

TESIS

**EVALUASI KOMPARATIF LOGISTIC REGRESSION, LIGHTGBM,
DAN FT-TRANSFORMER UNTUK PREDIKSI RISIKO KREDIT:
VALIDASI BERBASIS UJI STATISTIK NON-PARAMETRIK DAN
INTERPRETABILITAS SHAP**



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Oleh:

FIRDA MAULIDA PRAYETNO

NIM: 24206051020

PROGRAM STUDI INFORMATIKA

**PROGRAM MAGISTER FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA**

YOGYAKARTA

2026

PERNYATAAN BEBAS PLAGIASI

PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan di bawah ini:

Nama : Firda Maulida Prayetno
NIM : 24206051020
Jenjang : Magister
Program Studi : Informatika

Menyatakan bahwa naskah tesis ini secara keseluruhan benar-benar bebas dari plagiasi. Jika di kemudian hari terbukti melakukan plagiasi, maka saya siap ditindak sesuai ketentuan hukum yang berlaku.

Yogyakarta, 29 Juni 2026

Saya yang menyatakan,



Firda Maulida Prayetno

NIM: 24206051020

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PENGESAHAN



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-2071/Un.02/DST/PP.00.9/08/2026

Tugas Akhir dengan judul : EVALUASI KOPERATIF LOGISTIK REGRESSION, LIGHT GBM, FT-TRANSFORMER UNTUK PREDIKSI RESIKO KREDIT, VALIDASI UJI STATISTIK NON PARAMETRIK DAN INTERPRETABILITAS SHAP.

yang dipersiapkan dan disusun oleh:

Nama : FIRDA MAULIDA PRAYETNO, S.Kom
Nomor Induk Mahasiswa : 24206051020
Telah diujikan pada : Rabu, 12 Agustus 2026
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D.
SIGNED

Valid ID: 6a88094db377



Penguji I

Dr. Agung Fatwanto, S.Si., M.Kom.
SIGNED

Valid ID: 6a86d9fa596d



Penguji II

Siti Mutmainah, S.Kom., M.Cs., Ph.D.
SIGNED

Valid ID: 6a86f852b923



Yogyakarta, 12 Agustus 2026
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wārdati, M.Si.
SIGNED

Valid ID: 6a8ba6fa52987

NOTA DINAS PEMBIMBING

NOTA DINAS PEMBIMBING

Kepada Yth.,
Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga
Yogyakarta

Assalamu 'alaikum wr. wb.

Setelah melakukan bimbingan, arahan, dan koreksi terhadap penulisan tesis yang berjudul:

**EVALUASI KOMPARATIF LOGISTIC REGRESSION, LIGHTGBM, DAN
FT-TRANSFORMER UNTUK PREDIKSI RISIKO KREDIT: VALIDASI
BERBASIS UJI STATISTIK NON-PARAMETRIK DAN
INTERPRETABILITAS SHAP**

Yang ditulis oleh:

Nama : Firda Maulida Prayetno
NIM : 24206051020
Jenjang : Magister
Prodi : Informatika

Saya berpendapat bahwa tesis tersebut sudah dapat diajukan kepada Magister Informatika UIN Sunan Kalijaga untuk diujikan dalam rangka memperoleh gelar Magister Informatika.

Wassalamu'alaikum wr. wb.

Yogyakarta, 21 April 2026
Pembimbing,


Ir. Maria Ulfah Siregar, S.Kom., MIT., Ph.D.
NIP. 19780106 200212 2 001

ABSTRAK

Penelitian ini melakukan evaluasi komparatif antara Regresi Logistik, LightGBM, dan FT-Transformer dalam memprediksi risiko gagal bayar menggunakan dataset *Give Me Some Credit* (150.000 sampel) dengan tingkat ketidakseimbangan kelas ekstrem (1:13,96). Guna mengatasi bias data, teknik *Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan secara lokal dalam skema 15-fold Stratified Cross-Validation, dengan optimasi hiperparameter berbasis *framework* Optuna. Hasil pengujian menunjukkan LightGBM mencapai performa superior dengan AUC-ROC akhir sebesar 0,8631, mengungguli FT-Transformer (0,8369) dan Regresi Logistik (0,7959). Validasi inferensial melalui Friedman Test ($X^2 = 30, p < 0,001$) dan uji post-hoc Wilcoxon Signed-Rank mengonfirmasi bahwa perbedaan performa antar-model tersebut signifikan secara statistik dengan ukuran efek (Cliff's Delta) yang besar. Untuk menembus sifat *black-box* model, penerapan *Explainable Artificial Intelligence* (XAI) berbasis SHAP (*SHapley Additive Explanations*) secara global berhasil mengidentifikasi riwayat keterlambatan pembayaran sebagai prediktor risiko utama pada LightGBM dan FT-Transformer, sementara analisis lokal via *waterfall plot* mampu mendiagnosis penyebab kegagalan prediksi pada kasus *False Negative*. Penelitian ini membuktikan bahwa integrasi validasi statistik non-parametrik dan teknik interpretabilitas SHAP sangat krusial dalam menghasilkan model credit scoring yang akurat dan transparan.

Kata Kunci: Credit Scoring, Logistic Regression, LightGBM, FT-Transformer, Uji Non-Parametrik, SHAP.

MOTTO

“Allah Tidak Akan Membebani Sesorang Melainkan Sesuai Dengan
Kesanggupannya”

(QS: Al-Baqarah Ayat 286)

"Allah akan meninggikan orang-orang yang beriman di antaramu dan orang-orang yang diberi ilmu beberapa derajat."

(QS. Al-Mujadilah [58]: 11)

Knowledge grows when curiosity meets perseverance.

“Ilmu pengetahuan berkembang ketika rasa ingin tahu dipadukan dengan
ketekunan”

(Firda Maulida Prayetno)

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

KATA PENGANTAR

Bismillahirrahmanirrahim

Puji dan syukur penulis panjatkan ke hadirat Allah Swt. atas segala limpahan rahmat, taufik, hidayah, serta karunia-Nya, sehingga penulis dapat menyelesaikan tesis yang berjudul "Evaluasi Komparatif Logistic Regression, LightGBM, dan FT-Transformer untuk Prediksi Risiko Kredit: Validasi Berbasis Uji Statistik Non-Parametrik dan Interpretabilitas SHAP" sebagai salah satu syarat untuk memperoleh gelar Magister pada Program Studi Magister Informatika, Universitas Islam Negeri Sunan Kalijaga Yogyakarta.

Tesis ini disusun sebagai bentuk kontribusi ilmiah dalam pengembangan metode prediksi risiko kredit (credit risk prediction) melalui pendekatan komparatif terhadap tiga algoritma machine learning yang memiliki karakteristik berbeda, yaitu Logistic Regression sebagai model statistik klasik yang banyak digunakan dalam industri keuangan, LightGBM sebagai algoritma berbasis gradient boosting yang dikenal memiliki performa tinggi, serta FT-Transformer sebagai pendekatan deep learning modern yang dirancang untuk pengolahan data tabular. Penelitian ini tidak hanya berfokus pada aspek akurasi prediksi, tetapi juga mengedepankan validasi ilmiah melalui uji statistik non-parametrik serta interpretabilitas model menggunakan SHAP (SHapley Additive Explanations), sehingga hasil penelitian diharapkan mampu memberikan kontribusi baik secara teoritis maupun praktis dalam pengembangan sistem penilaian risiko kredit yang lebih akurat, transparan, dan dapat dipertanggungjawabkan.

Penulis menyadari bahwa penyusunan tesis ini tidak akan dapat terselesaikan tanpa bantuan, dukungan, bimbingan, doa, serta motivasi dari

berbagai pihak. Oleh karena itu, dengan segala kerendahan hati penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Rektor Universitas Islam Negeri Sunan Kalijaga Yogyakarta beserta seluruh jajaran pimpinan universitas.
2. Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
3. Ketua Program Studi Magister Informatika beserta seluruh dosen yang telah memberikan ilmu, pengalaman, serta wawasan akademik selama masa perkuliahan.
4. Bapak/Ibu Dosen Pembimbing Tesis yang telah meluangkan waktu, tenaga, pikiran, serta memberikan arahan, kritik, dan saran yang sangat berharga sehingga tesis ini dapat diselesaikan dengan baik.
5. Seluruh dosen, tenaga kependidikan, dan civitas akademika Universitas Islam Negeri Sunan Kalijaga Yogyakarta yang telah memberikan pelayanan akademik selama proses studi.
6. Kedua orang tua tercinta, keluarga, serta seluruh saudara yang senantiasa memberikan doa, kasih sayang, semangat, dan dukungan tanpa henti kepada penulis.
7. Rekan-rekan mahasiswa Magister Informatika dan semua pihak yang tidak dapat disebutkan satu per satu yang telah memberikan bantuan, motivasi, serta kebersamaan selama proses penyusunan tesis ini.

Penulis menyadari sepenuhnya bahwa tesis ini masih memiliki berbagai keterbatasan dan jauh dari kata sempurna. Oleh karena itu, kritik dan saran yang bersifat membangun sangat diharapkan demi penyempurnaan penelitian ini pada masa mendatang.

Akhir kata, penulis berharap semoga tesis ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan, khususnya di bidang kecerdasan buatan (Artificial Intelligence), machine learning, analisis data, serta

manajemen risiko kredit. Penulis juga berharap penelitian ini dapat menjadi salah satu referensi bagi akademisi, praktisi industri keuangan, maupun peneliti selanjutnya dalam mengembangkan model prediksi risiko kredit yang lebih efektif, andal, dan dapat diinterpretasikan.

Wassalamu'alaikum warahmatullahi wabarakatuh.

Yogyakarta, 05 Agustus 2026

Hormat Penulis,



Firda Maulid Prayetno



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

TESIS	1
PERNYATAAN BEBAS PLAGIASI.....	i
HALAMAN PENGESAHAN	ii
NOTA DINAS PEMBIMBING	iii
ABSTRAK	iv
MOTTO	v
KATA PENGANTAR.....	vi
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN	10
A. Latar Belakang.....	10
B. Rumusan Masalah	15
C. Batasan Masalah	15
D. Tujuan Penelitian.....	17
E. Manfaat Penelitian.....	17
BAB II KAJIAN PUSTAKA DAN KAJIAN TEORI	19
A. Kajian Pustaka	19
1. Uji Statistik Non Parametrik	19
2. Interpretabilitas SHAP.....	23
3. Model.....	27
4. Kredit Scoring	33
B. Landasan Teori	37
1. Credit Scoring.....	37
2. Peer-to-Peer (P2P) Lending.....	38
3. Machine Learning untuk Credit Scoring	39
4. Regresi Logistik (LR).....	40
5. Gradient Boosting dan LightGBM	41

6.	Deep Learning untuk Data Tabular	42
7.	FT-Transformer	43
8.	Penanganan Ketidakseimbangan Kelas dengan SMOTE.....	44
9.	Hyperparameter Tuning dengan Optuna	45
10.	AUC-ROC	46
11.	K-fold	48
12.	Kolmonogrov.....	49
13.	Friedman Test.....	50
14.	Wilconxon Signed-Rank Test.....	51
15.	Effect Size Cliff's Delta	53
16.	Interpretabilitas Model dengan SHAP.....	55
BAB III METODE PENELITIAN.....		57
A.	Alat dan Bahan	57
B.	Data Penelitian.....	58
C.	Tahapan Penelitian	61
1.	Konfigurasi Parameter Eksperimen.....	63
2.	Pemuatan dan Validasi Dataset	64
3.	Eksplorasi Data (<i>Exploratory Data Analysis</i>)	65
a.	Analisis Distribusi Kelas Target.....	65
b.	Analisis Statistik Deskriptif Fitur Numerik.....	66
c.	Analisis Nilai Hilang (<i>Missing Value</i>)	66
d.	Analisis Korelasi Antar Fitur.....	67
4.	Pembagian Data (Data Splitting).....	68
5.	Praproses Data	69
6.	Pelatihan dan Optimasi Model	71
a.	Class Imbalance.....	71
b.	Konfigurasi Eksperimen dan Penyetelan Hiperparameter	72
7.	Evaluasi model pada data uji (test set)	75
8.	Analisis Struktur Data untuk Perbandingan Performa Model	77
9.	Kolmonogrov.....	78

10. Friedman Test.....	80
11. Wilconxon Signe Rank Test.....	81
12. Cliff delta.....	82
13. Interpretabilitas SHAP.....	82
BAB IV HASIL DAN PEMBAHASAN.....	85
A. Hasil Penelitian.....	85
1. Analisis Data (Exploratory Data Analysis)	85
2. Pembagian Data dan Preprocessing.....	87
3. Hasil Hyperparameter Tuning dan Cross-Validasi 15-Fold.....	88
4. Evaluasi pada Data Uji (Test Set)	90
5. Hasil Pengujian Statistik Non-Parametrik.....	92
6. Hasil Analisis SHAP dan Interpretabilitas Model.....	97
B. Pembahasan	103
1. Analisis Performa Prediksi dan Uji Signifikansi Statistik Antar Model	103
2. Analisis Besaran Efek Performa Menggunakan Cliff's Delta	105
3. Interpretabilitas SHAP.....	106
BAB V PENUTUP	109
A. Kesimpulan.....	109
B. Saran	111
DAFTAR PUSTAKA.....	115
DAFTAR RIWAYAT HIDUP	125

DAFTAR TABEL

Tabel 3. 1 Daftar Fitur Dataset Give Me Some Credit.....	59
Table 3.2 Class Imbalance.....	60
Tabel 3.3 Parameter Konfigurasi Eksperimen.....	63
Tabel 3.4 Identifikasi Nilai Hilang (Missing Value).....	67
Tabel 3.5 Hasil Pembagian Data Latih dan Data Uji	68
Tabel 3.6 Rancangan Pipeline Praproses Data per Model.....	69
Tabel 4.1 Hasil Pembagian Data Latih dan Data Uji	87
Tabel 4.2 Hasil Hyperparameter Tuning dan Cross-Validasi 15-Fold.....	88
Tabel 4.3 Skor AUC-ROC per Fold (15-Fold Cross-Validation)	89
Tabel 4.4 Hasil Evaluasi pada Data Uji	90
Tabel 4.5 Statistik Deskriptif AUC-ROC per Fold	92

DAFTAR GAMBAR

Gambar 2.1 Kajian Pustaka	37
Gambar 3.1 Tahapan penelitian.....	62
Gambar 3.2 Hyperparameter Logistic Regression	73
Gambar 3.3 Hyperparameter LightGBM	74
Gambar 3.4 Hyperparameter FT-Transformer	75
Gambar 4.1 Distribusi Kelas Target (Non-Default vs Default).....	85
Gambar 4.2 Proporsi Missing Value	86
Gambar 4.3 Matriks Korelasi Antar Fitur Numerik	86
Gambar 4.4 ROC Curve pada Data Uji	91
Gambar 4.5 Distribusi AUC-ROC per Fold Boxplot 3 Model.....	92
Gambar 4.6 Kolmogorov-Smirnov.....	93
Gambar 4.7 Hasil Friedman Test	94
Gambar 4.8 Hasil Post-Hoc Wilcoxon Signed-Rank Test dengan Koreksi Holm ...	95
Gambar 4.9 Effect Size Cliff's Delta	96
Gambar 4.10 SHAP Global Feature Importance.....	97
Gambar 4.11 SHAP Directional Impact Analysis	99

BAB I

PENDAHULUAN

A. Latar Belakang

Perkembangan teknologi keuangan (*fintech*) dalam satu dekade terakhir telah mentransformasi model intermediasi keuangan global secara fundamental (Monje *et al.*, 2025). Platform *peer-to-peer* (P2P) *lending* menjadi salah satu model yang paling disruptif (Souadda *et al.*, 2025), menghubungkan pemberi pinjaman dengan peminjam tanpa perantara bank konvensional dan memperluas akses keuangan bagi segmen *underbanked* yang sulit dilayani perbankan tradisional (Sanz-guerrero and Arroyo, 2025). Namun di balik pertumbuhan pesat ini tersimpan risiko kredit yang signifikan.

Data P2P *lending* memiliki karakteristik yang kompleks karena banyak peminjam memiliki riwayat kredit yang terbatas (*thin-file*) atau bahkan tidak memiliki riwayat kredit sama sekali (*no-file*) (Friedrich, Winkel and Reichenbach, 2024). Data yang digunakan sebagai input model terdiri dari berbagai jenis tipe data, yaitu data angka (numerik) dan data kategori (kategorikal), distribusi kelas sangat tidak seimbang dengan proporsi gagal bayar yang rendah (Soni and Jethava, 2026), serta sinyal alternatif berupa data perilaku digital menjadi prediktor penting (Xu, Tian and Fu, 2022). Kondisi ini menantang efektivitas model statistik klasik seperti regresi logistik yang lama menjadi tulang punggung credit scoring (Dumitrescu *et al.*, 2022). Meskipun regresi logistik masih banyak digunakan karena interpretabilitasnya, kompleksitas data kredit modern dan hubungan nonlinear antar fitur menyebabkan keterbatasan performa

model statistik klasik tersebut dibandingkan pendekatan machine learning modern (Banna, 2025).

Berbagai algoritma *machine learning* telah diaplikasikan secara luas dalam *credit scoring*, mencakup Random Forest, XGBoost, CatBoost *upport Vector Machine*, *Multilayer Perceptron*, hingga *Linear Discriminant Analysis*. Keberagaman model tersebut mencerminkan bahwa tidak ada satu algoritma yang secara universal unggul pada seluruh konteks *credit scoring* performa setiap model sangat bergantung pada karakteristik dataset, tingkat ketidakseimbangan kelas, dan kompleksitas hubungan antar fitur (Roy *et al.*, 2026). Meskipun demikian, sebagian besar studi komparatif dalam domain ini cenderung membandingkan model-model yang berasal dari paradigma arsitektur yang sama, sehingga nilai ilmiah dari perbandingan tersebut menjadi terbatas (Cogent Economics & Finance, 2025). Penelitian ini hadir dengan pendekatan yang berbeda: alih-alih membandingkan model dalam paradigma yang redundan, dipilih tiga model yang masing-masing merepresentasikan tiga paradigma pemodelan yang berbeda secara fundamental. Logistic Regression dipilih sebagai representasi paradigma statistik linear yang menjadi *baseline* industri kredit karena interpretabilitasnya yang tinggi dan kompatibilitasnya dengan regulasi perbankan (Mestiri, 2024). LightGBM dipilih sebagai representasi paradigma *gradient boosting* bukan XGBoost atau Random Forest yang berbagi paradigma serupa karena tiga alasan: LightGBM terbukti secara konsisten mengungguli XGBoost, AdaBoost, dan CatBoost dalam prediksi *personal default* pada berbagai dataset kredit (Cogent Economics & Finance, 2025); strategi pertumbuhan pohon *leaf-wise* yang dimilikinya menghasilkan efisiensi komputasi superior pada dataset berskala 150.000 sampel (Nature HSS, 2025); serta kemampuan penanganan *missing*

value secara *native* yang relevan untuk karakteristik dataset *Give Me Some Credit* (Zhu *et al.*, no date). FT-Transformer dipilih sebagai representasi paradigma *deep learning* berbasis mekanisme *self-attention* untuk data tabular, dengan pertimbangan bahwa dalam perbandingan langsung terhadap XGBoost dan CatBoost, FT-Transformer terbukti mengungguli keduanya pada 7 dari 11 kasus yang diuji (Gorishniy *et al.*, 2021), serta menunjukkan keunggulan dalam metrik *accuracy* dan AUC pada tugas prediksi risiko finansial (ACM, 2025) mengindikasikan bahwa arsitektur ini merupakan representasi *deep learning* yang paling kompetitif dan relevan untuk dievaluasi dalam konteks *credit scoring* saat ini. Dengan demikian, pemilihan ketiga model ini bukan berarti mengabaikan model lain, melainkan secara strategis memilih satu representan terbaik dari masing-masing paradigma yang paling kontras, sehingga menghasilkan perbandingan yang bermakna secara ilmiah dan tidak redundan.

Selain aspek prediktif, tuntutan akan transparansi dalam industri kredit mengharuskan model *credit scoring* memiliki tingkat interpretabilitas yang memadai agar dapat dijelaskan kepada regulator maupun pemangku kepentingan (Straus *et al.*, 2026a). Dalam konteks ini, pemahaman terhadap faktor-faktor yang memengaruhi prediksi model menjadi sama pentingnya dengan tingkat akurasi yang dihasilkan (Zheng *et al.*, 2026). Meskipun Logistic Regression, LightGBM, (Souadda *et al.*, 2025) dan FT-Transformer memiliki mekanisme pembelajaran yang berbeda (Choi *et al.*, 2026), belum terdapat konsensus mengenai apakah ketiganya mengandalkan pola prediktif yang serupa atau justru menghasilkan karakteristik pengambilan keputusan yang berbeda. Untuk menjembatani celah tersebut, penelitian ini memanfaatkan analisis SHAP guna mengidentifikasi kontribusi fitur terhadap prediksi risiko gagal

bayar, sehingga karakteristik pengambilan keputusan dari setiap model dapat dipahami secara mendalam (Straus et al., 2026b).

Perbandingan performa model machine learning pada data tabular masih menunjukkan hasil yang beragam. Logistic Regression, sebagai model linear yang paling banyak diterapkan dalam industri kredit, kerap dijadikan *baseline* karena interpretabilitasnya yang tinggi dan stabilitasnya pada data berdimensi rendah. Namun, model berbasis gradient boosting seperti LightGBM dilaporkan mampu melampaui Logistic Regression pada sebagian besar *benchmark* data tabular berkat efisiensi komputasi dan ketahanannya terhadap fitur yang tidak berkorelasi (Park et al., 2026). Di sisi lain, FT-Transformer menunjukkan performa yang kompetitif, bahkan melampaui model berbasis pohon pada dataset tertentu berkat kemampuannya menangkap interaksi kompleks antarfitur melalui mekanisme *self-attention* (Abbas, Ying and Iqbal, 2026). Hasil yang berbeda tersebut menunjukkan bahwa belum terdapat satu model yang secara konsisten unggul pada seluruh dataset, sehingga performa setiap model sangat dipengaruhi oleh karakteristik data dan konteks permasalahan (Roy et al., 2026) (Qu, Ao and Xu, 2026).

Perbandingan performa antara Logistic Regression, LightGBM, dan FT-Transformer dalam studi sebelumnya sering kali hanya dilakukan berdasarkan metrik evaluasi seperti Precision, F1-Score, atau Recall secara deskriptif (Study, 2026). Pendekatan tersebut memiliki keterbatasan karena belum mampu mengonfirmasi apakah perbedaan performa yang teramati merupakan hasil yang signifikan secara statistik atau sekadar variasi akibat pemilihan sampel data dan proses validasi. Ketika performa model yang dibandingkan relatif berdekatan, klaim keunggulan suatu model tanpa dukungan analisis inferensial yang ketat menjadi rentan

terhadap bias interpretasi dan kurang memiliki validasi yang kuat. Padahal, ketika performa model yang dibandingkan relatif berdekatan, perbedaan tersebut belum tentu menunjukkan keunggulan yang nyata tanpa didukung oleh analisis statistik yang memadai. Oleh karena itu, Penelitian ini mengevaluasi performa model menggunakan AUC-ROC pada setiap fold dalam 15-fold stratified cross-validation. Nilai AUC-ROC antar-model pada fold yang sama dibandingkan dan diberi peringkat sebagai pengamatan berpasangan.(Solution, 2026). Dengan demikian, pemilihan uji statistik tidak hanya didasarkan pada distribusi data, tetapi juga mempertimbangkan struktur pengamatan dan peringkat performa model. Friedman Test digunakan untuk membandingkan performa ketiga model, sedangkan Wilcoxon Signed-Rank Test digunakan untuk membandingkan setiap pasangan model.(Hengsomboon, Hengsomboon and Harutaichun, 2026).

Selain aspek prediktif, tuntutan akan transparansi dalam industri kredit mengharuskan model credit scoring memiliki tingkat interpretabilitas yang memadai agar dapat dijelaskan kepada regulator maupun pemangku kepentingan (Straus et al., 2026a). Dalam konteks ini, pemahaman terhadap faktor-faktor yang memengaruhi prediksi model menjadi sama pentingnya dengan tingkat akurasi yang dihasilkan (Zheng et al., 2026). Meskipun Logistic Regression, LightGBM, (Souadda et al., 2025) dan FT-Transformer memiliki mekanisme pembelajaran yang berbeda (Choi et al., 2026), belum terdapat konsensus mengenai apakah ketiganya mengandalkan pola prediktif yang serupa atau justru menghasilkan karakteristik pengambilan keputusan yang berbeda. Untuk menjembatani celah tersebut, penelitian ini memanfaatkan analisis SHAP guna mengidentifikasi kontribusi fitur terhadap prediksi risiko gagal

bayar, sehingga karakteristik pengambilan keputusan dari setiap model dapat dipahami secara mendalam (Straus *et al.*, 2026b).

B. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana hasil uji statistik non-parametrik terhadap kinerja model Logistic Regression, LightGBM, dan FT-Transformer dalam memprediksi risiko kredit berdasarkan metrik AUC-ROC?
2. Seberapa besar besaran efek (effect size) perbedaan performa antar ketiga model berdasarkan Cliff's Delta?
3. Bagaimana pola interpretabilitas ketiga model berdasarkan tiga dimensi analisis SHAP dan sejauh mana konsistensi atau perbedaan faktor prediktif yang diandalkan?

C. Batasan Masalah

Agar penelitian terfokus dan dapat dipertanggungjawabkan secara metodologi dilakukan pembatasan sebagai berikut:

1. Penelitian ini hanya membandingkan tiga algoritma yaitu Logistic Regression, LightGBM, dan FT-Transformer, algoritma lain seperti XGBoost, Random Forest, atau arsitektur deep learning tabular lainnya tidak diikutsertakan.
2. Dataset yang digunakan dibatasi pada dataset publik *Give Me Some Credit* dari Kaggle yang terdiri atas 150.000 sampel dan 10 variabel prediktor bertipe numerik. Pemilihan dataset ini mempertimbangkan

kesesuaian karakteristik data serta efisiensi komputasi selama proses pelatihan dan optimasi model. Oleh karena itu, hasil penelitian tidak dapat digeneralisasikan secara langsung pada dataset kredit dengan karakteristik yang berbeda.

3. Pengujian signifikansi perbedaan performa dilakukan menggunakan tiga tahap, yaitu Friedman Test sebagai uji global, Wilcoxon Signed-Rank Test dengan koreksi Holm sebagai uji *post-hoc* pasangan, dan Cliff's Delta sebagai ukuran *effect size*. Uji statistik parametrik tidak digunakan karena distribusi nilai AUC-ROC lintas *fold* tidak diasumsikan normal.
4. Evaluasi performa model hanya menggunakan satu metrik, yaitu *Area Under the Receiver Operating Characteristic Curve* (AUC-ROC). Metrik ini dipilih karena bersifat *threshold-independent* dan mampu mengukur kemampuan diskriminasi model secara menyeluruh tanpa bergantung pada penentuan ambang batas klasifikasi tertentu, sehingga relevan untuk tugas prediksi risiko kredit pada dataset *Give Me Some Credit*. Metrik lain seperti F1-Score, Precision, Recall, atau Brier Score tidak digunakan dalam penelitian ini.
5. Analisis SHAP dilakukan pada 300 sampel data uji sebagai aproksimasi, bukan perhitungan eksak pada seluruh data uji. Metode *explainer* disesuaikan per model: LinearExplainer untuk Logistic Regression, TreeExplainer untuk LightGBM, dan KernelExplainer untuk FT-Transformer. Interpretabilitas dibatasi pada tiga dimensi: *Global Feature Importance*, *Local Explanation*, dan *Directional Impact Analysis*.

D. Tujuan Penelitian

Berdasarkan rumusan masalah, penelitian ini bertujuan untuk:

1. Menguji apakah terdapat perbedaan kinerja yang signifikan secara statistik antara model Logistic Regression, LightGBM, dan FT-Transformer dalam memprediksi risiko kredit menggunakan uji non-parametrik Friedman Test dan Wilcoxon Signed-Rank Test, serta mengukur besaran efek perbedaannya menggunakan Cliff's Delta.
2. Mengevaluasi dan membandingkan kinerja prediksi ketiga model Logistic Regression, LightGBM, dan FT-Transformer berdasarkan metrik AUC-ROC pada skema 15-fold stratified cross-validation dan evaluasi akhir pada data uji, untuk mengidentifikasi model dengan performa diskriminasi terbaik secara deskriptif.
3. Menganalisis pola interpretabilitas ketiga model melalui tiga dimensi SHAP, Global Feature Importance, Local Explanation, dan Directional Impact Analysis untuk membandingkan mekanisme pengambilan keputusan masing-masing model dalam konteks prediksi risiko kredit.

E. Manfaat Penelitian

1. Manfaat Teoretis

Menganalisis pola interpretabilitas ketiga model melalui tiga dimensi SHAP. Global Feature Importance, Local Explanation, dan Directional Impact Analysis untuk membandingkan mekanisme pengambilan keputusan masing-masing model dalam konteks prediksi risiko kredit.

2. Manfaat Praktis

Bagi industri fintech dan P2P lending, hasil penelitian ini memberikan panduan berbasis bukti empiris dalam memilih model credit scoring yang paling sesuai dengan mempertimbangkan aspek akurasi prediksi (AUC-ROC), validitas statistik perbedaan performa, dan interpretabilitas secara bersamaan. Hasil analisis SHAP tiga dimensi juga memberikan masukan bagi regulator dalam memahami faktor-faktor prediktif yang digunakan model dan implikasinya terhadap transparansi keputusan kredit.



BAB V PENUTUP

A. Kesimpulan

Berdasarkan hasil penelitian evaluasi komparatif model *Logistic Regression*, *LightGBM*, dan *FT-Transformer* untuk prediksi risiko kredit pada *dataset Give Me Some Credit* menggunakan pengujian statistik non-parametrik dan interpretabilitas SHAP, dapat ditarik kesimpulan sebagai berikut.

1. Terdapat perbedaan performa yang signifikan secara statistik di antara ketiga model. Uji Friedman menghasilkan nilai $\chi^2 F = 30$ ($p < 0,01$) nilai maksimum teoretis yang dapat dicapai pada desain tiga model dengan 15 *fold*, yang menunjukkan bahwa urutan peringkat performa *LightGBM* , *FT-Transformer*, *Logistic Regression* berlaku secara sempurna dan konsisten di seluruh 15 *fold* tanpa satu pun pengecualian. Uji *post-hoc Wilcoxon Signed-Rank* dengan koreksi Holm mengonfirmasi bahwa seluruh tiga pasangan model berbeda secara signifikan ($p_{\text{Holm}} = 0,02$ untuk semua pasangan), dengan *W Statistic* = 0,0 mengindikasikan dominasi lengkap tanpa pembalikan hierarki pada satu pun *fold*. Pengujian statistik bertahap ini memberikan validasi inferensial yang diperlukan agar klaim mengenai keunggulan relatif antar model tidak sekadar bersifat deskriptif, tetapi dapat dipertahankan secara statistik.
2. Perbedaan performa tidak hanya signifikan secara statistik tetapi juga bermakna secara praktis. Ukuran efek *Cliff's Delta* menghasilkan nilai $|\delta| = 1,000$ untuk pasangan LR vs *LightGBM* dan LR vs *FT-Transformer*, serta $|\delta| = 0,920$ untuk pasangan *LightGBM* vs *FT-Transformer* semua tergolong kategori large

berdasarkan ambang batas Romano et al. ($|\delta| \geq 0,474$). Besaran efek ini mengindikasikan bahwa perbedaan performa antar model bukan sekadar artefak ukuran sampel yang besar, melainkan mencerminkan perbedaan kapasitas pemodelan yang fundamental. Secara deskriptif, LightGBM mencapai rerata AUC-ROC terbaik pada 15-fold CV ($0,8596 \pm 0,0097$) dan pada data uji ($0,8631$), diikuti FT-Transformer (CV: $0,8340 \pm 0,0075$; Uji: $0,8369$) dan Logistic Regression (CV: $0,7821 \pm 0,0100$; Uji: $0,7959$).

3. Ketiga model menunjukkan pola interpretabilitas yang berbeda secara substantif pada tiga dimensi analisis SHAP. Pada dimensi Global Feature Importance, ketiga model mengidentifikasi fitur-fitur keterlambatan pembayaran dan usia sebagai prediktor utama, namun dengan distribusi kepentingan yang berbeda LightGBM menunjukkan konsentrasi kepentingan pada fitur keterlambatan (*NumberOfTime30-59DaysPastDueNotWorse*: mean SHAP = $1,3371$), sementara FT-Transformer mendistribusikan kepentingan secara lebih merata melalui mekanisme *self-attention*. Pada dimensi *Local Explanation*, analisis waterfall mengungkap bahwa kasus False Negative pada LightGBM terjadi akibat dominasi kontribusi negatif fitur-fitur yang berlawanan dengan profil risiko umum pada sampel individual tertentu, memberikan pemahaman tentang mekanisme kegagalan model yang tidak dapat diperoleh dari metrik agregat. Pada dimensi Directional Impact Analysis, ketiga model menunjukkan konsistensi arah pengaruh yang selaras dengan teori manajemen risiko kredit: nilai fitur keterlambatan pembayaran yang tinggi secara konsisten meningkatkan risiko *default*, sementara usia dan

pendapatan yang tinggi menurunkan risiko, memberikan validasi konvergensi antara pembelajaran mesin dan pengetahuan domain.

4. Penelitian ini menegaskan urgensi evaluasi komparatif yang melampaui perbandingan deskriptif nilai metrik. Tanpa pengujian statistik inferensial, klaim mengenai keunggulan LightGBM atas FT-Transformer yang memiliki selisih AUC-ROC hanya $\approx 0,026$ akan rentan terhadap interpretasi yang tidak berdasar, karena perbedaan sekecil itu dapat saja merupakan variasi *sampling* belaka. Prosedur bertahap Friedman, Wilcoxon dengan koreksi Holm Cliff's Delta yang diterapkan dalam penelitian ini memberikan kerangka evaluasi yang lebih ketat dan dapat dipertanggung jawabkan secara ilmiah, sejalan dengan rekomendasi best practice dalam evaluasi perbandingan algoritma *machine learning* pada data kredit.

B. Saran

Berdasarkan temuan dan keterbatasan penelitian ini, beberapa saran diajukan untuk penelitian dan pengembangan selanjutnya.

1. Validasi pada Dataset Multi-Sumber (Multi-Dataset Benchmarking). Penelitian ini dibatasi pada satu dataset publik (Give Me Some Credit) sehingga temuan mengenai hierarki performa ketiga model tidak dapat secara langsung digeneralisasi ke konteks kredit lain. Penelitian selanjutnya disarankan mengadopsi pendekatan *multi-dataset benchmarking* dengan dataset dari berbagai platform P2P lending dan yurisdiksi berbeda, termasuk dataset yang mengandung data perilaku digital (*alternative data*) yang kini semakin umum digunakan dalam penilaian risiko kredit modern. Pendekatan ini akan meningkatkan

validitas eksternal kesimpulan penelitian dan memberikan pemahaman yang lebih lengkap tentang kondisi di mana masing-masing model unggul.

2. Eksplorasi Arsitektur Deep Learning Tabular Terkini. Penelitian ini menggunakan FT-Transformer dengan implementasi rtdl yang terbatas pada konfigurasi tertentu. Penelitian selanjutnya disarankan mengeksplorasi arsitektur tabular terkini seperti TabPFN, SAINT (Self-Attention and Intersample Attention Transformer), atau TabNet yang menggabungkan mekanisme *attention* dengan seleksi fitur eksplisit. Khususnya, TabPFN yang menggunakan *prior-fitted networks* menunjukkan potensi yang sangat menjanjikan pada dataset berukuran kecil hingga sedang dan dapat menjadi alternatif yang lebih efisien secara komputasi dibandingkan FT-Transformer konvensional.
3. Penerapan Teknik Kalibrasi Probabilitas Pasca-Pelatihan. Penelitian ini hanya menggunakan AUC-ROC sebagai metrik evaluasi, sehingga kualitas kalibrasi probabilitas yang dihasilkan model tidak dievaluasi secara eksplisit. Penelitian selanjutnya dapat mengeksplorasi teknik kalibrasi *post-hoc* seperti *Platt Scaling*, *Isotonic Regression*, atau *Temperature Scaling* untuk meningkatkan kualitas estimasi probabilitas tanpa mengorbankan kapasitas diskriminasi. Kalibrasi yang lebih baik sangat bermanfaat untuk penetapan suku bunga berbasis risiko (*risk-based pricing*) dan pengelolaan portofolio kredit di industri P2P lending.
4. Integrasi Evaluasi Keadilan Algoritma (*Algorithmic Fairness*). Penelitian ini berfokus pada aspek akurasi dan interpretabilitas, namun belum mempertimbangkan potensi bias algoritmik

terhadap kelompok demografis tertentu. Penelitian selanjutnya disarankan mengintegrasikan evaluasi *fairness* menggunakan metrik seperti *Demographic Parity*, *Equalized Odds*, atau *Calibration within Groups*, sejalan dengan prinsip AI yang bertanggung jawab dan tuntutan regulasi OJK melalui POJK Nomor 40 Tahun 2024 yang menekankan perlindungan konsumen dan non-diskriminasi dalam layanan keuangan berbasis teknologi.

5. Perluasan Ruang Pencarian Hiperparameter dan Jumlah Trial Optuna. Pembatasan pada 10 trial Optuna per model dalam penelitian ini merupakan kompromi antara kualitas optimasi dan efisiensi komputasi. Penelitian selanjutnya disarankan meningkatkan jumlah trial (misalnya 50–100 trial) dan memperluas ruang pencarian hiperparameter untuk FT-Transformer, khususnya parameter d_token (dimensi embedding), jumlah lapisan Transformer, dan konfigurasi *feedforward network*, yang dalam penelitian ini dibatasi pada konfigurasi minimal. Peningkatan ruang eksplorasi ini berpotensi meningkatkan performa FT-Transformer secara substansial dan memberikan perbandingan yang lebih adil antara model *deep learning* dan *gradient boosting*.
6. Pengembangan Kerangka Evaluasi Komprehensif yang Lebih Sistematis. Penelitian ini dapat dikembangkan dengan menambahkan dimensi evaluasi efisiensi inferensi (*inference latency* dan *throughput*) secara terukur, serta analisis stabilitas model terhadap pergeseran distribusi data (*concept drift*) yang umum terjadi pada data kredit seiring waktu. Kerangka evaluasi yang mencakup performa prediksi, kalibrasi probabilitas, interpretabilitas, *fairness*, dan efisiensi inferensi secara

terintegrasi akan memberikan panduan yang lebih komprehensif bagi praktisi industri keuangan dalam memilih model *credit scoring* yang tepat, melampaui fokus sempit pada nilai AUC-ROC yang masih mendominasi sebagian besar literatur komparatif saat ini.

Secara keseluruhan, penelitian ini berkontribusi pada literatur evaluasi komparatif model *machine learning* untuk prediksi risiko kredit dengan menunjukkan bahwa validasi statistik inferensial dan interpretabilitas multi-dimensional merupakan komponen yang tidak dapat diabaikan dalam proses evaluasi model. Temuan ini diharapkan dapat mendorong adopsi pendekatan evaluasi yang lebih ketat dan komprehensif, baik dalam penelitian akademis maupun dalam praktik industri keuangan dan *fintech* P2P lending di Indonesia dan konteks global.



DAFTAR PUSTAKA

- Abbas, G., Ying, Z. and Iqbal, M. (2026) 'From Binary Scores to Risk Tiers : An Interpretable Hybrid Stacking Model for Multi-Class Loan Default Prediction', pp. 1–24.
- Abubakar, H. (2026) 'Hybrid Multiobjective Framework for Boolean Satisfiability Optimization With Application to Financial Risk Prediction', pp. 1–28. Available at: <https://doi.org/10.1155/jama/9145091>.
- Airlangga, T., Sunitiyoso, Y. and Sudrajad, O.Y. (2026) 'Social Sciences & Humanities Open Who gets the credit ? Digitizing psychometric and character-based credit scoring in fintech microlending', *Social Sciences & Humanities Open*, 13(September 2025), p. 102715. Available at: <https://doi.org/10.1016/j.ssaho.2026.102715>.
- Ajibode, A., Shu, T. and Gulsher, L. (2022) 'Effectively Combining Risk Evaluation Metrics for Precise Fault Localization', *matematic*, 10(21), pp. 1–24. Available at: <https://doi.org/10.3390/math10213924>.
- Alalawi, K. (2026) 'Comparative Explainable AI for Breast Cancer Classification : Cross-Model SHAP Agreement Analysis Using XGBoost and Logistic Regression', pp. 1–22.
- Allbright, J. and Merz, C. (2025) *Temporal Boosting for Incremental Tree-based Learning on Tabular Data, 13th ACM IKDD International Conference on Data Science (CODS 2025), December 17–20, 2025, Pune, India*. Association for Computing Machinery. Available at: <https://doi.org/10.1145/3799830.3799877>.
- Alvi, J., Arif, I. and Nizam, K. (2024) 'Heliyon Advancing financial

- resilience : A systematic review of default prediction models and future directions in credit risk management', *Heliyon*, 10(21), p. e39770. Available at: <https://doi.org/10.1016/j.heliyon.2024.e39770>.
- Asal, B. and Yalciner, B. (2026) 'Benchmarking TabNet , NODE , and FT-Transformer for Software Defect Prediction: An Empirical Comparison and Explainability Analysis', *IEEE Access*, 14(January), pp. 11660–11681. Available at: <https://doi.org/10.1109/ACCESS.2026.3656247>.
- Banna, H. Al (2025) 'Preliminary research of P2P lending bankruptcy in emerging country: the role of financial literacy', (November). Available at: <https://doi.org/10.1108/JEBDE-12-2024-0055>.
- 'Basel Committee on Banking Supervision Principles for the management of credit risk' (2025), (April).
- Bougha, P. *et al.* (2026) 'Machine Learning-Based Dry Gas Reservoirs Z-Factor Prediction for Sustainable Energy Transitions to Net Zero', pp. 1–23.
- Calabrese, R., Osmetti, S.A. and Zanin, L. (2024) 'Sample selection bias in non-traditional lending: A copula-based approach for imbalanced data', *Socio-Economic Planning Sciences*, 95, p. 102045. Available at: <https://doi.org/https://doi.org/10.1016/j.seps.2024.102045>.
- Choi, G. *et al.* (2026) 'LFTD: Transformer-Enhanced Diffusion Model for Realistic Financial Time-Series Data Generation', pp. 1–23.
- Dasril, Y. *et al.* (2023) 'Credit Risk Assessment in P2P Lending Using LightGBM and Particle Swarm Optimization', 9(January), pp. 18–28.
- Deng, C. *et al.* (2025) 'Prediction of retail commodity hot-spots : a machine learning approach', 8(February), pp. 414–422. Available at:

<https://doi.org/10.1016/j.dsm.2025.02.003>.

- Dumitrescu, E. *et al.* (2022) ‘Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects’, *European Journal of Operational Research*, 297(3), pp. 1178–1192. Available at: <https://doi.org/https://doi.org/10.1016/j.ejor.2021.06.053>.
- Ehiro, T. (2026) ‘Leveraging SHAP values for superior prediction and efficient Bayesian optimization in material chemistry’, pp. 1–20.
- Fan, Y. and Waldmann, P. (2024) ‘Tabular deep learning : a comparative study applied to multi - task genome - wide prediction’, *BMC Bioinformatics*, pp. 1–20. Available at: <https://doi.org/10.1186/s12859-024-05940-1>.
- Friedrich, G., Winkel, B. and Reichenbach, F. (2024) *Improving credit risk assessment in P2P lending with explainable machine learning survival analysis*, *Digital Finance*. Springer International Publishing. Available at: <https://doi.org/10.1007/s42521-024-00114-3>.
- Gabrielli, G. (2026) *Corporate financial distress prediction : a machine learning*. Available at: <https://doi.org/10.1108/JAOC-05-2025-0166>.
- Gandhi, M., Chaudhari, A.Y. and Mulla, R. (2026) ‘A data-driven framework for estimating remaining driving range in cargo electric vehicles’.
- Ganie, S.M., Kanti, P. and Pramanik, D. (2023) ‘An ensemble learning approach for diabetes prediction using boosting techniques’, (October), pp. 1–15. Available at: <https://doi.org/10.3389/fgene.2023.1252159>.
- Garg, S. *et al.* (2026) ‘Multifaceted structural classification of Indian legal judgments’.
- Ghahremani, M. *et al.* (2026) ‘Machine Learning-Based Loan Approval Automation: Enhancing Efficiency, Accuracy and Fairness in Credit

- Decision-Making', *Expert Systems*, 43(7), p. e70307. Available at: <https://doi.org/https://doi.org/10.1111/exsy.70307>.
- Ghasemzadeh, H. (2024) 'Toward Generalizable Machine Learning Models in Speech , Language , and Hearing Sciences : Estimating Sample Size and Reducing Overfitting', 67(March).
- Gupta, A., Chug, A. and Singh, A.P. (2027) 'Feature extraction in sensor plant disease datasets using reformed membership functions independent of class variables', (2026), pp. 1–27.
- He, Z. and Moreira, C. (2026) 'TabAttackBench : A benchmark for adversarial attacks on tabular data', *Expert Systems With Applications*, 301(November 2025), p. 130491. Available at: <https://doi.org/10.1016/j.eswa.2025.130491>.
- Hengsomboon, P., Hengsomboon, N. and Harutaichun, P. (2026) 'Effects of forward shoulder elevation on spinal excursion at three different speeds', pp. 1–8.
- Hjelkrem, L.O. (2023) 'Explaining Deep Learning Models for Credit Scoring with SHAP : A Case Study Using Open Banking Data'.
- Jahan, J. (2026) 'Biomedical Signal Processing and Control Hybrid attention-based temporal convolutional networks for driver fatigue detection using physiological signal', *Biomedical Signal Processing and Control*, 113(PD), p. 109125. Available at: <https://doi.org/10.1016/j.bspc.2025.109125>.
- Jain, A. *et al.* (2026) 'stacking ensemble with Boruta- based feature selection for phishing URL detection'.
- Jallow, B. and Seong, J. (2026) 'FT-AFM: A Hybrid Transformer-Factorization Framework for Interpretable Click-Through Rate

- Prediction’, *IEEE Access*, PP, p. 1. Available at: <https://doi.org/10.1109/ACCESS.2026.3695521>.
- Ko, J. *et al.* (2026) ‘Interpretable Feature Selection and Hybrid Deep Learning Models for Depressive Symptoms Prediction from Wearable Device Data’, pp. 1–16.
- Koganti, H. and Anne, S. (2026) ‘Operational Excellence Through AI and ML in Financial Services : A Comprehensive Review of Applications , Challenges , and Future Directions’.
- Kpatcha, E. (2025) ‘Balancing Fairness and Accuracy in Machine Learning-Based Probability of Default Modeling via Threshold Optimization’.
- Kyeong, S. and Shin, J. (2022) ‘Two - stage credit scoring using Bayesian approach’, *Journal of Big Data* [Preprint]. Available at: <https://doi.org/10.1186/s40537-022-00665-5>.
- Li, L., Kumar, A. and Cheng, S. (2025) ‘Intelligent Systems with Applications Explainable AI based LightGBM prediction model to predict default borrower in social lending platform’, *Intelligent Systems with Applications*, 26(March), p. 200514. Available at: <https://doi.org/10.1016/j.iswa.2025.200514>.
- Liu, E. and Cho, K. (2026) ‘Deep learning-based labor relations prediction system with multi-source data fusion and early warning mechanisms’, pp. 1–25.
- Liu, Y. *et al.* (2024) ‘Leveraging network topology for credit risk assessment in P2P lending : A comparative study under the lens of machine learning’, *Expert Systems With Applications*, 252(PB), p. 124100. Available at: <https://doi.org/10.1016/j.eswa.2024.124100>.
- Loss, G.H.M. and Optimization, F. (2024) ‘XGBoost-B-GHM : An Ensemble

- Model with Feature Selection’, pp. 1–26.
- Malave, S. *et al.* (2026) ‘MethodsX Transforming financial documents into credit decisions using explainable artificial intelligence and optical character recognition’, *MethodsX*, 16(May), p. 103962. Available at: <https://doi.org/10.1016/j.mex.2026.103962>.
- Mathibela, M.R. and Maposa, D. (2026) ‘Predictive Modelling of Credit Default Risk Using Machine Learning and Ensemble Techniques’, pp. 1–20.
- Mathur, D., Kanwar, N. and Goyal, S.K. (2026) ‘Intelligent energy management of coordinated community microgrid systems using metaheuristic optimization and deep learning’, pp. 1–19.
- Mestiri, S. (2024) ‘Credit scoring using machine learning and deep Learning-Based models’, 4(December 2023), pp. 236–248.
- Mohamed, A. and Lee, K.H. (2026) ‘Gradient boosting with knockoff filters : a biostatistical approach to variable selection’.
- Monje, L. *et al.* (2025) ‘Machine Learning XAI for Early Loan Default Prediction’, *Computational Economics*, (June), pp. 4033–4062. Available at: <https://doi.org/10.1007/s10614-025-10962-9>.
- Ohue, M. and Furui, K. (2026) ‘PBPredictor . net : GBDT - based model and web tool for prediction of blood – placental barrier permeability of small molecules’, pp. 1–12.
- Park, K.H. *et al.* (2026) ‘Transformer-based feature extraction approach for hematopoietic cancer’, *Computers in Biology and Medicine*, 203(November 2025), p. 111466. Available at: <https://doi.org/10.1016/j.combiomed.2026.111466>.

- Qian, D. *et al.* (2026) ‘An enhanced connected banking system optimizer incorporating triple mechanism for solving global optimization problems’, pp. 1–19.
- Qu, M.I.N., Ao, C. and Xu, T. (2026) ‘Robust prediction of shopper intent via heterogeneous feature interaction and gradient calibration’, *IEEE Access*, PP, p. 1. Available at: <https://doi.org/10.1109/ACCESS.2026.3692241>.
- Rao, R.V.R. *et al.* (2026) ‘Unified approach for accurate heart disease prediction using machine learning techniques’.
- Rohini, G. *et al.* (2026) ‘OPEN A stacked Gradient Boosting – XGBoost ensemble with ridge meta-learner for accurate short- term solar PV power forecasting in smart grids’, pp. 1–32.
- Roy, J. *et al.* (2026) ‘Customer Churn Prediction on Structured Data Using FT-Transformer and Stacking Ensembles’, *IEEE Access*, PP, p. 1. Available at: <https://doi.org/10.1109/ACCESS.2026.3686374>.
- Sanz-guerrero, M. and Arroyo, J. (2025) ‘Credit Risk Meets Large Language Models : Building a Risk Indicator from Loan Descriptions in P2P Lending’, 28(75), pp. 220–247. Available at: <https://doi.org/10.4114/intartif.vol28iss75pp220-247>.
- Shi, X.I. and Tang, D. (2025) ‘Credit Scoring Prediction Using Deep Learning Models in the Financial Sector’, *IEEE Access*, 13(June), pp. 130731–130746. Available at: <https://doi.org/10.1109/ACCESS.2025.3591005>.
- Solution, L. (2026) ‘Automated Categorization of Bug Reports as Security Related or Non-security Related : A Machine’, 13(2), pp. 183–211. Available at: <https://doi.org/10.1142/S2196888825500022>.
- Soni, U. and Jethava, G. (2026) ‘Enhancing credit risk prediction via voting

- classifier and meta-dynamic ensemble selection on imbalanced data', 5.
- Souadda, L.I. *et al.* (2025) 'Optimizing Credit Risk Prediction for Peer-to-Peer Lending Using Machine Learning', pp. 1–31.
- Straus, J.V.D. *et al.* (2026a) 'Explaining to defend: SHAP-based defense mechanism against adversarial attacks in P2P credit risk', *Expert Systems With Applications*, 301(November 2025), p. 130432. Available at: <https://doi.org/10.1016/j.eswa.2025.130432>.
- Straus, J.V.D. *et al.* (2026b) 'Explaining to defend: SHAP-based defense mechanism against adversarial attacks in P2P credit risk', *Expert Systems With Applications*, 301(September 2025), p. 130432. Available at: <https://doi.org/10.1016/j.eswa.2025.130432>.
- Straus, J.V.D. *et al.* (2026c) 'Explaining to defend: SHAP-based defense mechanism against adversarial attacks in P2P credit risk', *Expert Systems With Applications*, 301(October 2025), p. 130432. Available at: <https://doi.org/10.1016/j.eswa.2025.130432>.
- Study, O. (2026) 'Enhancing Korean adolescent suicide risk prediction with TabR', 13(February).
- Teodorescu, V. and Obreja Braşoveanu, L. (2025) 'Assessing the Validity of k-Fold Cross-Validation for Model Selection: Evidence from Bankruptcy Prediction Using Random Forest and XGBoost', *Computation*, p. 127. Available at: <https://doi.org/10.3390/computation13050127>.
- Tran, V.Q., Choi, Y. and Byeon, H. (2024) 'Explainable stacking ensemble with feature tokenizer transformers for men ' s diabetes prediction', (MI). Available at: <https://doi.org/10.22514/jomh.2024.184>.
- Vakrani, D.S. *et al.* (2026) 'Evaluating AI-driven credit scoring models versus

- traditional statistical techniques', 1, pp. 1–14.
- Wang, H., Xu, Q. and Zhou, L. (2015) 'Lasso-Logistic Regression Ensemble', pp. 1–20. Available at: <https://doi.org/10.1371/journal.pone.0117844>.
- Wen, Y. *et al.* (2026) 'Machine learning model optimization with optuna for accurate prediction of strength and crack behavior in prestressed concrete beams', pp. 1–15.
- Xia, Y. (2019) 'A Novel Reject Inference Model Using Outlier Detection and Gradient Boosting Technique in Peer-to-Peer Lending', *IEEE Access*, 7, pp. 92893–92907. Available at: <https://doi.org/10.1109/ACCESS.2019.2927602>.
- Xing, J. *et al.* (2026) 'Interpretable Machine Learning for Predicting Splitting Strength of Asphalt Concrete : Insights from SHAP Analysis', pp. 1–31.
- Xu, M., Tian, B. and Fu, Y. (2022) 'Default prediction of online credit loans based on mobile application usage behaviors', *Journal of Intelligent & Fuzzy Systems*, 43(3), pp. 2253–2264. Available at: <https://doi.org/10.3233/JIFS-211825>.
- Xu, N. *et al.* (2026) 'High-Confidence Computing SparseLinTab : Sparse linear self-attention for efficient feature interaction in tabular data', *High-Confidence Computing*, 6(2), p. 100379. Available at: <https://doi.org/10.1016/j.hcc.2025.100379>.
- Ying, C., Shi, A. and Li, X. (2025) 'framework for enhanced credit risk assessment in', pp. 1–13.
- Zheng, C. *et al.* (2026) 'Credit scoring modeling based on an improved two-stage method with fragmentary data'.

Zhu, M. *et al.* (no date) 'Ensemble Methodology : Innovations in Credit Default Prediction Using LightGBM , XGBoost , and LocalEnsemble'.

