

Pengaruh Prediksi Kredit bermasalah Terhadap Keputusan Bank Dengan Komparasi Metode ANN, CNN, dan Random Forest

Gupita Nurmalita Sari^{*1}, Rujianto Eko Saputro², Giat Karyono³

^{1,2,3}Universitas Amikom Purwokerto, Indonesia

Email: ¹gupitasari96@gmail.com, ²rujianto@amikompurwokerto.ac.id,
³giatkaryono@amikompurwokerto.ac.id

Abstrak

Kredit bermasalah menjadi salah satu tantangan utama sektor perbankan karena dapat mengganggu stabilitas keuangan dan mempengaruhi keputusan strategis. Oleh karena itu, kemampuan memprediksi potensi kredit bermasalah secara akurat sangat penting dalam pengelolaan risiko. Penelitian ini membandingkan kinerja tiga algoritma kecerdasan buatan ANN (*Artificial Neural Network*) dengan arsitektur MLP (*Multilayer Perceptron*), CNN (*Convolutional Neural Network*), dan RF (*Random Forest*) untuk klasifikasi risiko kredit berdasarkan data nominatif kredit. Dataset mencakup fitur-fitur utama seperti plafon pinjaman, jangka waktu, penghasilan, baki debit, kolektibilitas dan total tunggakan. Data diklasifikasikan dalam lima kategori risiko berdasarkan status pembayaran antara lain L (Lancar), DP (Dalam Perhatian), KL (Kurang Lancar), D (Diragukan), dan M (Macet). Setiap model dievaluasi berdasarkan akurasi keseluruhan dan kemampuannya mengidentifikasi kelas minoritas dan mayoritas. Hasil menunjukkan bahwa *Random Forest* unggul dengan akurasi 91,42%, efektif dalam mengklasifikasikan kategori “Lancar” dan “Macet”. Model ANN MLP mencapai akurasi 89,90%, namun kurang optimal untuk kelas minoritas. Sementara CNN hanya mencatat akurasi 58,58% dan mengalami overfitting terhadap kelas mayoritas. Studi ini menyimpulkan bahwa *Random Forest* adalah metode paling tepat untuk data tabular yang seimbang, sementara CNN memiliki potensi dalam konteks data spasial atau sekuensial. Dan hasil yang akurat dari prediksi *Random forest* diharapkan dapat menjadi pertimbangan bagi lembaga keuangan dalam pengambilan keputusan terkait kebijakan di bank Pemerintah di Purbalingga.

Kata kunci: ANN, CNN, kebijakan, Kredit bermasalah, Random Forest.

The Impact of Non-Performing Loan Prediction on Bank Decisions Through a Comparison of ANN, CNN, and Random Forest Methods

Abstract

Non performing loans are one of the main challenges in the banking sector, as they can disrupt financial stability and affect strategic decision-making. Therefore, the ability to accurately predict the potential for non performing loans is crucial in risk management. This study compares the performance of three artificial intelligence algorithms Artificial Neural Network (ANN) with a Multilayer Perceptron (MLP) architecture, Convolutional Neural Network (CNN), and Random Forest (RF) for credit risk classification based on nominative credit data. The dataset includes key features such as loan amount, term, income, outstanding balance, collectability and total arrears. The data is classified into five risk categories based on payment status: Current (L), Special Mention (DP), Substandard (KL), Doubtful (D), and Loss (M). Each model is evaluated based on overall accuracy and its ability to identify both minority and majority classes. The results show that Random Forest outperforms the others with an accuracy of 91.42%, effectively classifying both "Current" and "Loss" categories. The ANN MLP model achieved an accuracy of 89.90%, but performed less effectively in handling minority classes. Meanwhile, CNN recorded only 58.58% accuracy and exhibited overfitting to the majority class. The study concludes that Random Forest is the most appropriate method for balanced tabular data, while CNN has potential for use with spatial or sequential data. Furthermore, the accurate predictions obtained from Random Forest are expected to provide valuable insights for financial institutions in making policy-related decisions at government banks in Purbalingga.

Keywords: ANN, CNN, policy, Non-performing Loans, Random Forest.

1. PENDAHULUAN

Dalam sektor perbankan, kredit bermasalah adalah faktor penting yang berdampak langsung dan mempengaruhi kesehatan keuangan dan kestabilan lembaga keuangan. Permasalahan kredit memperlihatkan ketidakmampuan bank dalam mengelola resiko kredit. Selain itu juga dapat berdampak buruk terhadap operasional bank, yang memicu kerugian finansial yang signifikan[1]. Oleh karena itu, memprediksi resiko kredit secara akurat menjadi elemen strategis dalam pengambilan keputusan kebijakan bank. NPL (*Non Performing Loan*) atau biasa disebut kredit bermasalah pada suatu bank yang baik adalah dibawah ambang batas yaitu 5% yang ditentukan oleh BI (Bank Indonesia), maka bank tersebut dinilai mampu menerapkan manajemen risiko yang baik, sehingga bank mampu mempertahankan kinerja yang baik dan tidak mengganggu *profitabilitas*[2]. Selain itu bank juga harus berhati-hati dalam mengelola perkreditannya untuk mencegah besarnya angka kegagalan kredit. Atas dasar tersebut hal ini dimaksudkan untuk mengukur tingkat prediksi kredit bermasalah. Harapannya adalah agar prediksi tersebut dapat menjadi pertimbangan bagi bank untuk mengambil keputusan sebelum terjadinya gagal bayar kredit[3].

Seiring dengan kemajuan teknologi informasi, didunia perbankan yang dipenuhi dengan data yang semakin banyak, kehadiran *Artificial Intelligence* (AI) dan algoritma *machine learning* mulai dibutuhkan. Karena dipandang sangat membantu dalam memprediksi kredit bermasalah. Pendekatan tradisional dalam penilaian risiko kredit kini mulai dilengkapi, bahkan digantikan, oleh metode berbasis AI[4]. Model *machine learning* dan *deep learning* memiliki potensi besar untuk mengenali pola tersembunyi pada data nominatif kredit yang bersifat kompleks dan memiliki banyak variabel, sehingga mampu meningkatkan akurasi prediksi risiko secara lebih objektif dan adaptif terhadap perubahan data. Beberapa algoritma populer yang banyak digunakan dalam konteks ini antara lain ANN, CNN, dan RF[5].

AI dianggap sebagai dasar utama dari pengembangan *deep learning*. AI adalah bidang dari ilmu komputer yang berorientasi pada pengembangan sistem pintar yang mampu menirukan cara berfikir dan kemampuan manusia. ANN termasuk dalam salah satu pendekatan AI yang dirancang seperti struktur serta mekanisme kerja otak manusia. ANN terdiri dari sejumlah *node* (neuron) yang tersusun dalam tiga bagian utama diantaranya lapisan masukan, lapisan tersembunyi dan keluaran. Dimana setiap masing-masingnya saling terhubung melalui bobot tertentu yang disesuaikan melalui proses pembelajaran (*training*) untuk mengoptimalkan hasil prediksi atau klasifikasi[6]. Pada penelitian sebelumnya model ANN dibandingkan dengan model tradisional seperti *Decision Tree* dan *Discriminant Analysis*, ANN lebih efektif dalam mengevaluasi risiko kredit nasabah di bank komersial, dengan indikator laporan keuangan perusahaan. Hal ini menunjukkan bahwa ANN lebih unggul daripada model tradisional dalam memprediksi risiko kredit perusahaan[7]. Pada penelitian lain, penggunaan ANN dalam bidang perbankan juga dilakukan, terutama dalam pengembangan proses klasifikasi nasabah untuk kelayakan kredit. Pengembangan model klasifikasi persetujuan kredit dengan membandingkan tiga algoritma, yaitu ANN, *Decision Tree*, *Naive Bayes*. Hasilnya diketahui bahwa ANN memiliki performa paling unggul[8]. Dari hal ini membuat penulis mengadopsi model ANN untuk membandingkan dengan metode lain dalam prediksi kredit bermasalah.

Selain itu dengan menggunakan data keuangan perusahaan yang mengusulkan model prediksi risiko kredit, peneliti lain menggabungkan model CNN untuk mengekstrak fitur spasial dari data masukan, LSTM untuk menangkap dependensi urutan/waktu dari data keuangan historis, dan mekanisme perhatian (*Attention Mechanism*) untuk memberikan bobot berbeda pada informasi yang relevan selama proses prediksi. Model ini dirancang untuk memperoleh akurasi prediksi yang lebih tinggi dengan memanfaatkan kemampuan CNN dalam ekstraksi karakteristik data serta keunggulan LSTM dalam mengelola data berurutan[9]. Studi lain mengusulkan pendekatan *deep learning* untuk memprediksi gagal bayar obligasi kredit menggunakan CNN. Model ini menunjukkan kemampuan yang kuat dalam menangkap interaksi fitur kompleks dari data keuangan Perusahaan. CNN merupakan model *neural network* yang dirancang secara khusus untuk mengolah data berbentuk 2 dimensi, seperti foto atau gambar digital. Dalam konteks penelitian ini, CNN digunakan pada data tabular dengan melakukan transformasi input menjadi bentuk sekuensial satu dimensi (1D). Setiap baris data yang merepresentasikan satu nasabah kredit dianggap sebagai urutan fitur numerik yang diproses oleh CNN 1D. Pemilihan CNN 1D dilakukan untuk mengeksplorasi potensi jaringan konvolusional dalam mengekstraksi pola-pola lokal dari data numerik. Dikarenakan CNN sangat populer dalam tugas-tugas klasifikasi citra, deteksi objek, dan pengenalan pola visual, maka CNN juga mulai banyak digunakan dalam sektor keuangan. Oleh karena itu CNN dianggap mampu menangkap pola kompleks antar fitur keuangan untuk mendukung keputusan klasifikasi dalam sistem peringkat kredit untuk memprediksi gagal bayar[10]. Dari berbagai studi yang telah dikaji ANN dan CNN merupakan metode yang sangat potensial dalam berbagai aplikasi perbankan dan keuangan terutama dalam klasifikasi kredit dan analisis perilaku nasabah, sehingga memberikan fleksibilitas dan daya saing tinggi bagi model pembelajaran mesin dalam sektor perbankan[11].

Adapun penelitian lain dengan pendekatan prediksi menggunakan *Random Forest*. Penelitian dengan mengadopsi *Historical Random forest* (HRF) tersebut memiliki akurasi lebih baik daripada model parameter klasik. Penelitian ini memberi landasan empirik kuat untuk evaluasi risiko kredit UKM dan model keuangan

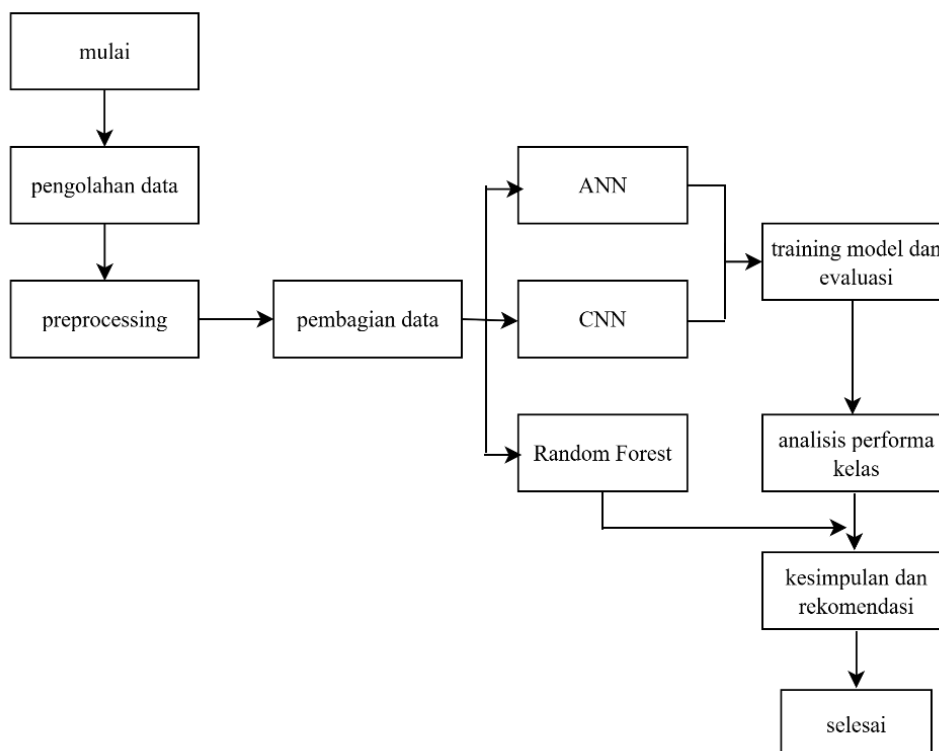
adaptif berbasis data historis. Interpretabilitas model meningkatkan transparansi dan kepercayaan pengguna, terutama dalam konteks keuangan[12]. Kemudian peneliti lain membandingkan dua metode yaitu RF dan KNN (*K-Nearest Neighbor*) untuk memprediksi tingkat kelancaran pembayaran kredit nasabah dengan hasil RF unggul karena merupakan *ensemble model*, mampu mengatasi *overfitting*, dan efisien pada dataset besar dalam menentukan kategori lancar atau tidaknya disaat pembayaran sesuai data-data yang digunakan[13]. Selain itu penelitian sebelumnya juga ada yang melakukan perbandingan performa lima algoritma *machine learning* yaitu *Logistic Regression*, KNN, *Naive Bayes*, RF, *Decision Tree* dengan tujuan untuk membangun model klasifikasi risiko kredit berdasarkan data pinjaman nasabah bank. RF unggul sebagai model prediksi risiko kredit yang paling akurat, stabil, dan seimbang. Penelitian ini menunjukkan pentingnya menggunakan model ML modern untuk sistem evaluasi kredit di industri keuangan[14]. Dalam mengembangkan sistem prediksi risiko kredit menggunakan algoritma gabungan beberapa model dan metode seleksi fitur berbasis *Information Gain* (IG), penelitian terdahulu dengan model *Extreme Gradient Boosting*, RF, *Gradient Boosting* dibandingkan dengan KNN, *Decision Tree*, ANN. Hasil akurasi tertinggi diperoleh model RF dengan gabungannya sebagai yang lebih unggul[15]. Studi lain dalam prediksi default kartu kredit membandingkan metode *Support Vector Machine*, *Deep Neural Network*, dan *Logistic Regression*. Hasilnya menunjukkan bahwa metode *machine learning* modern, termasuk RF, memiliki performa prediktif yang lebih baik dibandingkan metode statistik tradisional, terutama dalam metrik seperti F1 score, G-mean, dan AUC[16].

Berdasarkan kajian dari berbagai literatur dan hasil temuan studi sebelumnya, penulis menyimpulkan untuk menggunakan tiga metode utama dalam penelitian yaitu ANN, CNN, dan RF. Ketiga algoritma ini dipilih karena masing masing memiliki keunggulan yang spesifik. Dan diharapkan penelitian ini dapat membandingkan performa mana yang lebih unggul akurasi dalam memprediksi kredit bermasalah dalam perbankan.

2. METODE PENELITIAN

Dalam Studi ini mengadopsi metode eksperimen komputasional dengan pendekatan kuantitatif untuk membandingkan kinerja tiga algoritma kecerdasan buatan dalam memprediksi risiko kredit bermasalah, yaitu ANN (*Artificial Neural Network*) MLP (*Multilayer Perceptron*), CNN (*Convolutional Neural Network*), dan RF (*Random Forest*) dalam mengklasifikasikan data kolektibilitas kredit.

Berikut adalah gambar diagram alur proses tahapan penelitiannya :



Gambar 1. Diagram Alur Penelitian

2.1. Pengolahan Data.

Data yang digunakan adalah data primer yang berasal dari salah satu bank Pemerintah di Purbalingga. Penulis melakukan wawancara terhadap beberapa petugas bank kemudian melakukan observasi dan dokumentasi sehingga data tersebut telah melalui proses verifikasi oleh pihak bank yang bersangkutan dan mewakili kondisi aktual nasabah selama periode tertentu.

Dari penghimpunan data yang ada kemudian dijadikan acuan dalam studi yaitu data nominatif kredit berupa file excel yang berjumlah 2.253 baris data nasabah yang berisi informasi pinjaman nasabah sampai dengan awal tahun 2025. Fitur yang digunakan adalah jangka waktu, penghasilan, plafon, baki debit, kolektibilitas dan total tunggakan, sedangkan target klasifikasi adalah Kolek. Nilai target dikodekan menggunakan *LabelEncoder* agar dapat digunakan dalam model klasifikasi. Data diklasifikasikan ke dalam lima kelas kolektibilitas, yaitu: L atau lancar, DP atau Dalam Pengawasan, KL atau Kurang lancar, D atau Diragukan, dan M atau Macet. Pemodelan menggunakan bahasa pemrograman *Python*. Seluruh prosesnya di dalam pengolahan data, pelatihan model, hingga evaluasi kinerja dalam studi ini dilakukan dengan penggunaan *Python* sebagai bahasa pemrogramannya, karena memiliki pustaka (*library*) yang kuat dan mendukung komputasi ilmiah serta pembelajaran mesin (*machine learning*)[17].

2.2. Preprocessing Data

Fitur numerik distandarisasi menggunakan *StandardScaler* untuk memastikan model *neural network* bekerja optimal. Target dikonversi ke bentuk *one hot encoding* agar kompatibel dengan arsitektur ANN dan CNN yang menggunakan fungsi aktivasi *softmax*.

2.3. Pembagian Data

Sebanyak 70% dari keseluruhan data digunakan untuk pelatihan model, sementara sisanya 30% dialokasikan sebagai data pengujian menggunakan *train_test_split*. Pembagian ini dilakukan secara acak namun konsisten dengan menggunakan parameter *random_state=42*.

2.4. Training dan Evaluasi Model

a. *Random Forest*.

Model RF dibangun dengan 100 pohon keputusan. Karena RF adalah model berbasis *ensemble* dari *decision tree*, proses training dilakukan secara langsung melalui pemanggilan *fit()* tanpa *epoch*. Model langsung siap untuk evaluasi. Hasil evaluasi menunjukkan akurasi tinggi dan *confusion matrix* menunjukkan klasifikasi yang akurat di setiap kelas.

b. *Artificial Neural Network* (ANN).

Model ANN dibangun dengan 2 lapisan tersembunyi yang memanfaatkan fungsi aktivasi *ReLU* dan lapisan keluarannya menggunakan *softmax*. Model dilatih selama 50 *epoch* dan dievaluasi menggunakan data uji. Hasilnya ditampilkan dalam bentuk *confusion matrix* dan grafik akurasi/loss.

c. *Convolutional Neural Network* (CNN).

Model CNN 1D digunakan dengan mengubah input menjadi bentuk 3D. Model terdiri dari satu lapisan *MaxPooling1D*, dilanjutkan *Flatten* dan *Dense layer*. CNN juga dilatih selama 50 *epoch* dan menunjukkan hasil akurasi yang kompetitif dengan ANN dan RF.

Evaluasi performa model klasifikasi dilakukan menggunakan metrik. Berikut Adalah tabel Confusion Matrix:

Tabel 1. *Confusion Matrix*

<i>Confusion Matrix</i>		<i>Actual Class</i>	
		<i>Positive</i>	<i>Negative</i>
<i>Prediction Class</i>	<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
	<i>Negative</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

- a. *Confusion Matrix* yaitu visualisasi distribusi prediksi antar kelas.
TN (True negative) berarti bahwa model memprediksi data ada dikelas negative, dan yang benar data harusnya ada di kelas negative.
TP (True Positive) berarti bahwa model memprediksi data ada dikelas positive, dan yang benar harusnya data ada dikelas positive.
FN (False Negative) berarti bahwa model memprediksi data ada dikelas negative dan yang benar harusnya data ada dikelas positive.
FP (False Positive) berarti bahwa model memprediksi data ada dikelas positive dan yang benar harusnya data ada dikelas negative.
- b. *Accuracy* yaitu proporsi prediksi benar terhadap seluruh data. Untuk *Accuracy* dapat dihitung dengan rumus pada persamaan (1).

$$Accuracy = \frac{(TP+TN)}{TP+TN+FP+FN} \quad (1)$$

- c. *Precision* yaitu kemampuan model menghindari kesalahan positif. Untuk *Precision* dapat dihitung dengan rumus pada persamaan (2).

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

- d. *Recall* yaitu kemampuan model menangkap semua data positif yang benar. Untuk *Recall* dapat dihitung dengan rumus pada persamaan (3).

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

2.5. Analisis Performa

Ketiga model yaitu ANN, CNN dan *Random Forest* dievaluasi berdasarkan metrik *accuracy*, *precision*, *recall*, *f-1 score* dan *confusion matrix*. Hasil evaluasi menunjukkan bahwa *Random Forest* mencapai akurasi tinggi secara langsung tanpa proses *epoch*.

2.6. Kesimpulan Dan Rekomendasi

Proses penelitian ini dilakukan secara sistematis mulai dari pengumpulan dan pengolahan data hingga evaluasi model. Data yang digunakan berupa data nominatif kredit dari salah satu bank pemerintah di Purbalingga, yang telah diproses menggunakan teknik *preprocessing* seperti standardisasi fitur numerik dan konversi target menjadi *one hot encoding*. Pembagian datanya dilakukan secara acak namun tetap konsisten yang kemudian dibagi menjadi data pelatihan model dan data pengujiannya dengan proporsi 70% dan 30%. Tiga model klasifikasi diterapkan, yaitu *Random Forest*, ANN, dan CNN. *Random Forest* menunjukkan performa tinggi tanpa pelatihan berbasis *epoch*, sedangkan ANN dan CNN dilatih selama 50 epoch dan menghasilkan akurasi yang kompetitif. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *confusion matrix*. Secara keseluruhan, *Random Forest* unggul dalam akurasi langsung, sedangkan ANN dan CNN menunjukkan potensi yang baik dalam klasifikasi multikelas setelah proses pelatihan. Hasil ini menegaskan bahwa kombinasi pengolahan data yang tepat dan pemilihan model yang sesuai dapat meningkatkan performa klasifikasi dalam sistem kredit perbankan.

3. HASIL DAN PEMBAHASAN

3.1. HASIL

Perbandingannya tiga metode dalam hal ini untuk menyelesaikan tugas klasifikasi multikelas. Evaluasi dilakukan berdasarkan metrik seperti *accuracy*, *precision*, *recall*, *f1-score*, serta *confusion matrix*.

Sumber data yang digunakan berasal dari dataset internal nominatif kredit salah satu bank Pemerintah di Purbalingga. Data mencakup fitur-fitur yang relevan dalam penilaian risiko kredit yaitu Plafon Kredit, Jangka Waktu, Penghasilan Bulanan, Baki Debet, Kolektibilitas dan Total Tunggakan. Berikut Adalah hasil evaluasi dari perbandingan 3 metode ANN (*Artificial Neural Network*) MLP (*Multilayer Perceptron*), CNN (*Convolutional Neural Network*), dan RF (*Random Forest*). Masing masing menunjukan Tingkat *accuracy*, *precision*, *recall* dan *f1-score*.

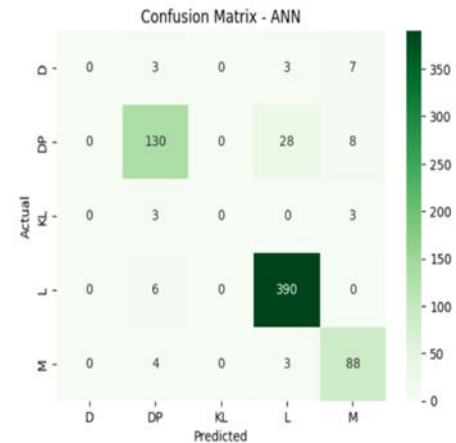
a. Hasil ANN MLP

Dari analisis performa masing masing, hasil dari pemrosesan ANN MLP adalah sebagai berikut :

=== ANN / MLP ===

Accuracy: 0.8994082840236687

	precision	recall	f1-score	support
D	0.00	0.00	0.00	13
DP	0.89	0.78	0.83	166
KL	0.00	0.00	0.00	6
L	0.92	0.98	0.95	396
M	0.83	0.93	0.88	95
accuracy			0.90	676
macro avg	0.53	0.54	0.53	676
weighted avg	0.87	0.90	0.88	676



Gambar 2. Gambar Hasil ANN MLP.

Dari gambar tersebut, hasil Akurasi model ANN MLP menunjukkan angka 89.90%, dengan *Precision (macro avg)* adalah 0.53, kemudian *Recall (macro avg)* menunjukkan angka 0.54, dan *F1-score (macro avg)* adalah 0.53. Kemudian untuk hasil *confusion matrix* ANN, maka diketahui bahwa kolektabilitas aktual D, belum dapat diprediksi dengan baik. Begitu juga dengan KL nya. Sedangkan kolektabilitas actual L mempunyai prediksi nilai yang lebih baik atau mudah dikenali. Untuk DP dan M cenderung mudah dikenali.

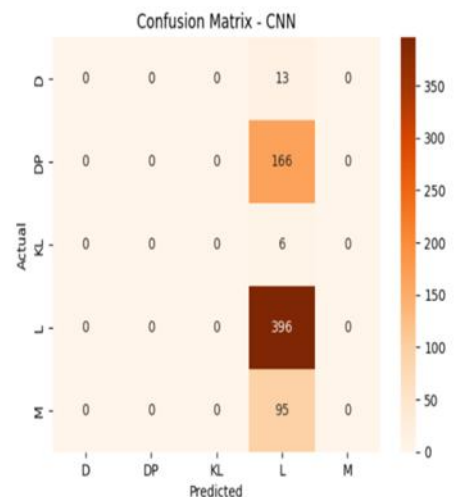
b. Hasil CNN

Berikut adalah gambar hasil pemrosesan CNN :

=== CNN ===

Accuracy: 0.5857988165680473

	precision	recall	f1-score	support
D	0.00	0.00	0.00	13
DP	0.00	0.00	0.00	166
KL	0.00	0.00	0.00	6
L	0.59	1.00	0.74	396
M	0.00	0.00	0.00	95
accuracy			0.59	676
macro avg	0.12	0.20	0.15	676
weighted avg	0.34	0.59	0.43	676

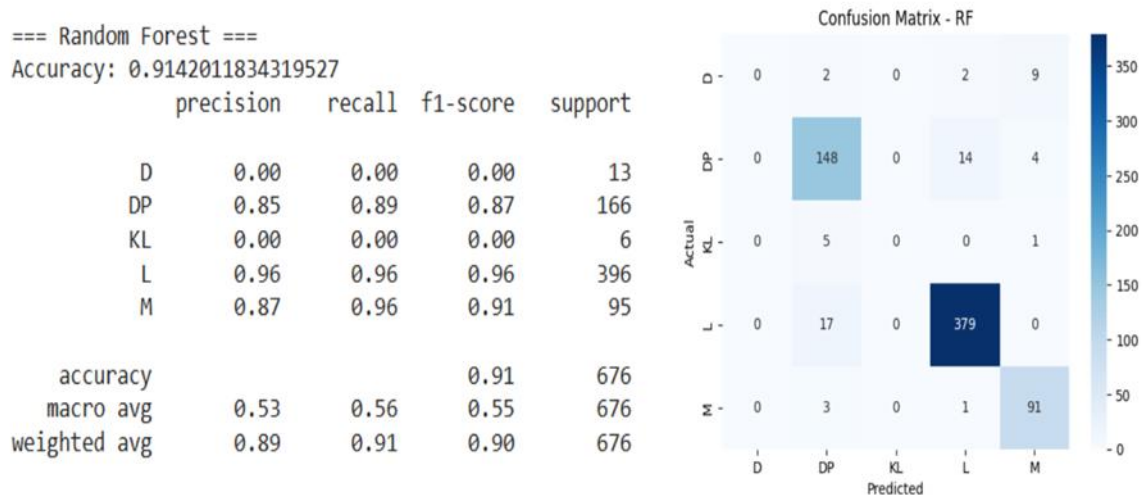


Gambar 3. Gambar Hasil CNN.

Kemudian hasil model dari CNN menunjukkan bahwa tingkat Akurasi sebesar 58.58%, sedangkan untuk *Precision (macro avg)* menunjukkan angka 0.12, kemudian hasil *Recall (macro avg)*: 0.20 dan untuk *F1-score (macro avg)*: 0.15. Sedangkan *confusion matrix* CNN maka diketahui bahwa prediksi pada seluruh kelas sangat buruk dan cenderung bias.

c. Hasil *Random Forest*

Berikut adalah gambar pemrosesan *Random forest* :



Gambar 4. Gambar hasil *Random Forest*

Untuk hasil *Random forest* hasil Akurasinya 91.42%, kemudian *Precision (macro avg)* adalah 0.53, dan untuk *Recall (macro avg)* menunjukkan angka 0.56 dan *F1-score (macro avg)* adalah 0.55. Dari gambar *confusion matrix* RF maka diketahui bahwa prediksi performa terbaik untuk semua kelas adalah L, DP dan M, tetapi untuk kelas D dan KL sulit untuk dikenali.

3.2. PEMBAHASAN

Label target berupa klasifikasi risiko kredit berdasarkan kolektibilitas (KOLEK), yang terbagi ke dalam lima kelas yaitu: Kategori L yaitu dimana debitur membayar angsuran tepat waktu sesuai tanggal. Kategori DP yaitu dimana debitur mengalami terlambat bayar antara 1 sampai dengan 90 hari. Kategori KL yaitu dimana debitur yang terlambat bayar antara 91 sampai dengan 120 hari. Kategori D dimana debitur yang terlambat bayar antara 121 sampai dengan 180 hari. Kategori M dimana debitur yang mengalami terlambat bayar lebih dari 180 hari. Berikut Adalah tabel perbandingan hasil evaluasi dari ketiga model ANN, CNN dan RF :

Tabel 2. Hasil Komparasi CNN, ANN, dan RF berdasarkan tingkat akurasi

Model	Akurasi	Precision (macro avg)	Recall (macro avg)	F1-score (macro avg)	keterangan
RF	91.42%	0.53	0.56	0.55	- Performa terbaik secara keseluruhan. - Sangat baik untuk kelas L dan M - Tidak mendeteksi kelas D dan KL
ANN/MLP	89.90%	0.53	0.54	0.53	- Performa mirip dengan RF - Agak lebih buruk dalam mengenali DP dan M dibanding FR
CNN	58.58%	0.12	0.20	0.15	- Sangat buruk untuk semua kelas kecuali L - Cenderung bias hanya ke1 kelas (semua diprediksi sebagai L) - Overfitting kemungkinan besar

Dari hasil RF menunjukkan performa terbaik secara keseluruhan, kemudian menunjukan hasil yang sangat baik untuk kelas L dan M, namun tidak mendeteksi kelas D dan KL, sedangkan kelas DP menunjukan hasil yang cukup baik. Kemudian dari hasil ANN akurasi cukup tinggi untuk kelas L dan M, kemudian untuk kelas D dan KL tidak terdeteksi, dan lebih buruk dalam mengenali DP. Kemudian untuk hasil CNN adalah sangat buruk untuk

semua kelas kecuali L, cenderung bias hanya ke satu kelas karena semua diprediksi sebagai L. Kemungkinan besar *overfitting*. Sedangkan berdasarkan perbandingan kelas L, DP, KL, D, dan M dapat dilihat pada tabel perbandingan dibawah ini :

Tabel 3. Hasil Komparasi ANN, CNN dan RF berdasarkan kelas

Aspek	Random Forest	ANN	CNN
Kelas D, KL	Tidak terdeteksi	Tidak Terdeteksi	Tidak Terdeteksi
Kelas DP	Baik (148/166 benar)	Menurun (130/166 benar)	Semua salah
Kelas L	Sangat baik (379/396)	Sangat baik (390/396)	Semua diprediksi diarahkan ke L
Kelas M	Baik (91/95)	Cukup baik (88/95)	Semua prediksi salah
Overfitting	Tidak terlihat	Tidak terlihat	Kemungkinan besar overfitting/ learning stagnan

Hasil dari tiga model klasifikasi yaitu RF, ANN, dan CNN telah diterapkan untuk menyelesaikan tugas klasifikasi multikelas pada dataset kredit kolektibilitas. Untuk model diuji performanya menggunakan sejumlah metrik *accuracy*, *precision*, *recall* serta *f1-score*, serta diperkuat lebih lanjut melalui *confusion matrix*. Selain itu, analisis terhadap performa tiap kelas juga menunjukkan perbedaan signifikan. Kelas L dikenali dengan sangat baik oleh RF dan ANN, namun CNN cenderung memprediksi semua data ke kelas ini (bias ekstrem).

Kelas M dikenali oleh RF dan ANN mampu mengklasifikasikan dengan akurasi tinggi, sedangkan CNN gagal sepenuhnya. Kelas DP dikenali oleh RF menunjukkan performa tinggi (148/166 benar), ANN mengalami penurunan (130/166 benar), sementara CNN tidak berhasil mengenali sama sekali. Kelas D dan KL menunjukkan bahwa tidak ada model yang berhasil mendeteksi kedua kelas ini secara efektif.

Random Forest dengan hasilnya menunjukkan performa terbaik secara keseluruhan. Dengan akurasi mencapai 91.42%, RF secara konsisten berhasil membedakan sebagian besar kelas, terutama kelas L, M, dan DP. Kelebihan RF terletak pada kemampuannya menangani dataset tabular tanpa memerlukan arsitektur kompleks, serta kestabilan dalam menghindari *overfitting*. Namun, model ini gagal mendeteksi kelas minoritas seperti D dan KL, yang kemungkinan besar disebabkan oleh ketidakseimbangan distribusi kelas.

Model ANN mencapai akurasi 89.90%, sedikit di bawah RF. Grafik akurasi dan loss menunjukkan bahwa model ini berhasil konvergen dengan baik. ANN cukup kompeten dalam mengenali kelas utama, tetapi cenderung mengalami penurunan performa pada kelas DP dan gagal total pada kelas minoritas. Meskipun arsitekturnya sederhana, ANN masih memerlukan tuning lebih lanjut agar mampu menangani data yang tidak seimbang.

Model CNN menunjukkan hasil yang paling buruk, dengan akurasi hanya 58.58%. Grafik *akurasi* dan *loss* menunjukkan stagnasi (*underfitting*), dan *confusion matrix* memperlihatkan bahwa model ini cenderung hanya memprediksi kelas L untuk semua data. Hal ini menandakan arsitektur CNN tidak cocok untuk data tabular yang digunakan dalam studi ini. CNN lebih sesuai untuk data sekuensial atau spasial seperti teks dan gambar, bukan data numerik tabular biasa seperti dataset yang digunakan.

4. KESIMPULAN

Dari hasil evaluasi dan analisis kinerja terhadap ketiga model yang diuji, diperoleh kesimpulan bahwa RF merupakan model yang paling unggul dalam tugas klasifikasi multikelas pada dataset kolektibilitas kredit, dengan akurasi tertinggi dan kemampuan klasifikasi yang baik terhadap kelas mayoritas. Kemudian untuk ANN (MLP) menunjukkan performa yang kompetitif dan stabil, namun memerlukan peningkatan terutama dalam menangani kelas minoritas. Dan hasil dari CNN 1D tidak cocok digunakan pada dataset tabular seperti ini, karena menunjukkan performa yang rendah dan cenderung bias terhadap satu kelas.

Temuan ini mempertegas pentingnya pemilihan algoritma yang sesuai dengan karakteristik struktur data untuk mencapai hasil klasifikasi yang optimal dalam konteks industri keuangan. Penerapan metode ensemble seperti *Random Forest* menunjukkan potensi tinggi dalam membedakan berbagai tingkat risiko, khususnya ketika dihadapkan pada data numerik yang kompleks. Rekomendasi praktis dari studi ini adalah hasil prediksi yang akurat dapat dimanfaatkan oleh lembaga keuangan untuk mempertimbangkan keputusan dalam mengambil kebijakan dan dapat merancang sistem evaluasi kelayakan kredit yang lebih presisi dan tangguh terhadap variabilitas nasabah. Selanjutnya, eksplorasi pendekatan penyeimbangan kelas serta integrasi teknik optimasi fitur diharapkan mampu meningkatkan deteksi risiko secara menyeluruh pada segmen data yang kurang terwakili.

Untuk penelitian selanjutnya dalam penanganan data tidak seimbang, kelas minoritas seperti “Kurang Lancar” atau “Diragukan” sulit dikenali oleh model karena jumlahnya jauh lebih sedikit. Penanganan bisa dilakukan dengan resampling (*oversampling/undersampling*), *Synthetic Minority Over-sampling Technique* (SMOTE), atau Pemberian bobot kelas (*class weighting*) saat pelatihan model.

DAFTAR PUSTAKA

- [1] L. J. Sinay, S. J. Latupeirissa, S. M. Pelu, and M. I. Tilukay, "the Impact of Bank-Specific Factors on Non-Performing Loan in Indonesia: Evidence From Ardl Model Approach," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 16, no. 2, pp. 675–686, 2022, doi: 10.30598/barekengvol16iss2pp675-686.
- [2] D. Zahra Yuniar, E. Suherman, D. Epty, U. Buana, and P. Karawang, "Analisis Non Performing Loan pada PT Bank BRI tbk Info Artikel ABSTRAK Sejarah artikel," *J. Ilm. Akunt. dan Keuang.*, vol. 5, no. 5, p. 2022, 2022, [Online]. Available: www.idx.co.id
- [3] N. Shonhadji, "What Most Influence on Non-Performing Loan in Indonesia? Bank Accounting Perspective with MARS Analysis," *J. Account. Strateg. Financ.*, vol. 3, no. 2, pp. 136–153, 2020, doi: 10.33005/jasf.v3i2.85.
- [4] S. I. Serengil, S. Imece, U. G. Tosun, E. B. Buyukbas, and B. Koroglu, "A Comparative Study of Machine Learning Approaches for Non Performing Loan Prediction with Explainability," *Int. J. Mach. Learn. Comput.*, vol. 12, no. 5, 2022, doi: 10.18178/ijmlc.2022.12.5.1102.
- [5] E. Ismanto and M. Novalia, "Komparasi Kinerja Algoritma C4.5, Random Forest, dan Gradient Boosting untuk Klasifikasi Komoditas Performance Comparison Between C4.5 Algorithm, Random Forests, and Gradient Boosting for Commodity Classification," *Techno.COM*, vol. 20, no. 3, pp. 400–410, 2021.
- [6] L. Alzubaidi *et al.*, *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*, vol. 8, no. 1. Springer International Publishing, 2021. doi: 10.1186/s40537-021-00444-8.
- [7] Y. Hu and J. Su, "Research on Credit Risk Evaluation of Commercial Banks Based on Artificial Neural Network Model," *Procedia Comput. Sci.*, vol. 199, pp. 1168–1176, 2021, doi: 10.1016/j.procs.2022.01.148.
- [8] P. Rahmawati, A. Larasati, and M. Marsono, "Pengembangan Model Persetujuan Kredit Nasabah Bank Dengan Algoritma Klasifikasi Naïve Bayes, Decision Tree, Dan Artificial Neural Network," *J@ti Undip J. Tek. Ind.*, vol. 17, no. 1, pp. 1–12, 2022, doi: 10.14710/jati.1.1.1-12.
- [9] J. Li, C. Xu, B. Feng, and H. Zhao, "Credit Risk Prediction Model for Listed Companies Based on CNN-LSTM and Attention Mechanism," *Electron.*, vol. 12, no. 7, pp. 1–18, 2023, doi: 10.3390/electronics12071643.
- [10] B. Feng, W. Xue, B. Xue, and Z. Liu, "Every Corporation Owns Its Image: Corporate Credit Ratings via Convolutional Neural Networks," *2020 IEEE 6th Int. Conf. Comput. Commun. ICC3 2020*, pp. 1578–1583, 2020, doi: 10.1109/ICC351575.2020.9344973.
- [11] J. Ye, Z. Zhao, E. Ghafourian, A. R. Tajally, H. A. Alkhazaleh, and S. Lee, "Optimizing the topology of convolutional neural network (CNN) and artificial neural network (ANN) for brain tumor diagnosis (BTD) through MRIs," *Heliyon*, vol. 10, no. 16, p. e35083, 2024, doi: 10.1016/j.heliyon.2024.e35083.
- [12] A. Bitetto, P. Cerchiello, S. Filomeni, A. Tanda, and B. Tarantino, "Machine learning and credit risk: Empirical evidence from small- and mid-sized businesses," *Socioecon. Plann. Sci.*, vol. 90, no. October, p. 101746, 2023, doi: 10.1016/j.seps.2023.101746.
- [13] B. Aji Santoso and A. Dwi Hartanto, "Comparison of Accuracy Levels of Random Forest and K-Nearest Neighbor (Knn) Algorithms for Classifying Smooth Bank Credit Payments," *J. Tek. Inform.*, vol. 5, no. 1, pp. 77–87, 2024, [Online]. Available: <https://doi.org/10.52436/1.jutif.2024.5.1.1195>
- [14] Y. Wang, Y. Zhang, Y. Lu, and X. Yu, "A Comparative Assessment of Credit Risk Model Based on Machine Learning —a case study of bank loan data," *Procedia Comput. Sci.*, vol. 174, pp. 141–149, 2020, doi: 10.1016/j.procs.2020.06.069.
- [15] I. Emmanuel, Y. Sun, and Z. Wang, "A machine learning-based credit risk prediction engine system using a stacked classifier and a filter-based feature selection method," *J. Big Data*, vol. 11, no. 1, 2024, doi: 10.1186/s40537-024-00882-0.
- [16] R. Bhandary and B. Ghosh, "Credit Card Default Prediction: An Empirical Analysis on Prediction Performance Between Statistical and Machine Learning Methods," vol. 2005, 2024, [Online]. Available: www.preprints.org
- [17] Bhavya Pratap Singh, "Python and Its Future Scope," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 2, no. 5, pp. 400–403, 2022, doi: 10.48175/ijarsct-4829.