghindari overfitting, sementara *Adaptive Boosting* secara iteratif menyesuaikan bobot untuk data yang sulit diklasifikasikan. Sinergi kedua metode tersebut menghasilkan model yang lebih robust dan mampu menangani kompleksitas data yang tinggi, seperti atribut-atribut yang digunakan dalam analisis kelayakan kredit. Sedangkan, meskipun SVM merupakan *classifier* yang kuat terutama dalam menangani data dengan dimensi tinggi, performanya sangat bergantung pada pemilihan kernel dan parameter optimasinya. Kombinasi SVM dengan *Adaptive Boosting* telah memberikan peningkatan, namun belum mampu mengatasi tantangan heterogenitas data seefektif *Random Forest* dalam konteks ini [19].

Temuan penelitian ini memberikan kontribusi penting dalam pengembangan sistem penilaian risiko kredit berbasis teknologi informasi, yang mana pemilihan algoritma yang tepat sangat krusial dalam mendukung pengambilan keputusan yang cepat dan akurat di industri keuangan. Dengan mengimplementasikan model *Random Forest + Adaptive Boosting*, lembaga peminjam dapat meningkatkan efektivitas dalam mengidentifikasi kelayakan kredit nasabah, sehingga mengurangi risiko keterlambatan pembayaran dan potensi gagal bayar.

5. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma *Random Forest* yang diperkuat dengan *Adaptive Boosting* menunjukkan performa yang lebih unggul dibandingkan dengan *Support Vector Machine* yang juga ditingkatkan dengan *Adaptive Boosting* dalam menentukan kelayakan kredit nasabah. Model *Random Forest* mencapai akurasi yang lebih tinggi, yakni sebesar 92,35%, dengan nilai presisi dan *recall* yang masing-masing mencapai 90,12% dan 89,87%, serta tingkat *missclassification error* yang lebih rendah, yaitu hanya 7,65%. Hasil ini menandakan bahwa *Random Forest* lebih efektif dalam mengklasifikasikan calon peminjam yang layak serta mengurangi kesalahan prediksi, yang sangat krusial dalam konteks pengambilan keputusan kredit. Temuan ini menegaskan pentingnya penerapan metode *ensemble learning* untuk meningkatkan akurasi dan keandalan sistem penilaian risiko kredit, sehingga dapat mendukung lembaga keuangan dalam mengambil keputusan yang lebih tepat dan meminimalkan potensi risiko gagal bayar.

6. Daftar Pustaka

[1] Ardiyansyah, R. Sa'adah, Lisnawanty, and D. Purwaningtias, "Peningkatan Akurasi Metode C4 . 5 Untuk Memprediksi Kelayakan Kredit Berbasis Stratified Sampling Dan Optimize Selection," *Kumpul. J. Ilmu Komput.*, vol. 10, no. 2, pp. 239–249, 2023.

[2] D. N. Sholihaningtias, "Rekomendasi Kelayakan Penerima Kredit Menggunakan Metode TOPSIS dengan Pembobotan ROC," *J. SAINTEKOM*, vol. 13, no. 1, pp. 88–99, 2023, doi: 10.33020/saintekom.v13i1.376.

[3] A. Muzakir and R. A. Wulandari, "Model Data Mining Sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree," *Sci. J. Informatics*, vol. 3, no. 1, pp. 19–26, 2016.

[4] M. A. Mulyana, Y. Religia, and A. Suwarno, "Optimasi Berbasis Particle Swarm Optimization Pada Algoritma Naive Bayes Dalam Memprediksi Penyakit Stroke," *Pelita Teknol. J. Ilm. Inform. Arsit. dan Lingkung.*, vol. 14, no. 1, pp. 1–15, 2019.

[5] I. Nawangsih and P. Purnamasari, "Analisis Pola Pembelian Produk Kecantikan Menggunakan Algoritma Apriori," *J. Teknol. Inform. dan Komput.*, vol. 9, no. 1, pp. 537–546, 2023, doi: 10.37012/jtik.v9i1.1614.

[6] I. M. S. Ramayu, F. Susanto, and G. S. Mahendra, "Penerapan Data Mining Dengan Algoritma C4.5 Dalam Pemesanan Obat Guna Meningkatkan Keuntungan Apotek," in *Prosiding Seminar Nasional Manajemen, Desain & Aplikasi Bisnis Teknologi (SENADA)*, 2022, vol. 5, pp. 237–245, [Online]. Available: http://senada.idbbali.ac.id.

[7] B. Pernama and H. D. Purnomo, "Analisis Risiko Pinjaman dengan Metode Support Vector Machine, Artificial Neural Network dan Naïve Bayes," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 1, pp. 92–99, 2023, doi: 10.35870/jtik.v7i1.693.

[8] K. P. Ulandari, N. Chamidah, and A. Kurniawan, "Prediksi Risiko Gagal Bayar Kredit Kepemilikan Rumah Dengan Pendekatan Metode Random Forest," *SAINSMAT J. Ilm. Ilmu Pengetah. Alam*, vol. 13, no. 2, pp. 162–170, 2024.

[9] Robet, J. T. K. P. Angin, and O. Pribadi, "Implementation of Deep Learning Model for Classification of Household Trash Image," *Sink. J. dan Penelit. Tek. Inform. Vol. 8, Nomor 4, Oct. 2024 DOI https//doi.org/10.33395/sinkron.v8i4.14198 e-*, vol. 8, no. 4, pp. 2575–2583, 2024.

[10] Amna *et al.*, *Data Mining*. Padang: PT Global Eksekutif Teknologi, 2023.

[11] F. Diba, M. S. Lydia, and P. Sihombing, "Analisis Random Forest Menggunakan Principal Component Analysis Pada Data Berdimensi Tinggi," *Indones. J. Comput. Sci.*, vol. 12, no. 4, pp. 2152–2160, 2023, doi: 10.33022/ijcs.v12i4.3329.

[12] R. A. Rizal, I. S. Girsang, and S. A. Prasetiyo, "Klasifikasi Wajah Menggunakan Support Vector Machine (SVM)," *REMIK (Riset dan E-Jurnal Manaj. Inform. Komputer)*, vol. 3, no. 2, p. 1, 2019, doi: 10.33395/remik.v3i2.10080.

[13] M. R. Raihan, Y. H. Chrisnanto, and A. K. Ningsih, "KLASIFIKASI PENENTUAN KELAYAKAN PINJAMAN KOPERASI DENGAN ALGORITMA CART MENGGUNAKAN ALGORITMA ADABOOST," *INFOTECH J.*, vol. 8, no. 2, pp. 74–83, 2022.

[14] M. A. Muslim *et al.*, *Data Mining Algoritma C4.5*. 2019.

[15] A. Andi, C. Juliandy, and D. David, "Clustering Analysis of Tweets About COVID-19 Using the K-Means Algorithm," *Sinkron*, vol. 8, no. 1, pp. 543–533, 2023, doi: 10.33395/sinkron.v8i1.12145.

[16] A. E. Karrar, "The Effect of Using Data Pre-Processing by Imputations in Handling Missing Values," *Indones. J. Electr. Eng. Informatics*, vol. 10, no. 2, pp. 375–384, 2022, doi: 10.52549/ijeei.v10i2.3730.

[17] M. R. A. Prasetya and A. M. Priyatno, "Penanganan Imputasi Missing Values pada Data Time Series dengan Menggunakan Metode Data Mining," *J. Inf. dan Teknol.*, vol. 5, no. 2, pp. 56–62, 2023, doi: 10.37034/jidt.v5i1.324.

[18] M. Sholeh, D. Andayati, and R. Y. Rachmawati, "Data Mining Model Klasifikasi Menggunakan Algoritma K-Nearest Neighbor Dengan Normalisasi Untuk Prediksi Penyakit Diabetes," *TeIKa*, vol. 12, no. 02, pp. 77–87, 2022, doi: 10.36342/teika.v12i02.2911.

[19] Andi, Thamrin, A. Susanto, E. Wijaya, and D. Djohan, "Analysis of the random forest and grid search algorithms in early detection of diabetes mellitus disease," *J. Mantik*, vol. 7, no. 2, pp. 2685–4236, 2023, doi: 10.35335/mantik.v7i2.3981.