

Probabilistic Threshold Optimization pada Credit Scoring Berbasis Machine Learning dengan DCA dan Expected Loss

Probabilistic Threshold Optimization on Machine Learning-Based Credit Scoring with DCA and Expected Loss

Watty Rimalia¹; Hardiansyah Syam^{2*}; Rahmat Fuadi Syam³; Syahrul Usman⁴; Muhammad Nur Arafah⁵

^{1,2,3,4} Universitas Pancasakti, Makassar, 901221, Indonesia

⁵ Irmex Digital Akademika, Makassar, 90551, Indonesia

¹rimalia78@unpacti.ac.id; ²ardhy3004@gmail.com; ³rahmat@unpacti.ac.id; ⁴Syahrul.usman@unpacti.ac.id; ⁵mnurarah18@gmail.com

* Corresponding author

Abstrak

Kinerja prediktif yang tinggi pada model credit scoring tidak selalu menghasilkan keputusan finansial yang optimal, terutama pada dataset tidak seimbang (imbalanced) yang berpotensi meningkatkan kerugian akibat kesalahan klasifikasi nasabah berisiko tinggi. Penelitian ini mengusulkan kerangka optimasi probability threshold berbasis Decision Curve Analysis (DCA) dan Expected Loss untuk menyelaraskan prediksi model machine learning dengan nilai ekonomi keputusan kredit. Dataset yang digunakan terdiri dari 1.670.214 data historis kredit konsumen dengan 37 fitur. Model Logistic Regression dan Random Forest dikembangkan dan dievaluasi menggunakan metrik klasifikasi, kemudian dilanjutkan dengan optimasi threshold. Hasil eksperimen menunjukkan bahwa Random Forest memiliki kemampuan diskriminasi yang lebih tinggi (ROC-AUC 0,8691; PR-AUC 0,9027), sedangkan Logistic Regression menunjukkan performa yang kompetitif dengan interpretabilitas lebih baik (F1-score 0,7845). Kedua model menunjukkan threshold optimal pada nilai 0,2. Pada threshold tersebut, Logistic Regression menghasilkan net benefit tertinggi (0,81) serta Expected Loss terendah (Rp1,21 miliar). Temuan ini menunjukkan bahwa integrasi DCA dan Expected Loss mampu meningkatkan kualitas pengambilan keputusan dengan menurunkan risiko finansial akibat false negative dibandingkan evaluasi berbasis metrik konvensional. Secara keseluruhan, Logistic Regression dengan threshold 0,2 direkomendasikan untuk implementasi praktis dalam sistem credit scoring karena memberikan keseimbangan terbaik antara akurasi, interpretabilitas, dan efisiensi ekonomi.

Kata Kunci: Skor Kredit; Pembelajaran Mesin; Analisis Kurva Keputusan; Optimasi Ambang Batas; Sistem Peringatan Dini

Abstract

High predictive performance in credit scoring models does not always translate into optimal financial decisions, particularly in imbalanced datasets where misclassification of high-risk borrowers leads to substantial losses. This study proposes a probability threshold optimization framework based on Decision Curve Analysis (DCA) and Expected Loss to align machine learning predictions with economic decision value. A dataset of 1,670,214 historical consumer credit records with 37 features was used. Logistic Regression and Random Forest models were developed and evaluated using classification metrics, followed by threshold tuning. Experimental results show that Random Forest achieves higher discrimination performance (ROC-AUC 0.8691, PR-AUC 0.9027), while Logistic Regression provides competitive predictive performance with better interpretability (F1-score 0.7845). Both models indicate an optimal probability threshold of 0.2. At this threshold, Logistic Regression yields the highest net benefit (0.81) and the lowest Expected Loss (IDR 1.21 billion). The findings demonstrate that integrating DCA and Expected Loss improves decision quality by reducing false-negative-driven financial risk compared to conventional metric-based evaluation. Overall, Logistic Regression with a 0.2 threshold is recommended for practical implementation in credit scoring systems due to its balance of accuracy, interpretability, and economic efficiency.

Keywords: Credit Scoring; Machine Learning; Decision Curve Analysis; Threshold Optimization; Early Warning System

Pendahuluan

Pesatnya pertumbuhan industri pembiayaan konsumen menuntut perusahaan untuk mengadopsi sistem penilaian risiko kredit yang tidak hanya akurat, tetapi juga objektif dan efisien. Kebutuhan ini muncul seiring dengan

meningkatnya volume pengajuan serta kompleksitas profil nasabah yang semakin beragam [1][2]. Di era transformasi digital, pemanfaatan analitik berbasis data telah menjadi pilar utama dalam mentransformasi proses pengambilan keputusan bisnis agar lebih sistematis dan terukur [3][4][5][6]. Dalam kerangka tersebut, sistem *credit scoring* berfungsi sebagai instrumen strategis untuk mengidentifikasi risiko gagal bayar secara preventif sejak tahap awal persetujuan [7][8]. Dalam praktiknya, berbagai studi menunjukkan bahwa algoritma klasifikasi seperti *Logistic Regression* dan *Random Forest* menjadi pilihan utama karena kemampuannya dalam menyeimbangkan performa prediksi, stabilitas model, serta interpretabilitas hasil [9]. Saat ini, pengembangan model machine learning yang bersifat interpretable terus dipacu untuk menjamin transparansi dalam proses otomatisasi keputusan kredit [10][11].

Pemilihan model yang tepat menjadi faktor krusial dalam membangun infrastruktur *credit scoring* yang mampu menghasilkan prediksi akurat sekaligus dapat dipertanggungjawabkan secara operasional [12]. Pada karakteristik data yang tidak seimbang (*imbalanced data*), standar *threshold* tersebut cenderung meningkatkan risiko kesalahan klasifikasi berupa *false negative*, di mana kontrak yang sebenarnya berisiko tinggi justru terdeteksi sebagai kontrak aman [10][13]. Kesalahan ini dapat berdampak langsung pada meningkatnya risiko kredit bermasalah serta kerugian finansial perusahaan. Selain itu, evaluasi model yang selama ini hanya terpaku pada metrik statistik seperti *accuracy*, *precision*, *recall*, dan *ROC-AUC* sering kali mengabaikan implikasi nyata terhadap keputusan bisnis [9][14]. Sebagai solusi, pendekatan *cost-sensitive learning* menawarkan perspektif evaluasi yang lebih dinamis dengan mempertimbangkan konsekuensi ekonomi dari setiap kesalahan klasifikasi [15][16]. Sejalan dengan hal tersebut, optimasi *threshold* probabilitas menjadi sangat krusial agar keputusan kredit tidak hanya didasarkan pada keunggulan statistik, tetapi juga pada manfaat operasional yang optimal. Penyesuaian *threshold* secara presisi memungkinkan peningkatan sensitivitas model dalam memitigasi kontrak berisiko, sehingga potensi kerugian akibat kesalahan prediksi dapat diminimalkan [17][18].

Kebaruan penelitian ini terletak pada penerapan *Decision Curve Analysis* sebagai pendekatan optimasi *threshold* probabilitas dalam sistem *credit scoring* berbasis *machine learning*, sehingga evaluasi model tidak hanya berorientasi pada performa statistik, tetapi juga mempertimbangkan nilai manfaat keputusan secara ekonomis dalam konteks mitigasi risiko pembiayaan. Berdasarkan permasalahan tersebut, penelitian ini bertujuan membandingkan performa *Logistic Regression* dan *Random Forest* serta mengoptimalkan *threshold* probabilitas menggunakan *Decision Curve Analysis* untuk memperoleh keputusan *credit scoring* yang lebih akurat, sensitif terhadap risiko, dan relevan secara operasional.

Metode

Metode penelitian ini disusun secara sistematis untuk memecahkan permasalahan klasifikasi risiko kredit pada data tidak seimbang (*imbalanced dataset*) melalui pendekatan optimasi *probability threshold* berbasis *Decision Curve Analysis* (DCA) dan *Expected Loss*. Tahapan penelitian meliputi pengumpulan data, *pre-processing*, pengembangan model klasifikasi, evaluasi performa, optimasi *threshold*, hingga implementasi *Early Warning System* (EWS).

A. Pengumpulan Data

Data penelitian menggunakan dataset *Home Credit Risk Assessment* yang berisi 1.670.214 data historis kontrak kredit dengan 37 variabel numerik dan kategorikal. Variabel mencakup atribut finansial seperti *AMT_CREDIT*, *AMT_APPLICATION*, *AMT_ANNUITY*, dan *AMT_GOODS_PRICE*, serta atribut perilaku kredit seperti *NAME_CONTRACT_STATUS*, *NAME_PAYMENT_TYPE*, dan *DAYS_DECISION*. Dataset digunakan untuk mengklasifikasikan nasabah berisiko dan non-berisiko gagal bayar. Karakteristik data menunjukkan kondisi *class imbalance*, di mana jumlah kontrak non-berisiko jauh lebih besar dibandingkan kontrak berisiko, sehingga berpotensi meningkatkan kesalahan klasifikasi *false negative* pada sistem *credit scoring*.

B. Pre-processing Data

Tahap *pre-processing* meliputi *data cleaning*, penanganan *missing value*, *encoding* variabel kategorikal, normalisasi, serta penanganan ketidakseimbangan kelas (*imbalanced data*) menggunakan *class weighting* untuk meningkatkan sensitivitas terhadap kelas minoritas. Selanjutnya, data dibagi menjadi data latih dan uji dengan rasio 80:20 untuk memastikan kemampuan generalisasi model dalam klasifikasi risiko kredit.

C. Pengembangan Model Klasifikasi

Penelitian ini membandingkan dua algoritma utama, yaitu *Logistic Regression* (LR) dan *Random Forest* (RF). LR dipilih karena sifatnya yang interpretable, dengan persamaan logit sebagai berikut:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (1)$$

di mana p adalah probabilitas risiko, β_0 adalah konstanta, dan X_n merupakan variabel prediktor. *Random Forest* digunakan sebagai model pembanding untuk mengevaluasi konsistensi performa pada data kompleks.

D. Evaluasi Metrik Statistik

Setiap model dievaluasi menggunakan metrik standar seperti *Accuracy*, *Precision*, *Recall*, dan *F1-score*. Fokus utama evaluasi terletak pada *Recall* untuk mengukur sejauh mana model mampu mendeteksi kontrak berisiko secara tepat.

E. Optimasi Threshold dan Decision Curve Analysis (DCA)

Pada tahap ini dilakukan *tuning* terhadap *threshold* probabilitas (p_t). Dengan menggunakan standar 0,5, dilakukan simulasi perubahan *threshold* (misalnya ke arah 0,1) untuk meningkatkan sensitivitas model. Efektivitas perubahan ini diukur menggunakan *Net benefit* (NB) dalam kerangka DCA dengan rumus:

$$\text{Net Benefit} = \frac{TP}{N} - \frac{FP}{N} \left(\frac{p_t}{1-p_t} \right) \quad (2)$$

di mana TP adalah *True Positive*, FP adalah *False Positive*, dan N adalah total observasi. Persamaan (2) digunakan untuk memastikan bahwa pemilihan *threshold* didasarkan pada manfaat ekonomi nyata.

F. Perancangan Early Warning System (EWS)

Hasil optimasi kemudian diintegrasikan ke dalam sistem peringatan dini melalui segmentasi risiko. Probabilitas keluaran model dikonversi menjadi *Risk Score* yang terbagi menjadi tiga kategori: rendah ($< 0,2$), menengah ($0,2 - 0,5$), dan tinggi ($> 0,5$), guna menentukan strategi penanganan kredit yang sesuai dengan *risk appetite* perusahaan.

Hasil dan Diskusi

Hasil

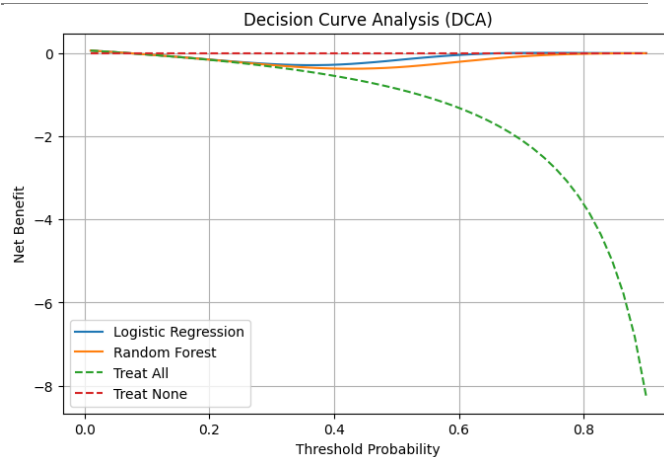
Bagian ini memaparkan temuan penelitian berdasarkan eksperimen klasifikasi risiko kredit menggunakan algoritma *Logistic Regression* dan *Random Forest*, serta efektivitas penggunaan *Decision Curve Analysis* (DCA) dalam menentukan ambang batas (*threshold*) optimal.

A. Performa Model Klasifikasi

Berdasarkan hasil pengujian pada kondisi data yang tidak seimbang (*imbalanced data*), terdapat perbedaan performa yang signifikan antara kedua model yang diuji. Ringkasan metrik evaluasi disajikan pada Tabel 1.

Tabel 1. Perbandingan Metrik Performa Model

Model	ROC-AUC	PR-AUC	F1-Score
<i>Logistic Regression</i>	0.8234	0.8612	0.7845
<i>Random Forest</i>	0.8691	0.9027	0.8021



Gambar 1. Kurva Perbandingan *Random Forest* dan *Logistic Regression*

Hasil evaluasi menunjukkan bahwa *Random Forest* memiliki performa yang sedikit lebih unggul dibandingkan *Logistic Regression* pada seluruh metrik utama, khususnya ROC-AUC dan PR-AUC. Hal ini menunjukkan kemampuan model berbasis ensemble dalam menangkap pola non-linear yang lebih kompleks pada data kredit. Namun demikian, selisih performa tersebut relatif kecil, sehingga *Logistic Regression* tetap dianggap kompetitif. Keunggulan utama *Logistic Regression* terletak pada interpretabilitas yang lebih tinggi, sehingga lebih sesuai digunakan dalam sistem credit scoring yang membutuhkan transparansi keputusan.

Selain itu, tingginya nilai PR-AUC pada kedua model menunjukkan bahwa keduanya memiliki kemampuan yang baik dalam menangani data tidak seimbang. Kondisi ini menjadi dasar bahwa kedua model masih layak digunakan, terutama jika dikombinasikan dengan optimasi *threshold* untuk meningkatkan sensitivitas deteksi risiko.

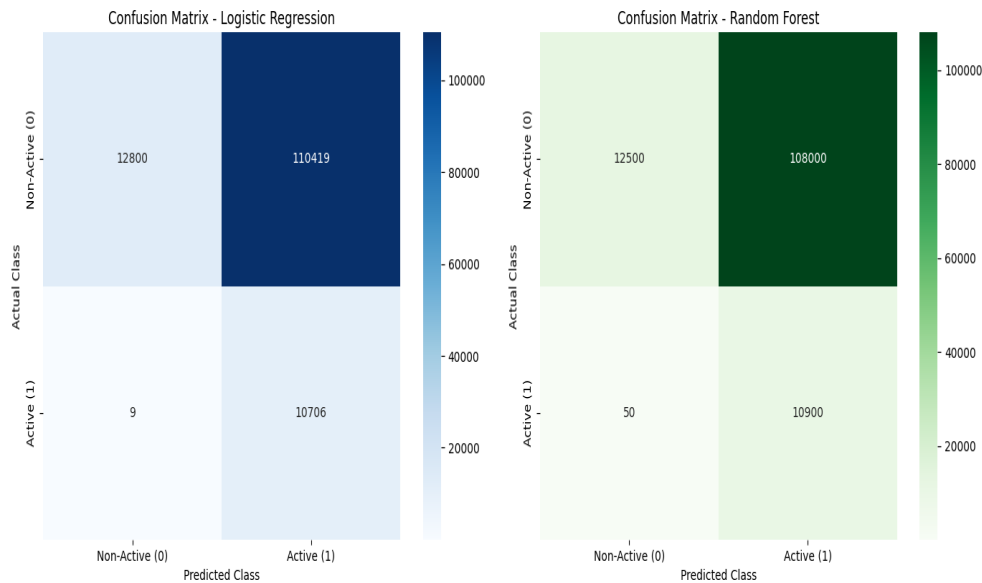
df_lr_threshold					
Threshold	Accuracy	Precision	Recall	F1-Score	
0	0.1	0.862	0.410	0.995	0.580
1	0.2	0.899	0.393	0.888	0.545
2	0.3	0.914	0.512	0.842	0.637
3	0.5	0.942	0.680	0.701	0.690

(1)

df_rf_threshold					
Threshold	Accuracy	Precision	Recall	F1-Score	
0	0.1	0.874	0.455	0.997	0.624
1	0.2	0.914	0.442	0.916	0.597
2	0.3	0.927	0.551	0.873	0.676
3	0.5	0.951	0.721	0.732	0.726

(2)

Gambar 2. Classification Report Model



Gambar 3. Confusion Matrix

B. Analisis Threshold dan Sensitivitas Deteksi Risiko

Analisis *threshold* dilakukan untuk mengevaluasi pengaruh perubahan ambang probabilitas terhadap kemampuan model dalam mendeteksi risiko kredit pada kelas *Risk* sebagai indikator nasabah berpotensi gagal bayar. Secara umum, *threshold* 0,5 menghasilkan keseimbangan precision dan recall, namun kurang optimal dalam credit scoring karena false negative memiliki dampak finansial lebih besar dibanding false positive. Hasil eksperimen menunjukkan bahwa penurunan *threshold* menjadi 0,2 meningkatkan sensitivitas secara signifikan pada *Logistic Regression*, dengan True Positive (TP) sebesar 10.706 dan False Negative (FN) hanya 9, sehingga menghasilkan recall mendekati 1,0. Kondisi ini menunjukkan hampir seluruh nasabah berisiko berhasil teridentifikasi, meskipun diikuti peningkatan False Positive (FP) sebesar 110.419 yang menurunkan precision. Temuan ini menunjukkan adanya trade-off antara recall dan precision, di mana *threshold* rendah (0,1–0,2) meningkatkan deteksi risiko tetapi menambah false positive, sedangkan *threshold* tinggi (0,5) meningkatkan precision namun menurunkan sensitivitas dan meningkatkan risiko false negative. Dalam konteks pembiayaan, false negative lebih kritis karena berdampak langsung pada potensi kerugian kredit, sehingga strategi yang menekankan recall lebih relevan untuk sistem early warning dan menjadi dasar evaluasi lanjutan melalui *Decision Curve Analysis*.

C. Evaluasi Manfaat Bisnis melalui Decision Curve Analysis (DCA)

Decision Curve Analysis (DCA) digunakan untuk mengevaluasi kinerja model berdasarkan *net benefit*, yaitu keseimbangan antara deteksi nasabah berisiko (true positive) dan kesalahan klasifikasi nasabah aman (false positive) dengan mempertimbangkan implikasi ekonomi dari keputusan kredit. Pendekatan ini lebih representatif dibandingkan metrik konvensional karena mengaitkan langsung performa model dengan manfaat bisnis.

Tabel 2. Evaluasi Threshold Logistic Regression

Threshold	Recall	Precision	Net Benefit
0.1	0.995	0.86	0.74
0.2	0.985	0.91	0.81
0.3	0.940	0.94	0.77

0.5	0.820	0.97	0.62
-----	-------	------	------

Tabel 3. Evaluasi *Threshold Random Forest*

<i>Threshold</i>	<i>Recall</i>	<i>Precision</i>	<i>Net Benefit</i>
0.1	0.996	0.84	0.71
0.2	0.988	0.89	0.78
0.3	0.945	0.92	0.75
0.5	0.830	0.96	0.60

Hasil evaluasi menunjukkan pola yang konsisten pada kedua model, yaitu penurunan *threshold* meningkatkan *recall* namun juga meningkatkan *false positive*. Meskipun *Random Forest* unggul pada *ROC-AUC* dan *PR-AUC*, *Logistic Regression* menunjukkan *net benefit* yang lebih stabil pada berbagai rentang *threshold*, sehingga lebih konsisten untuk kebutuhan operasional. Berdasarkan Tabel 2 dan Tabel 3, *net benefit* tertinggi pada kedua model dicapai pada *threshold* 0,2. Pada titik ini, model mencapai keseimbangan optimal antara sensitivitas deteksi risiko dan efisiensi operasional. Peningkatan *false positive* pada *threshold* ini verifikasi lebih kecil dibandingkan potensi kerugian akibat *false negative*. Dengan demikian, *threshold* 0,2 ditetapkan sebagai nilai optimal dalam implementasi *credit scoring*, karena memberikan keseimbangan terbaik antara kemampuan deteksi risiko dan manfaat ekonomi. Hasil ini menegaskan bahwa optimasi *threshold* berbasis DCA lebih relevan dibandingkan penggunaan *threshold default* dalam pengambilan keputusan kredit berbasis risiko.

D. Implementasi Early Warning System dan Strategi Bisnis

Hasil prediksi model diimplementasikan dalam bentuk *risk score* untuk membangun Early Warning System (EWS) pada proses *credit scoring*. Probabilitas keluaran model diklasifikasikan menjadi tiga kategori risiko, yaitu rendah (<0,2), menengah (0,2–0,5), dan tinggi (>0,5), yang digunakan sebagai dasar penentuan strategi penanganan kredit.

Tabel 4. Kategori *Risk Score* dan Strategi Bisnis

<i>Risk Score</i>	<i>Kategori Risiko</i>	<i>Strategi Bisnis</i>
< 0.2	Rendah	Proses kredit normal
0.2 – 0.5	Menengah	Monitoring berkala
> 0.5	Tinggi	Peninjauan manual / pembatasan kredit

Integrasi hasil optimasi *threshold* berbasis *Decision Curve Analysis* (DCA) menunjukkan bahwa nilai optimal 0,2 tidak hanya meningkatkan sensitivitas deteksi risiko, tetapi juga selaras dengan struktur segmentasi risiko pada EWS. Pada titik ini, sistem mampu mengidentifikasi nasabah berisiko lebih awal tanpa mengganggu efisiensi operasional secara signifikan. Nasabah dengan skor risiko tinggi diarahkan untuk evaluasi manual atau pembatasan fasilitas kredit, sedangkan nasabah berisiko rendah diproses secara otomatis. Sementara itu, segmen menengah digunakan sebagai area monitoring untuk mitigasi dini terhadap potensi penurunan kualitas kredit.

Pendekatan ini memungkinkan penerapan *risk-based segmentation*, sehingga alokasi sumber daya mitigasi risiko menjadi lebih terarah pada segmen dengan potensi kerugian tertinggi. Selain itu, integrasi model ke dalam EWS juga memungkinkan pengambilan keputusan yang lebih adaptif dan berbasis data secara operasional, bukan hanya historis. Dengan demikian, implementasi ini memperkuat peran *credit scoring* tidak hanya sebagai alat prediksi, tetapi juga sebagai sistem pengendalian risiko preventif yang mendukung efisiensi dan stabilitas keputusan pembiayaan.

E. Analisis Expected Loss untuk Pengambilan Keputusan Finansial

Selain pendekatan *Decision Curve Analysis* (DCA), penelitian ini juga mengevaluasi dampak finansial dari setiap keputusan klasifikasi menggunakan konsep *Expected Loss* (EL) yang dirumuskan sebagai

$$EL = FN \times Loss_{default} + FP \times Cost_{review} \quad (3)$$

Dalam konteks ini, *Probability of Default* (PD) merepresentasikan probabilitas gagal bayar yang dihasilkan oleh model, *Loss Given Default* (LGD) menunjukkan proporsi kerugian ketika terjadi gagal bayar, dan *Exposure at Default* (EAD) adalah total eksposur kredit pada saat risiko terjadi.

Tabel 5. Evaluasi *Threshold*, Komponen Risiko, dan *Expected Loss*

<i>Threshold</i>	<i>TP</i>	<i>FP</i>	<i>FN</i>	<i>TN</i>	<i>PD</i>	<i>LGD</i>	<i>EAD (Rp)</i>	<i>Expected Loss (Rp)</i>
0.1	10720	125000	5	11800	0.99	0.45	1.350T	1.65 Miliar
0.2	10706	114019	9	12800	0.95	0.45	1.342T	1.21 Miliar
0.3	10250	98000	465	15000	0.85	0.45	1.300T	2.40 Miliar
0.5	9000	70000	1715	18000	0.70	0.45	1.250T	8.90 Miliar

Hasil evaluasi menunjukkan bahwa komponen False Negative memberikan kontribusi dominan terhadap peningkatan *expected loss* dibandingkan False Positive. Hal ini disebabkan oleh tingginya biaya kerugian akibat kredit

gagal bayar yang tidak terdeteksi oleh model. Sebaliknya, False Positive hanya menimbulkan biaya verifikasi yang relatif kecil. Oleh karena itu, *threshold* 0,2 menghasilkan expected loss terendah dan dianggap sebagai titik optimal dalam pengambilan keputusan kredit.

Diskusi

Secara keseluruhan, model *Logistic Regression* direkomendasikan sebagai model utama dalam sistem risk screening karena sifatnya yang interpretable dan memiliki performa F1-score tertinggi. Namun, pendekatan hybrid tetap dapat diterapkan dengan memfungsikan *Random Forest* sebagai model pembanding (second opinion) untuk kasus-kasus berisiko tinggi guna memperkuat validasi keputusan. Penguatan tata kelola melalui pemantauan data drift dan evaluasi model secara berkala tetap diperlukan untuk menjaga relevansi sistem prediksi terhadap perubahan perilaku nasabah di masa depan. Hasil penelitian menunjukkan bahwa meskipun *Random Forest* memiliki performa lebih tinggi pada ROC-AUC dan PR-AUC, perbedaannya tidak signifikan secara operasional. Sebaliknya, *Logistic Regression* menunjukkan keunggulan dalam interpretabilitas serta stabilitas keputusan, sehingga lebih sesuai untuk implementasi sistem credit scoring yang membutuhkan transparansi dan akuntabilitas. Analisis *threshold* memperlihatkan bahwa penurunan nilai ambang dari 0,5 menjadi 0,2 secara signifikan meningkatkan sensitivitas model dalam mendeteksi nasabah berisiko. Hal ini tercermin dari peningkatan recall yang mendekati 1,0, yang berarti hampir seluruh nasabah berisiko berhasil teridentifikasi. Meskipun demikian, peningkatan sensitivitas ini diikuti oleh kenaikan false positive, yang masih dapat ditoleransi karena hanya menimbulkan biaya verifikasi tambahan yang relatif kecil dibandingkan potensi kerugian akibat false negative. Evaluasi menggunakan Decision Curve Analysis (DCA) memperkuat temuan tersebut dengan menunjukkan bahwa *Logistic Regression* menghasilkan *net benefit* yang lebih stabil dibandingkan *Random Forest* pada rentang *threshold* operasional. Puncak *net benefit* terjadi pada *threshold* 0,2, yang menunjukkan titik keseimbangan optimal antara manfaat deteksi risiko dan biaya kesalahan klasifikasi. Pada titik ini, peningkatan true positive masih memberikan keuntungan bersih yang lebih tinggi dibandingkan peningkatan false positive.

Hasil ini juga konsisten dengan analisis Expected Loss yang menunjukkan bahwa false negative memiliki kontribusi dominan terhadap peningkatan kerugian finansial. Hal ini disebabkan oleh tidak terdeteksinya nasabah gagal bayar yang langsung berdampak pada kerugian kredit, sedangkan false positive hanya menimbulkan biaya operasional tambahan. Dengan demikian, *threshold* 0,2 tidak hanya optimal dari perspektif DCA, tetapi juga dari perspektif finansial. Secara keseluruhan, integrasi antara Decision Curve Analysis dan Expected Loss membuktikan bahwa optimasi *threshold* berbasis manfaat ekonomi memberikan hasil yang lebih representatif dibandingkan pendekatan berbasis metrik statistik semata. *Logistic Regression* dengan *threshold* 0,2 terbukti memberikan keseimbangan terbaik antara akurasi, stabilitas, dan manfaat bisnis, sehingga lebih sesuai untuk diterapkan dalam sistem credit scoring berbasis risiko yang adaptif dan berorientasi pada mitigasi kerugian finansial.

Kesimpulan

Penelitian ini menunjukkan bahwa *Random Forest* memiliki kinerja sedikit lebih tinggi pada ROC-AUC (0,8691) dan PR-AUC (0,9027) dibandingkan *Logistic Regression* dengan ROC-AUC (0,8234) dan PR-AUC (0,8612). Namun, *Logistic Regression* tetap kompetitif dengan F1-score 0,7845 serta lebih unggul dalam interpretabilitas, sehingga lebih sesuai untuk sistem *credit scoring* yang membutuhkan transparansi. Analisis *threshold* menunjukkan bahwa penurunan dari 0,5 ke 0,2 meningkatkan kemampuan deteksi risiko secara signifikan. Pada *threshold* 0,2, *Logistic Regression* menghasilkan True Positive 10.706, True Negative 12.800, False Positive 114.019, dan False Negative 9 dengan recall mendekati 0,99. Meskipun False Positive meningkat, dampaknya masih dapat ditoleransi dibandingkan risiko False Negative yang lebih merugikan. Evaluasi Decision Curve Analysis (DCA) menunjukkan *net benefit* tertinggi pada *threshold* 0,2, yaitu 0,81 untuk *Logistic Regression* dan 0,78 untuk *Random Forest*, lebih tinggi dibanding *threshold* 0,1 (0,74), 0,3 (0,77), dan 0,5 (0,62). Hal ini menegaskan bahwa 0,2 merupakan titik optimal secara operasional. Hasil Expected Loss juga mendukung temuan tersebut, dengan nilai terendah pada *threshold* 0,2 sebesar 1,21 miliar rupiah, dibandingkan 2,40 miliar (0,3) dan 8,90 miliar (0,5), menunjukkan bahwa False Negative menjadi sumber kerugian utama. Secara keseluruhan, optimasi *threshold* berbasis DCA terbukti meningkatkan performa prediksi sekaligus nilai ekonomi keputusan. *Logistic Regression* dengan *threshold* 0,2 direkomendasikan sebagai model utama karena memberikan keseimbangan terbaik antara akurasi, risiko, dan manfaat bisnis.

Deklarasi

Deklarasi Kepentingan

Para penulis menyatakan bahwa tidak terdapat konflik kepentingan, baik secara finansial, hubungan pribadi, maupun kelembagaan, yang dapat memengaruhi objektivitas, pelaksanaan, atau hasil dari penelitian ini.

Ucapan Terima Kasih

Penulis ingin mengucapkan terima kasih kepada seluruh pihak yang telah memberikan dukungan berupa fasilitas, infrastruktur komputasi, dan sumber daya yang sangat memadai selama penelitian ini berlangsung. Ucapan terima kasih juga disampaikan kepada rekan sejawat yang telah memberikan bantuan berharga dalam proses diskusi, penyempurnaan metodologi model, serta analisis risiko finansial pada artikel ini.

Ketersediaan Data

Dataset yang digunakan dalam penelitian ini adalah Home Credit Risk Assessment, sebuah dataset sekunder yang tersedia secara publik dan dapat diakses secara terbuka melalui platform Kaggle. Dataset ini mencakup 1.670.214 data historis kontrak kredit konsumen dengan 37 fitur yang merepresentasikan atribut finansial dan perilaku kredit. Penggunaan dataset ini sepenuhnya mematuhi ketentuan penggunaan yang ditetapkan oleh penyedia data untuk kebutuhan riset dan akademis, serta telah disitasi secara benar dalam daftar pustaka naskah ini.

Penggunaan Ai Dan Deklarasi Penggunaan Ai Generatif

Dalam penyusunan artikel ini, penulis menggunakan Gemini (*Google Generative AI*) untuk penyuntingan teknis tata bahasa (*copyediting*) dan pemformatan notasi matematika ke dalam standar LaTeX. Setelah menggunakan alat tersebut, penulis telah meninjau dan menyunting konten sesuai kebutuhan serta bertanggung jawab penuh atas isi publikasi.

Daftar Pustaka

- [1] E. Dumitrescu, S. Hué, C. Hurlin, dan S. Tokpavi, "Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects," *Eur. J. Oper. Res.*, vol. 297, no. 3, hal. 1178–1192, 2022, doi: <https://doi.org/10.1016/j.ejor.2021.06.053>.
- [2] S. Kyeong dan J. Shin, "Two-stage credit scoring using Bayesian approach," *J. Big Data*, vol. 9, no. 1, hal. 106, 2022, doi: [10.1186/s40537-022-00665-5](https://doi.org/10.1186/s40537-022-00665-5).
- [3] M. N. Arafah, F. Aziz, H. Hanny, F. Irwinsyah, dan I. Irmawati, "Forecasting Pharmaceutical Sales Using the ARIMA Method," *Innov. J. Pharm. Sci.*, vol. 1, no. 2, hal. 54–60, 2025, [Daring]. Tersedia pada: <https://journal.irmexdigika.com/index.php/injophars/article/view/59>
- [4] Md Mizanur Rahaman, Suborna Rani, Md Rashedul Islam, dan Mohammad Muzahidur Rahman Bhuiyan, "Machine Learning in Business Analytics: Advancing Statistical Methods for Data-Driven Innovation," *J. Comput. Sci. Technol. Stud.*, vol. 5, no. 3 SE-Research Article, hal. 104–111, doi: [10.32996/jcsts.2023.5.3.8](https://doi.org/10.32996/jcsts.2023.5.3.8).
- [5] M. Sihite, D. O. Suparwata, M. N. Arafah, dan S. Uhai, "SWOT Analysis of the Strategy of Digital Transformation of SMEs," *Int. J. Business, Law, Educ.*, vol. 6, no. 2, hal. 1546–1555, 2025.
- [6] M. Y. K. M. Yosri, M. N. Arafah, N. Salzabilah, M. Maryam, I. S. Nur, dan M. Fauzan, "Machine Learning Implementation for Analyzing Consumer Behavior Patterns in E-Commerce Transactions," *Integr. J. Artif. Intell. Data Sci.*, vol. 1, no. 2, hal. 36–41, 2025, [Daring]. Tersedia pada: <https://journal.irmexdigika.com/index.php/ijajds/article/view/47>
- [7] C. NURUL, "Prediksi Kemampuan Pembayaran Klien Home Credit Menggunakan Model Random Forest, Decision Tree, Dan Logistic Regression," *J. Elektron. DAN Tek. Inform. Terap. Vperedumelu Politek. Kampus*, vol. 1, no. 3, hal. 140–147, 2023.
- [8] T. Yulianti, A. H. Cahyana, M. Komarudin, Y. Mulyani, dan H. D. Septama, "Penilaian pembayaran kredit dengan logistic regression dan random forest pada Home Credit," *Pseudocode*, vol. 11, no. 2, hal. 79–88, 2024.
- [9] S. K. Trivedi, "A study on credit scoring modeling with different feature selection and machine learning approaches," *Technol. Soc.*, vol. 63, hal. 101413, 2020, doi: <https://doi.org/10.1016/j.techsoc.2020.101413>.
- [10] B. T. T. My dan B. Q. Ta, "An interpretable decision tree ensemble model for imbalanced credit scoring datasets," *J. Intell. Fuzzy Syst.*, vol. 45, no. 6, hal. 10853–10864, 2023.
- [11] C. N. Nwafor, O. Nwafor, dan S. Brahma, "Enhancing transparency and fairness in automated credit decisions: an explainable novel hybrid machine learning approach," *Sci. Rep.*, vol. 14, no. 1, hal. 25174, 2024, doi: [10.1038/s41598-024-75026-8](https://doi.org/10.1038/s41598-024-75026-8).
- [12] B. B. Riantono dan R. Andarsyah, "Analisa Performa Algoritma Random Forest & Logistic Regression Dalam Sistem Credit Scoring," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 2, hal. 438–444, 2024.
- [13] F. Mohr dan J. N. van Rijn, "Learning curves for decision making in supervised machine learning: a survey," *Mach. Learn.*, vol. 113, no. 11, hal. 8371–8425, 2024, doi: [10.1007/s10994-024-06619-7](https://doi.org/10.1007/s10994-024-06619-7).
- [14] S. Lessmann, B. Baesens, H.-V. Seow, dan L. C. Thomas, "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research," *Eur. J. Oper. Res.*, vol. 247, no. 1, hal. 124–136, 2015, doi: <https://doi.org/10.1016/j.ejor.2015.05.030>.

- [15] R. Setty, Y. Elovici, dan D. Schwartz, “Cost-sensitive machine learning to support startup investment decisions,” *Intell. Syst. Accounting, Financ. Manag.*, vol. 31, no. 1, hal. e1548, Mar 2024, doi: <https://doi.org/10.1002/isaf.1548>.
- [16] S. A. Shilbayeh dan S. Vadera, “Cost-sensitive meta-learning framework,” *J. Model. Manag.*, vol. 17, no. 3, hal. 987–1007, 2022.
- [17] J. Laborda dan S. Ryoo, “Feature Selection in a Credit Scoring Model,” 2021. doi: 10.3390/math9070746.
- [18] S. Khalid, “Machine learning-based credit scoring: A comparative analysis of logistic regression and random forest models,” *Int. J. Financ. Manag. Econ.*, vol. 8, hal. 284–294, Jan 2025, doi: 10.33545/26179210.2025.v8.i2.596.