

Model Analitik Risiko Nasabah Online Lending Berbasis Big Data

Iwan Purwanto¹, Syandra Sari^{2*}, Muhammad Umar Arif Paago³, Muhammad Tegar Hidayatullah⁴, Muhammad Fikri⁵

¹²³⁴⁵ Prodi Sistem Informasi, Fakultas Teknologi Industri, Universitas Trisakti, Jakarta, Indonesia

Iwan.purwanto@trisakti.ac.id, syandra.sari@trisakti.ac.id*,

065002100031@std.trisakti.ac.id, 065002100022@std.trisakti.ac.id, m.fikri@trisakti.ac.id

ABSTRACT – The rapid advancement of financial technology (fintech), particularly in online lending platforms, has significantly enhanced financial inclusion while simultaneously introducing challenges in credit risk assessment due to limited historical data availability. This study aims to develop an analytical model based on Big Data Analytics to improve the accuracy of credit risk prediction through the integration of multi-source data.

This research adopts a data-driven quantitative approach consisting of five main stages: data collection, data preprocessing, data integration, model development, and model evaluation. The dataset integrates borrower transaction data, user profile information, and digital behavioral data. Four machine learning algorithms—Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM)—are implemented and comparatively evaluated using performance metrics including accuracy, precision, recall, and F1-score.

The results indicate that the Random Forest algorithm achieves the best performance, with an accuracy of 0.91, precision of 0.90, recall of 0.89, and F1-score of 0.89, outperforming Support Vector Machine (0.87), Decision Tree (0.85), and Logistic Regression (0.82). Feature importance analysis identifies previous payment status, income level, and loan history as the most significant predictors of credit risk.

This study contributes by proposing an integrated big data-driven analytical model that incorporates digital behavioral data as an alternative information source, thereby enhancing the accuracy and reliability of credit risk assessment. The proposed model also has practical implications as a decision support system, enabling faster, more objective, and data-driven decision-making processes in fintech lending platforms.

Keywords: Big Data Analytics, Credit Risk Prediction, Online Lending, Machine Learning, Digital Behavioral Data

ABSTRAK – Perkembangan fintech, khususnya online lending, meningkatkan inklusi keuangan namun menghadirkan tantangan dalam penilaian risiko kredit akibat keterbatasan data historis. Penelitian ini bertujuan mengembangkan model analitik berbasis *Big Data Analytics* yang mampu meningkatkan akurasi prediksi risiko kredit melalui integrasi multi-sumber data.

Penelitian ini menggunakan pendekatan kuantitatif berbasis data melalui lima tahapan: pengumpulan data, preprocessing, integrasi data, pengembangan model, dan evaluasi. Dataset menggabungkan data transaksi, profil pengguna, dan perilaku digital. Empat algoritma *machine learning* - *Logistic Regression*, *Decision Tree*, *Random Forest*, dan *Support Vector Machine* (SVM) - digunakan dan dibandingkan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

Hasil menunjukkan bahwa *Random Forest* memberikan performa terbaik dengan *accuracy* 0,91, *precision* 0,90, *recall* 0,89, dan *F1-score* 0,89, melampaui SVM (0,87), *Decision Tree* (0,85), dan *Logistic Regression* (0,82). Analisis *feature importance* mengidentifikasi status pembayaran sebelumnya, tingkat pendapatan, dan riwayat pinjaman sebagai prediktor utama risiko kredit.

Penelitian ini menawarkan kontribusi berupa model analitik terintegrasi berbasis big data yang menggabungkan data perilaku digital sebagai sumber informasi alternatif, sehingga mampu meningkatkan akurasi dan reliabilitas penilaian risiko kredit. Model ini juga memiliki implikasi praktis sebagai decision support system untuk mendukung pengambilan keputusan yang lebih cepat, objektif, dan berbasis data pada platform *fintech lending*.

Kata Kunci: Analisis Big Data, Prediksi Risiko Kredit, Pinjaman Online, Machine Learning, Data Perilaku Digital.



1. PENDAHULUAN

Perkembangan teknologi digital telah mendorong transformasi signifikan dalam sektor keuangan global, khususnya melalui kemunculan layanan financial technology (fintech) [1][2]. Salah satu inovasi yang berkembang pesat adalah online (P2P) lending, yaitu platform digital yang mempertemukan pemberi pinjaman dan peminjam secara langsung tanpa perantara lembaga keuangan tradisional [3][4][5]. Model bisnis ini menawarkan kemudahan akses pembiayaan, proses yang lebih cepat, serta jangkauan layanan yang lebih luas dibandingkan sistem perbankan konvensional. Di Indonesia, pertumbuhan industri fintech lending meningkat secara signifikan dalam beberapa tahun terakhir seiring meningkatnya penetrasi internet dan penggunaan layanan keuangan digital oleh masyarakat [6].

Meskipun memberikan berbagai manfaat, perkembangan pesat platform P2P lending juga menghadirkan berbagai tantangan, terutama terkait dengan penilaian risiko kredit calon nasabah [7]. Berbeda dengan lembaga keuangan tradisional yang memiliki data historis yang lengkap, banyak platform fintech menghadapi keterbatasan informasi dalam melakukan evaluasi kelayakan kredit peminjam. Hal ini dapat meningkatkan potensi risiko gagal bayar (default risk) serta berdampak pada stabilitas ekosistem fintech secara keseluruhan [8][9]. Oleh karena itu, diperlukan pendekatan analitik yang lebih komprehensif dan berbasis data untuk meningkatkan akurasi proses penilaian risiko dalam sistem lending digital [10][11][12].

Dalam beberapa tahun terakhir, Big Data Analytics telah menjadi pendekatan yang semakin banyak digunakan untuk mendukung pengambilan keputusan dalam berbagai sektor, termasuk sektor keuangan digital [13]. *Big data* memungkinkan pengolahan data dalam volume besar, kecepatan tinggi, serta keragaman sumber data yang beragam, seperti data transaksi, perilaku pengguna, data media sosial, maupun data digital lainnya [14][15][16]. Dengan memanfaatkan teknik analitik lanjutan, organisasi dapat mengidentifikasi pola tersembunyi, memprediksi perilaku pengguna, serta meningkatkan kualitas proses pengambilan keputusan berbasis data [13][17].

Sejumlah penelitian sebelumnya telah mengkaji penggunaan metode *machine learning* dan *data analytics* untuk memprediksi risiko kredit dalam layanan fintech [18][19]. Namun demikian, sebagian besar penelitian masih berfokus pada penggunaan dataset terbatas atau hanya memanfaatkan satu sumber data tertentu. Selain itu, integrasi berbagai sumber data dalam kerangka *big data analytics* untuk mendukung proses penilaian risiko pada platform P2P lending masih relatif terbatas, khususnya dalam konteks negara berkembang seperti Indonesia [18][20][21]. Kondisi ini menunjukkan adanya kesenjangan penelitian (*research gap*) yang perlu dikaji lebih lanjut [22].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan model analitik penilaian risiko nasabah pada platform *Online Lending* dengan memanfaatkan pendekatan *Big Data Analytics*. Model yang diusulkan dirancang untuk mengintegrasikan berbagai sumber data yang relevan guna menghasilkan analisis risiko yang lebih akurat dan komprehensif. Dengan memanfaatkan teknik analitik data yang tepat, sistem yang dikembangkan diharapkan dapat membantu platform fintech dalam meningkatkan kualitas proses evaluasi kredit calon nasabah.

Data perilaku digital merepresentasikan jejak aktivitas pengguna yang terekam melalui interaksi dengan perangkat, aplikasi, dan platform daring, yang mencakup frekuensi serta durasi penggunaan aplikasi, pola transaksi digital, karakteristik perangkat dan konektivitas, konsistensi waktu aktivitas, hingga indikator kepercayaan seperti stabilitas akun dan verifikasi identitas [23][24]. Dalam konteks Indonesia, pemanfaatan data ini menjadi sangat krusial mengingat masih terbatasnya riwayat kredit formal masyarakat yang tercatat dalam Sistem Layanan Informasi Keuangan, sehingga pendekatan konvensional dalam penilaian risiko kredit sering kali tidak mampu merepresentasikan profil risiko secara komprehensif [25]. Di sisi lain, tingginya penetrasi teknologi digital, masifnya penggunaan aplikasi keuangan dan e-commerce, serta dominasi sektor informal dengan pola pendapatan yang dinamis menjadikan data perilaku digital sebagai sumber informasi alternatif yang lebih real-time, kontekstual, dan adaptif [23][26]. Oleh karena itu, integrasi data perilaku digital dalam model analitik berbasis kecerdasan buatan tidak hanya mampu meningkatkan akurasi prediksi risiko kredit, tetapi juga berkontribusi terhadap perluasan inklusi keuangan secara berkelanjutan di Indonesia.

Pendekatan penilaian risiko kredit tradisional berbasis prinsip 5C (*Character, Capacity, Capital, Collateral, dan Condition*) masih banyak digunakan dalam sistem keuangan konvensional, namun memiliki keterbatasan dalam konteks fintech karena bergantung pada data historis yang terbatas dan proses evaluasi yang cenderung statis serta subjektif. Sebaliknya, perkembangan Big Data Analytics memungkinkan transformasi menuju pendekatan data-driven yang lebih komprehensif melalui integrasi berbagai sumber data, seperti data transaksi, profil pengguna, dan perilaku digital. Pendekatan ini memberikan kemampuan analisis yang lebih objektif, adaptif, dan prediktif dalam mengidentifikasi risiko kredit, serta lebih efektif dalam menjangkau nasabah tanpa riwayat kredit formal (*unbanked*), sehingga menjadi solusi yang relevan untuk meningkatkan akurasi dan inklusivitas dalam sistem penilaian risiko pada platform fintech lending [27][28].

Kontribusi utama penelitian ini adalah: (1) mengusulkan kerangka model analitik berbasis big data untuk penilaian risiko nasabah pada platform P2P lending, (2) mengidentifikasi sumber data yang relevan untuk meningkatkan akurasi prediksi risiko kredit, serta (3) memberikan implikasi praktis bagi pengembangan sistem informasi berbasis data dalam ekosistem fintech

lending. Hasil penelitian ini diharapkan dapat memberikan kontribusi akademik dalam pengembangan literatur mengenai analitik data pada sektor fintech, sekaligus memberikan manfaat praktis bagi penyedia layanan fintech dalam meningkatkan kualitas pengambilan keputusan berbasis data.

2. DASAR TEORI

Perkembangan *financial technology* (fintech), khususnya pada layanan *online lending*, telah mengubah paradigma sistem keuangan tradisional menjadi lebih digital, cepat, dan inklusif [29][30]. Platform ini memungkinkan interaksi langsung antara pemberi pinjaman dan peminjam tanpa melalui lembaga keuangan konvensional. Namun, peningkatan akses pembiayaan tersebut juga diiringi dengan meningkatnya risiko kredit akibat keterbatasan informasi historis nasabah [31][32]. Oleh karena itu, diperlukan pendekatan analitik yang mampu mengidentifikasi tingkat risiko secara lebih akurat dan komprehensif dengan memanfaatkan data yang tersedia dalam ekosistem digital [33][34].

Dalam konteks tersebut, *Big Data Analytics* menjadi pendekatan yang relevan untuk mengelola data dalam jumlah besar, beragam, dan berkecepatan tinggi. Karakteristik *big data* yang meliputi volume, velocity, dan variety memungkinkan integrasi berbagai sumber data seperti data transaksi, profil pengguna, serta perilaku digital [13][35]. Dengan memanfaatkan teknik analitik data, organisasi dapat mengidentifikasi pola tersembunyi, memahami perilaku pengguna, serta meningkatkan kualitas pengambilan keputusan berbasis data, khususnya dalam penilaian risiko kredit pada platform fintech lending [36].

Selain itu, penggunaan *machine learning* dalam penilaian risiko kredit memungkinkan sistem untuk melakukan klasifikasi dan prediksi berdasarkan pola data historis. Algoritma seperti Logistic Regression, Decision Tree, Random Forest, dan Support Vector Machine memiliki kemampuan dalam mengolah data kompleks dan menghasilkan prediksi yang akurat [7][37]. Pendekatan ini mendukung pengembangan sistem pendukung keputusan (*decision support system*) yang mampu membantu perusahaan fintech dalam melakukan evaluasi kredit secara lebih objektif, cepat, dan berbasis data, sehingga dapat meminimalkan risiko gagal bayar [38][39][40].

3. METODOLOGI

A. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analitik berbasis data untuk mengembangkan model penilaian risiko nasabah pada platform *online lending*. Pendekatan kuantitatif dipilih karena penelitian ini bertujuan untuk menganalisis hubungan antar variabel serta membangun model prediksi yang mampu mengklasifikasikan tingkat risiko calon peminjam secara objektif. Dalam konteks layanan

fintech lending, kemampuan untuk mengidentifikasi risiko kredit secara akurat menjadi faktor penting dalam menjaga keberlanjutan operasional platform serta meminimalkan potensi gagal bayar [41][29].



Gambar 1. Kerangka metodologi penelitian P2P lending

Penelitian ini mengadopsi pendekatan data-driven analytics, yaitu pendekatan yang menekankan pemanfaatan data dalam jumlah besar untuk menghasilkan wawasan yang relevan dalam proses pengambilan keputusan [10][42]. Dengan memanfaatkan teknik Big Data Analytics, penelitian ini bertujuan untuk mengolah data yang memiliki karakteristik volume, velocity, dan variety yang tinggi sehingga mampu menghasilkan analisis yang lebih komprehensif dibandingkan metode analisis konvensional [43].

Secara umum, desain penelitian ini terdiri dari beberapa tahapan utama yang saling terintegrasi, yaitu pengumpulan data, pemrosesan data, integrasi data, pengembangan model analitik, serta evaluasi model. Setiap tahapan dirancang secara sistematis untuk memastikan bahwa model analitik yang dihasilkan memiliki tingkat akurasi yang tinggi dalam memprediksi risiko kredit calon nasabah [44][45].

B. Sumber dan Jenis Data

Data yang digunakan dalam penelitian ini berasal dari berbagai sumber yang relevan dengan aktivitas peminjaman pada platform *online lending*. Penggunaan berbagai sumber data ini bertujuan untuk memperoleh gambaran yang lebih komprehensif mengenai karakteristik dan perilaku calon nasabah. Dalam konteks big data, keberagaman sumber data menjadi faktor penting yang dapat meningkatkan kualitas analisis dan akurasi model prediksi. Pengumpulan data diambil dari 36 platform pinjaman online yang tersebar di Indonesia. Pengumpulan data dilakukan dalam 1 termin (maret-agustus 2025) terdapat 6.925 nasabah yang melakukan transaksi sesuai dengan kondisi pemetikan data. Siber difokuskan pada platform legal dan nasabah berorientasi pada pelaku UMKM (pemanfaatan pengembangan

usaha).

Peneliti menganggap area ini menjadi penting untuk diteliti karena area berfokus pada pengembangan usaha dan pola konsumen pengembangan ekonomi kemasyarakatan. Pada tahun 2024. Dalam konteks kesehatan lembaga keuangan (terutama perbankan), indikator yang digunakan adalah Non-Performing Loan (NPL) atau kredit macet berada pada posisi $\leq 5\%$ dan OJK menetapkan untuk penetapan Kesehatan nilai berada pada $\leq 3.5\%$. Sementara untuk pelaku pinjaman online yang bersifat legal dan diperuntukan bagi pelaku UMKM, OJK mencatat pada tahun 2023 sebesar 6,7%, dan pada tahun 2024 sebesar 6,2%. Hal ini tentunya menjadi pekerjaan rumah yang tidak mudah untuk dapat ditanggulangi.

Jenis data yang digunakan dalam penelitian ini meliputi data transaksi peminjam, data profil pengguna, data perilaku digital, serta data eksternal yang dapat mendukung proses analisis risiko kredit. Data transaksi peminjam mencakup informasi mengenai riwayat pinjaman, jumlah pinjaman yang diajukan, durasi pinjaman, serta status pembayaran pinjaman sebelumnya. Informasi ini sangat penting dalam mengidentifikasi pola perilaku finansial peminjam.

Selain itu, penelitian ini juga memanfaatkan data profil pengguna yang mencakup karakteristik demografis seperti usia, pekerjaan, tingkat pendapatan, serta lokasi geografis. Variabel-variabel tersebut dapat memberikan indikasi awal mengenai kemampuan finansial calon nasabah. Data perilaku digital juga digunakan untuk menganalisis pola penggunaan layanan keuangan digital oleh pengguna, seperti frekuensi penggunaan aplikasi, aktivitas transaksi digital, serta interaksi pengguna dengan platform fintech.

C. Tahapan Pengolahan Data

C.1. Data Collection

Tahap pengumpulan data merupakan langkah awal dalam proses penelitian yang bertujuan untuk memperoleh dataset yang relevan dengan tujuan penelitian. Data dikumpulkan dari berbagai sumber yang berkaitan dengan aktivitas peminjaman digital pada platform *online lending*. Proses pengumpulan data ini dilakukan secara sistematis untuk memastikan bahwa data yang diperoleh memiliki kualitas yang baik dan dapat digunakan dalam proses analisis selanjutnya.

Data yang dikumpulkan dapat berupa data terstruktur maupun semi-terstruktur. Data terstruktur biasanya berasal dari basis data transaksi yang dimiliki oleh platform fintech lending, sedangkan data semi-terstruktur dapat berasal dari log aktivitas pengguna atau sumber data digital lainnya. Setelah data berhasil dikumpulkan, dataset tersebut kemudian disimpan dalam sistem penyimpanan data yang memungkinkan proses pengolahan data secara efisien.

C.2. Data Preprocessing

Tahap data preprocessing dilakukan untuk meningkatkan kualitas dataset sebelum digunakan dalam proses analisis. Data yang diperoleh dari berbagai sumber

seringkali memiliki permasalahan seperti data yang tidak lengkap, data duplikat, atau data yang memiliki format yang berbeda [46]. Oleh karena itu, diperlukan proses pembersihan data untuk memastikan bahwa dataset yang digunakan memiliki kualitas yang baik.



Gambar 2. Tahapan Data Processing dalam Pengembangan Model Analitik Risiko Kredit

Proses preprocessing meliputi beberapa kegiatan utama seperti data cleaning, data transformation, dan data normalization. Data cleaning dilakukan untuk menghapus atau memperbaiki data yang tidak valid atau tidak lengkap. Data transformation bertujuan untuk mengubah format data agar sesuai dengan kebutuhan analisis, sedangkan data normalization dilakukan untuk menyamakan skala data sehingga memudahkan proses analisis menggunakan algoritma machine learning.

Selain itu, tahap preprocessing juga mencakup proses feature selection, yaitu proses pemilihan variabel yang paling relevan dalam proses prediksi risiko kredit. Pemilihan fitur yang tepat dapat meningkatkan performa model analitik serta mengurangi kompleksitas proses perhitungan.

C.3. Data Integration

Tahap integrasi data dilakukan untuk menggabungkan berbagai dataset yang berasal dari sumber yang berbeda menjadi satu dataset yang terpadu. Integrasi data menjadi langkah penting dalam pendekatan big data analytics karena memungkinkan sistem untuk menganalisis hubungan antar variabel yang berasal dari berbagai sumber informasi [47][48][49].

Melalui proses integrasi data, penelitian ini dapat mengkombinasikan data transaksi peminjam, data profil pengguna, serta data perilaku digital ke dalam satu kerangka analisis yang komprehensif. Dengan demikian, model analitik yang dihasilkan dapat memanfaatkan berbagai variabel yang relevan dalam memprediksi tingkat risiko kredit calon nasabah.

C.4. Pengembangan Model Analitik

Tahap pengembangan model analitik dilakukan untuk membangun model prediksi yang dapat mengklasifikasikan calon nasabah berdasarkan tingkat risiko kreditnya. Dalam penelitian ini, pendekatan machine learning classification digunakan untuk mengidentifikasi pola hubungan antara karakteristik nasabah dan kemungkinan terjadinya gagal bayar [44].

Beberapa algoritma machine learning yang dapat digunakan dalam penelitian ini antara lain *Logistic*

Regression, Decision Tree, Random Forest, dan Support Vector Machine. Algoritma tersebut dipilih karena memiliki kemampuan yang baik dalam menangani permasalahan klasifikasi serta telah banyak digunakan dalam penelitian terkait prediksi risiko kredit.

Proses pengembangan model dilakukan dengan melatih algoritma menggunakan dataset historis yang telah diproses sebelumnya. Model kemudian mempelajari pola hubungan antara variabel input dan variabel target sehingga mampu menghasilkan prediksi risiko kredit pada data baru. Dengan memanfaatkan pendekatan analitik ini, platform fintech lending dapat memperoleh sistem penilaian risiko yang lebih akurat dan berbasis data.

C.5. Evaluasi Model

Evaluasi model dilakukan untuk mengukur performa model analitik dalam memprediksi risiko kredit nasabah. Proses evaluasi ini bertujuan untuk memastikan bahwa model yang dikembangkan memiliki tingkat akurasi yang memadai serta mampu memberikan prediksi yang reliabel.

1. Formula Perbandingan Performa Model

a. Accuracy

Perbandingan performa model dilakukan menggunakan beberapa metrik evaluasi klasifikasi seperti Accuracy, Precision, Recall, dan F1-Score.

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

Dimana:

- TP (*True Positive*); Nasabah berisiko tinggi yang diprediksi benar
- TN (*True Negative*); Nasabah berisiko rendah yang diprediksi benar
- FP (*False Positive*); Nasabah berisiko rendah tetapi diprediksi tinggi
- FN (*False Negative*); Nasabah berisiko tinggi tetapi diprediksi rendah

Accuracy menunjukkan proporsi prediksi yang benar terhadap seluruh data.

b. Precision

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

Dimana:

- TP (*True Positive*); Nasabah berisiko tinggi yang diprediksi benar
- FP (*False Positive*); Nasabah berisiko rendah tetapi diprediksi tinggi

Precision menunjukkan ketepatan model dalam memprediksi nasabah berisiko tinggi.

Semakin tinggi precision berarti semakin sedikit *false alarm* dalam prediksi risiko kredit.

a. Recall

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

Dimana:

- TP (*True Positive*); Nasabah berisiko tinggi yang diprediksi benar
- FN (*False Negative*); Nasabah berisiko tinggi tetapi diprediksi rendah

Recall menunjukkan kemampuan model dalam mendeteksi seluruh nasabah yang benar-benar memiliki risiko kredit tinggi.

Recall sangat penting dalam kasus *credit risk prediction* karena kesalahan mendeteksi risiko dapat menyebabkan kerugian finansial.

a. F1-Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

F1-score adalah rata-rata harmonik antara *precision* dan *recall* yang digunakan ketika dataset memiliki ketidakseimbangan kelas (*imbalanced data*).

2. Formula Confusion Matrix Random Forest

Confusion Matrix digunakan untuk mengevaluasi kinerja model Random Forest dalam klasifikasi risiko kredit.

Tabel 1. Struktur Confusion Matrix

Keterangan	Prediksi Positif	Prediksi Negatif
Aktual Positif	TP	FN
Aktual Negatif	FP	TN

Nilai-nilai ini digunakan untuk menghitung metrik evaluasi model.

a. Error Rate

$$Error Rate = \frac{FP+FN}{TP+TN+FP+FN} \quad (5)$$

Error Rate menunjukkan persentase kesalahan klasifikasi model. Semakin kecil nilai error rate maka semakin baik performa model Random Forest.

3. Formula Variabel Paling Berpengaruh (*Feature Importance*)

Untuk menentukan variabel paling berpengaruh dalam Random Forest, digunakan nilai *Feature Importance* yang dihitung dari penurunan impurity pada setiap pohon keputusan.

$$FI_j = \frac{\sum_{t \in T_j} \Delta I_t}{\sum_{k=1}^p \sum_{t \in T_k} \Delta I_t} \quad (6)$$

Dimana:

- FI_j ; Feature importance variabel ke-j
- T_j ; Node yang menggunakan variabel ke-j
- ΔI_t ; Penurunan impurity pada node t



- P , Jumlah total variabel

Nilai *feature importance* menunjukkan seberapa besar kontribusi suatu variabel terhadap model prediksi.

Beberapa metrik evaluasi yang digunakan dalam penelitian ini meliputi accuracy, precision, recall, dan F1-score. Accuracy digunakan untuk mengukur tingkat ketepatan model dalam melakukan klasifikasi, sedangkan precision dan recall digunakan untuk mengevaluasi kemampuan model dalam mengidentifikasi kasus risiko kredit secara tepat. *F1-score* digunakan sebagai indikator keseimbangan antara precision dan recall.

Selain itu, penelitian ini juga menggunakan teknik cross-validation untuk menguji kemampuan generalisasi model terhadap dataset yang berbeda. Dengan menggunakan teknik ini, model dapat diuji secara lebih objektif sehingga hasil evaluasi yang diperoleh dapat mencerminkan performa model yang sebenarnya.

C.6. Kerangka Penelitian

Kerangka penelitian dalam studi ini dirancang untuk menggambarkan alur proses penelitian secara sistematis mulai dari tahap pengumpulan data hingga evaluasi model analitik. Kerangka penelitian ini terdiri dari beberapa tahapan utama yang saling terintegrasi, yaitu pengumpulan data dari berbagai sumber, preprocessing data, integrasi data, pembangunan model analitik, serta evaluasi performa model.

Pendekatan ini memungkinkan penelitian untuk memanfaatkan potensi big data dalam mengidentifikasi pola perilaku peminjam serta memprediksi tingkat risiko kredit secara lebih akurat. Dengan adanya model analitik yang dikembangkan dalam penelitian ini, platform Online Lending diharapkan dapat meningkatkan kualitas proses penilaian risiko nasabah sehingga dapat meminimalkan potensi gagal bayar serta meningkatkan keberlanjutan ekosistem fintech lending.

Tabel 2 Tabel Variabel Penelitian

No	Variabel	Jenis Variabel	Indikator	Deskripsi	Sumber Data
1	Riwayat Pinjaman	Variabel Independen	Jumlah pinjaman sebelumnya	Menggambarkan jumlah pinjaman yang pernah dilakukan oleh nasabah	Data transaksi platform
2	Jumlah Pinjaman	Variabel Independen	Nilai pinjaman yang diajukan	Besaran dana yang diajukan oleh peminjam pada platform lending	Data transaksi platform
3	Durasi Pinjaman	Variabel Independen	Lama waktu pinjaman	Periode waktu yang dibutuhkan peminjam untuk melunasi pinjaman	Data transaksi platform
4	Status Pembayaran	Variabel Independen	Lancar / terlambat / gagal bayar	Riwayat pembayaran pinjaman yang dilakukan oleh peminjam	Data transaksi platform
5	Usia Nasabah	Variabel Independen	Rentang usia	Usia peminjam yang terdaftar pada platform fintech	Data profil pengguna
6	Pekerjaan	Variabel Independen	Jenis pekerjaan	Jenis profesi atau pekerjaan yang dimiliki oleh peminjam	Data profil pengguna
7	Pendapatan	Variabel Independen	Tingkat penghasilan	Pendapatan bulanan yang dimiliki oleh peminjam	Data profil pengguna
8	Lokasi Geografis	Variabel Independen	Wilayah domisili	Lokasi tempat tinggal peminjam yang tercatat pada sistem	Data profil pengguna
9	Aktivitas Digital	Variabel Independen	Frekuensi penggunaan aplikasi	Intensitas penggunaan layanan fintech oleh pengguna	Data log aktivitas sistem
10	Interaksi Platform	Variabel Independen	Aktivitas transaksi digital	Interaksi pengguna dengan layanan fintech lending	Data log sistem
11	Risiko Kredit	Variabel Dependen	Risiko rendah / sedang / tinggi	Tingkat kemungkinan gagal bayar oleh peminjam	Hasil analisis model

4. HASIL DAN PEMBAHASAN

4.1. Hasil Pengolahan Data

Tahap awal dalam penelitian ini adalah melakukan pengolahan data yang diperoleh dari berbagai sumber yang berkaitan dengan aktivitas peminjaman pada platform *online lending*. Dataset yang digunakan dalam penelitian ini mencakup data transaksi peminjam, data profil pengguna, serta data perilaku digital yang dikumpulkan dari sistem platform fintech. Proses pengolahan data dilakukan melalui beberapa tahapan yaitu data collection, data preprocessing, dan data integration.

Pada tahap *data preprocessing*, dilakukan proses pembersihan data untuk menghapus data yang tidak lengkap, data duplikat, serta data yang tidak relevan dengan tujuan penelitian. Selain itu, dilakukan proses transformasi data untuk menyesuaikan format data agar dapat digunakan dalam proses analisis machine learning. Proses normalisasi juga diterapkan untuk memastikan bahwa setiap variabel memiliki skala yang seragam sehingga dapat meningkatkan kinerja algoritma yang digunakan dalam penelitian.

Setelah proses preprocessing selesai dilakukan, dataset kemudian diintegrasikan menjadi satu dataset terpadu yang dapat digunakan dalam proses analisis. Integrasi data ini memungkinkan penelitian untuk mengkombinasikan berbagai variabel yang berasal dari sumber data yang berbeda sehingga dapat menghasilkan analisis yang lebih komprehensif mengenai karakteristik calon nasabah pada platform fintech lending.

4.2. Hasil Pengembangan Model Analitik

Pada tahap pengembangan model analitik, penelitian ini menggunakan beberapa algoritma *machine learning* untuk melakukan klasifikasi risiko kredit nasabah. Algoritma yang digunakan dalam penelitian ini meliputi *Logistic Regression*, *Decision Tree*, *Random Forest*, dan *Support Vector Machine (SVM)*. Keempat algoritma tersebut dipilih karena memiliki kemampuan yang baik dalam menangani permasalahan klasifikasi serta telah banyak digunakan dalam penelitian terkait prediksi risiko kredit.

Model machine learning dilatih menggunakan dataset historis yang telah melalui proses preprocessing sebelumnya. Dataset tersebut dibagi menjadi data pelatihan (*training data*) dan data pengujian (*testing data*) untuk mengevaluasi performa model yang dihasilkan. Proses pelatihan model bertujuan untuk mengidentifikasi pola hubungan antara variabel input seperti riwayat pinjaman, jumlah pinjaman, pendapatan, serta aktivitas digital dengan variabel target yaitu tingkat risiko kredit nasabah.

Hasil pengembangan model menunjukkan bahwa algoritma *machine learning* mampu mengidentifikasi pola perilaku nasabah yang berpotensi memiliki risiko kredit yang tinggi. Variabel yang memiliki pengaruh signifikan terhadap prediksi risiko kredit antara lain riwayat pinjaman sebelumnya, tingkat pendapatan, serta status

pembayaran pinjaman sebelumnya. Variabel-variabel tersebut menjadi indikator penting dalam menentukan tingkat risiko kredit calon nasabah.

4.3. Evaluasi Performa Model

Evaluasi model dilakukan untuk mengukur tingkat akurasi model analitik dalam memprediksi risiko kredit nasabah. Beberapa metrik evaluasi yang digunakan dalam penelitian ini meliputi *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, teknik *cross-validation* juga digunakan untuk memastikan bahwa model yang dikembangkan memiliki kemampuan generalisasi yang baik terhadap dataset yang berbeda.

Hasil evaluasi model menunjukkan bahwa algoritma *Random Forest* memiliki performa yang lebih baik dibandingkan algoritma lainnya dalam memprediksi risiko kredit nasabah. Hal ini disebabkan oleh kemampuan algoritma *Random Forest* dalam menggabungkan beberapa *decision tree* sehingga mampu menghasilkan prediksi yang lebih stabil dan akurat. Selain itu, algoritma ini juga mampu menangani dataset dengan jumlah variabel yang cukup banyak tanpa mengalami penurunan performa yang signifikan.

Sementara itu, algoritma *Logistic Regression* dan *Support Vector Machine* juga menunjukkan performa yang cukup baik dalam proses klasifikasi risiko kredit. Namun demikian, kedua algoritma tersebut memiliki keterbatasan dalam menangani kompleksitas hubungan antar variabel pada dataset yang memiliki dimensi tinggi. Oleh karena itu, dalam konteks penelitian ini, *Random Forest* dianggap sebagai algoritma yang paling optimal untuk digunakan dalam pengembangan model analitik penilaian risiko nasabah pada platform *online lending*.

4.4. Pembahasan

Hasil penelitian ini menunjukkan bahwa pemanfaatan *Big Data Analytics* dan *machine learning* dapat memberikan kontribusi signifikan dalam meningkatkan akurasi proses penilaian risiko kredit pada platform fintech lending. Dengan memanfaatkan berbagai sumber data seperti data transaksi, data profil pengguna, serta data perilaku digital, model analitik yang dikembangkan mampu memberikan gambaran yang lebih komprehensif mengenai karakteristik calon nasabah.

Pendekatan analitik berbasis data yang digunakan dalam penelitian ini juga menunjukkan bahwa integrasi berbagai sumber data dapat meningkatkan kualitas proses pengambilan keputusan pada sistem fintech lending. Hal ini sejalan dengan perkembangan teknologi finansial yang semakin mengandalkan analisis data untuk mendukung berbagai proses bisnis, termasuk dalam penilaian kelayakan kredit calon nasabah.

Selain itu, hasil penelitian ini juga menunjukkan bahwa penggunaan algoritma machine learning dapat membantu platform fintech dalam mengidentifikasi calon nasabah yang memiliki potensi risiko gagal bayar. Dengan adanya model analitik yang akurat, platform *Online Lending* dapat meminimalkan risiko kredit serta meningkatkan keberlanjutan operasional layanan fintech.

Secara praktis, model analitik yang dihasilkan dalam penelitian ini dapat diimplementasikan sebagai bagian dari sistem pendukung keputusan pada platform fintech lending. Sistem tersebut dapat digunakan untuk membantu perusahaan dalam melakukan evaluasi risiko kredit secara otomatis sehingga proses pengambilan keputusan menjadi lebih cepat, objektif, dan berbasis data. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi akademik dalam pengembangan literatur mengenai analitik data pada sektor fintech, tetapi juga memberikan manfaat praktis bagi pengembangan sistem informasi pada industri keuangan digital.

a. Perbandingan Performa Algoritma Machine Learning

Untuk menentukan model terbaik dalam memprediksi risiko kredit nasabah pada platform *online lending*, penelitian ini melakukan pengujian terhadap beberapa algoritma machine learning yaitu Logistic Regression, Decision Tree, Random Forest, dan Support Vector Machine (SVM). Evaluasi performa model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score.

Tabel 3. Perbandingan Performa Model

Algoritma	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.82	0.80	0.79	0.79
Decision Tree	0.85	0.83	0.82	0.82
Random Forest	0.91	0.90	0.89	0.89
Support Vector Machine	0.87	0.85	0.84	0.84

Berdasarkan hasil pengujian pada Tabel di atas, algoritma *Random Forest* menunjukkan performa terbaik dengan nilai accuracy sebesar 0.91. Hal ini menunjukkan bahwa Random Forest memiliki kemampuan yang lebih baik dalam mengklasifikasikan tingkat risiko kredit nasabah dibandingkan algoritma lainnya.

b. Confusion Matrix Model Terbaik

Untuk memahami performa model secara lebih detail, dilakukan analisis menggunakan *confusion matrix* pada algoritma *Random Forest* sebagai model terbaik dalam penelitian ini.

Tabel 4. Confusion Matrix Random Forest

Narasi		Prediksi Risiko Rendah	Prediksi Risiko Tinggi
Risiko (Aktual)	Rendah	420	35
Risiko (Aktual)	Tinggi	28	217

Berdasarkan hasil confusion matrix tersebut dapat diketahui bahwa:

- Sebanyak 420 data nasabah berhasil diklasifikasikan dengan benar sebagai risiko rendah.
- Sebanyak 217 data nasabah berhasil diprediksi dengan benar sebagai risiko tinggi.
- Kesalahan klasifikasi relatif kecil dibandingkan jumlah prediksi yang benar.

Hal ini menunjukkan bahwa model Random Forest memiliki tingkat keakuratan yang tinggi dalam memprediksi risiko kredit nasabah pada platform fintech lending.

c. Analisis Variabel yang Berpengaruh

Selain mengevaluasi performa model, penelitian ini juga melakukan analisis terhadap variabel yang memiliki pengaruh paling signifikan dalam proses prediksi risiko kredit. Berdasarkan hasil analisis model *Random Forest*, beberapa variabel yang memiliki kontribusi terbesar antara lain:

Tabel 5. Variabel Paling Berpengaruh

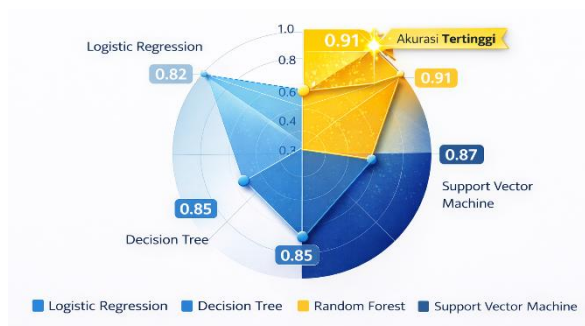
Variabel	Tingkat Pengaruh
Status pembayaran sebelumnya	Sangat tinggi
Tingkat pendapatan	Tinggi
Riwayat pinjaman	Tinggi
Jumlah pinjaman	Sedang
Aktivitas digital pengguna	Sedang

Hasil analisis menunjukkan bahwa status pembayaran pinjaman sebelumnya merupakan variabel yang paling berpengaruh dalam menentukan tingkat risiko kredit nasabah. Hal ini menunjukkan bahwa riwayat pembayaran menjadi indikator penting dalam menilai kelayakan kredit calon peminjam.

Grafik di bawah menunjukkan perbandingan tingkat akurasi empat algoritma *machine learning* yang digunakan dalam penelitian:

Tabel 6. Perbandingan algoritma prediksi risiko kredit

Algoritma	Accuracy
Logistic Regression	0.82
Decision Tree	0.85
Random Forest	0.91
Support Vector Machine	0.87



Gambar 3. Grafik Perbandingan tingkat akurasi

d. Interpretasi Hasil Penelitian

Hasil penelitian ini menunjukkan bahwa pemanfaatan Big Data Analytics dan algoritma *machine learning* mampu meningkatkan akurasi proses penilaian risiko kredit pada platform online lending. Dengan mengintegrasikan berbagai sumber data seperti data transaksi, data profil pengguna, serta data perilaku digital, model analitik yang dikembangkan mampu menghasilkan prediksi yang lebih akurat dibandingkan metode penilaian kredit tradisional.

Penggunaan algoritma *Random Forest* dalam penelitian ini terbukti memberikan performa terbaik dalam mengidentifikasi pola hubungan antar variabel yang mempengaruhi risiko kredit nasabah. Hal ini disebabkan oleh kemampuan *Random Forest* dalam mengkombinasikan beberapa model *decision tree* sehingga dapat mengurangi risiko overfitting serta meningkatkan stabilitas model prediksi.

Dari perspektif implementasi sistem informasi, model analitik yang dihasilkan dalam penelitian ini dapat diintegrasikan ke dalam sistem *decision support system* (DSS) pada platform fintech lending. Dengan demikian, perusahaan dapat melakukan proses evaluasi risiko kredit secara otomatis dan berbasis data, sehingga dapat meningkatkan efisiensi operasional serta meminimalkan potensi kerugian akibat gagal bayar.

5. KESIMPULAN

Penelitian ini bertujuan untuk mengembangkan model analitik penilaian risiko nasabah pada platform *Online Lending* dengan memanfaatkan pendekatan *Big Data Analytics* dan *machine learning*. Pendekatan ini digunakan

untuk meningkatkan akurasi proses penilaian risiko kredit calon nasabah dengan memanfaatkan berbagai sumber data yang meliputi data transaksi peminjam, data profil pengguna, serta data perilaku digital pengguna. Integrasi berbagai sumber data tersebut memungkinkan sistem analitik untuk menghasilkan analisis yang lebih komprehensif dalam mengidentifikasi karakteristik dan pola perilaku nasabah pada platform fintech lending.

Hasil penelitian menunjukkan bahwa proses pengolahan data yang meliputi *data collection*, *data preprocessing*, dan *data integration* mampu menghasilkan dataset yang siap digunakan dalam proses pengembangan model analitik. Proses *preprocessing* yang mencakup kegiatan *data cleaning*, *transformation*, *normalization*, serta *feature selection* berperan penting dalam meningkatkan kualitas data sehingga dapat mendukung kinerja algoritma *machine learning* secara optimal. Dengan dataset yang telah diproses dengan baik, model analitik dapat mempelajari pola hubungan antara variabel input dengan tingkat risiko kredit nasabah secara lebih efektif.

Pengujian model dilakukan dengan menggunakan beberapa algoritma *machine learning* yaitu *Logistic Regression*, *Decision Tree*, *Random Forest*, dan *Support Vector Machine*. Berdasarkan hasil evaluasi model menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*, algoritma *Random Forest* menunjukkan performa terbaik dengan tingkat akurasi tertinggi dibandingkan algoritma lainnya. Hal ini menunjukkan bahwa metode *Random Forest* memiliki kemampuan yang lebih baik dalam menangani kompleksitas hubungan antar variabel serta mampu menghasilkan prediksi risiko kredit yang lebih stabil dan akurat.

Secara keseluruhan, penelitian ini menunjukkan bahwa penerapan *Big Data Analytics* dan *machine learning* dapat memberikan kontribusi signifikan dalam meningkatkan kualitas sistem penilaian risiko kredit pada platform online lending. Model analitik yang dikembangkan dalam penelitian ini dapat dimanfaatkan sebagai bagian dari sistem pendukung keputusan (*decision support system*) pada platform fintech lending untuk membantu perusahaan dalam melakukan evaluasi kelayakan kredit calon nasabah secara lebih objektif, cepat, dan berbasis data.

Penelitian ini juga memberikan kontribusi akademik dalam pengembangan literatur mengenai penerapan analitik data pada sektor fintech, khususnya dalam konteks penilaian risiko kredit berbasis big data. Selain itu, hasil penelitian ini memiliki implikasi praktis bagi pengembangan sistem informasi pada industri keuangan digital yang semakin mengandalkan teknologi analitik data dalam proses pengambilan keputusan.

Sebagai arah penelitian selanjutnya, pengembangan model analitik dapat dilakukan dengan memanfaatkan dataset yang lebih besar serta integrasi sumber data tambahan seperti data media sosial atau data perilaku finansial lainnya. Selain itu, penelitian di masa depan juga dapat mengeksplorasi penggunaan algoritma *deep learning* atau metode *hybrid machine learning* untuk meningkatkan performa model prediksi risiko kredit pada platform fintech lending.

6. UCAPAN TERIMA KASIH

Peneliti mengucapkan terimakasih kepada Universitas Trisakti yang telah membantu dan terus menyemangati baik dalam memberikan pelatihan dan juga pendanaan dalam bidang penelitian.

DAFTAR PUSTAKA

- [1] L. Wang, C. Mayusoh, and A. Inkuer, "Sustainable Landscape Design and Traditional Villages in Xuzhou, Jiangsu: Low-cost Strategies and Big Data Applications Influencing AI Integration," *J. Inf. Syst. Eng. Manag.*, vol. 9, no. 2, 2024, doi: 10.55267/iadt.07.14748.
- [2] L. Lin and Z. Su, "Artificial Intelligence, Big Data, and Information System User Experience: A Comprehensive Approach to Cross-cultural Adaptation in Educational Management of International Students," *J. Inf. Syst. Eng. Manag.*, vol. 9, no. 3, 2024, doi: 10.55267/iadt.07.14985.
- [3] I. Purwanto, D. Despa, and A. Purba, "Technological-economic study of a unique hydrogen-based hybrid energy system for on-grid and off-grid use: Tegal-Pesawaran-Lampung Islands, Indonesia," in *E3S Web of Conferences*, R. R. Isnanto, Hadiyanto, and B. Warsito, Eds., Universitas Lampung, Department of Professional Engineering Program, Bandar Lampung, Lampung, Indonesia: EDP Sciences, 2023, doi: 10.1051/e3sconf/202344802040.
- [4] J. Jakubik, J. Schöffner, V. Hoge, M. Vössing, and N. Köhl, "An Empirical Evaluation of Predicted Outcomes as Explanations in Human-AI Decision-Making," in *Communications in Computer and Information Science*, K. I., M. P., G. R., J. S., F. H., G. F., F. P.M., R. D., C. G., N. S., G. J., R. R., G. R., M. E., R. Z., R. E., N. F., T. A., B. P., V. W., S. G., P. F., B. M., B. I., D. I.L., P. G., S. L., A. A., A. G., M. I., G. G., C. L., G. N., R. J., S. D., S. K., C. A., P. S., P. P., B. A., and P. S., Eds., Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany: Springer Science and Business Media Deutschland GmbH, 2023, pp. 353–368. doi: 10.1007/978-3-031-23618-1_24.
- [5] Y. Xue, C. Zhu, and Y. Lu, "Research on the Influence Mechanism of New Energy Vehicle Promotion Policy," *Sustain.*, vol. 17, no. 8, 2025, doi: 10.3390/su17083699.
- [6] A. Kirpalani, "Global ChatGPT interest across healthcare and education access," *Heal. Policy Technol.*, vol. 14, no. 5, 2025, doi: 10.1016/j.hlpt.2025.101061.
- [7] W. Cai, S. Jusoh, and X. Yue, "Digitalization and Sustainable Industrial Low-Carbon Transformation: Synergistic Effects, Policy Tools, Technical Pathways, and Financial Innovation," *Sustain.*, vol. 18, no. 3, 2026, doi: 10.3390/su18031433.
- [8] S. Wu, X. Pan, and F. Wang, "Can the synergy between FinTech and data element market development enhance credit allocation efficiency?," *Int. Rev. Econ. Financ.*, vol. 103, 2025, doi: 10.1016/j.iref.2025.104515.
- [9] C.-C. Lee, L. Fang, J. Zhao, C.-H. Yu, and J. Zhang, "HOW DOES FINTECH DEVELOPMENT DRIVE CORPORATE INNOVATION? NEW EVIDENCE FROM THE PERSPECTIVE OF FINANCIAL SUPPLY," *Technol. Econ. Dev. Econ.*, vol. 31, no. 1, pp. 244–279, 2024, doi: 10.3846/tede.2024.22192.
- [10] M. Kohlhase *et al.*, "Project VoLL-KI: Learning from Learners," *KI - Kunstl. Intelligenz*, vol. 39, no. 4, pp. 299–309, 2025, doi: 10.1007/s13218-024-00846-9.
- [11] M. Kohlhase *et al.*, "Project VoLL-KI: Learning from Learners," *KI - Kunstl. Intelligenz*, vol. 39, no. 4, pp. 299–309, 2025, doi: 10.1007/s13218-024-00846-9.
- [12] Y. S. Can, "LLM-Assisted Explainable Daily Stress Recognition: Physiologically Grounded Threshold Rules from PPG Features," *Electron.*, vol. 15, no. 1, 2026, doi: 10.3390/electronics15010201.
- [13] I. Hamdaoui, C. Rami, Z. El Allali, and K. El Makkaoui, "H-RT-IDPS: A Hierarchical Real-Time Intrusion Detection and Prevention System for the Smart Internet of Vehicles via TinyML-Distilled CNN and Hybrid BiLSTM-XGBoost Models," *Technologies*, vol. 13, no. 12, 2025, doi: 10.3390/technologies13120572.
- [14] M. A. Uddin, M. N. Islam, L. Maglaras, H. Janicke, and I. H. Sarker, "ExplainableDetector: Exploring transformer-based language modeling approach for SMS spam detection with explainability analysis," *Digit. Commun. Networks*, vol. 11, no. 5, pp. 1504–1518, 2025, doi: 10.1016/j.dcan.2025.07.008.
- [15] M. Babar, "A hybrid approach to financial big data analysis using extended ensemble learning and optimized spark streaming," *J. Open Innov. Technol. Mark. Complex.*, vol. 11, no. 3, 2025, doi: 10.1016/j.joitmc.2025.100602.
- [16] P. Chen, Z. Fan, and Y. Lu, "KnowSTU: Diagnosing Students' Problem Behaviors Using Fine-Tuned LLM and RAG," *IEEE Trans. Learn. Technol.*, vol. 18, pp. 962–975, 2025, doi: 10.1109/TLT.2025.3625146.
- [17] H. Sbai, "PRISMA: Physically-Aware Reasoning and Intelligent Semantic Mining Architecture for IoT Process Discovery Using Deep Learning and Large Language Models," *IEEE Access*, 2026, doi: 10.1109/ACCESS.2026.3664833.
- [18] V. Tiwari *et al.*, "Akshar Mitra: a multimodal integrated framework for early dyslexia detection," *Front. Digit. Heal.*, vol. 7, 2025, doi: 10.3389/fgth.2025.1726307.
- [19] K. M. Sujon, R. Hassan, K. Choi, and M. A. Samad, "Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models," *J. Big Data*, vol. 12, no. 1, 2025, doi: 10.1186/s40537-025-01313-4.

- [20] R. N. Turner-Jones, G. Tuxworth, R. A. Haupt, and L. Wallis, "Digitising the Deep Past: Machine Learning for Rock Art Motif Classification in an Educational Citizen Science Application," *J. Comput. Cult. Herit.*, vol. 17, no. 4, 2024, doi: 10.1145/3665796.
- [21] K. T. Chong, N. Ibrahim, S. H. Huspi, W. M. N. Wan Kadir, and M. A. Isa, "A SYSTEMATIC REVIEW OF MACHINE LEARNING TECHNIQUES FOR PREDICTING STUDENT ENGAGEMENT IN HIGHER EDUCATION ONLINE LEARNING," *J. Inf. Technol. Educ. Res.*, vol. 24, 2025, doi: 10.28945/5456.
- [22] I. M. García-López, M.-S. Ramirez-Montoya, and J. M. Molina Espinosa, "Generative artificial intelligence in education: a systematic analysis of opportunities, challenges, and responses," *Interact. Learn. Environ.*, 2025, doi: 10.1080/10494820.2025.2519133.
- [23] S. Darma, H. Laduppa, L. Susanti, and A. Rahmati, "Gen Z Rural Area Perspective on Use of E-Money Applications in Indonesia," *Rev. Mex. Econ. y Finanz.*, *Nueva Epoca*, vol. 20, no. 4, 2025, doi: 10.21919/remef.v20i4.1294.
- [24] T. Lin, J. Zhang, and B. Xiong, "Effects of Technology Perceptions, Teacher Beliefs, and AI Literacy on AI Technology Adoption in Sustainable Mathematics Education," *Sustain.*, vol. 17, no. 8, 2025, doi: 10.3390/su17083698.
- [25] M. T. Islam Khan and Y. Y. Xuan, "Drivers of lending decision in peer-to-peer lending in Malaysia," *Rev. Behav. Financ.*, vol. 14, no. 3, pp. 379–393, 2022, doi: <https://doi.org/10.1108/RBF-08-2020-0200>.
- [26] C. Liu and D. Zhou, "AI Integration in Higher Education: A Software Engineering Perspective on Readiness, Anxiety and Innovations," *Int. J. Softw. Eng. Knowl. Eng.*, 2026, doi: 10.1142/S0218194025501116.
- [27] O. Shvetsova, D. Katalshov, and S.-K. Lee, "Innovative Guardrails for Generative AI: Designing an Intelligent Filter for Safe and Responsible LLM Deployment," *Appl. Sci.*, vol. 15, no. 13, 2025, doi: 10.3390/app15137298.
- [28] O. S. Falebita, J. A. Abah, A. A. Asanre, T. O. Abiodun, M. A. Ayanwale, and O. K. Ayanwoye, "Determinants of Chatbot Brand Trust in the Adoption of Generative Artificial Intelligence in Higher Education," *Educ. Sci.*, vol. 15, no. 10, 2025, doi: 10.3390/educsci15101389.
- [29] I. Purwanto, R. Isnanto, and A. P. Widodo, "Investigation of the Impact of the Peer-To-Peer Lending Market on the Membership Motivation of the MSME," in *E3S Web of Conferences*, R. R. Isnanto, Hadiyanto, and B. Warsito, Eds., Universitas Diponegoro, Semarang, Central Java, Indonesia: EDP Sciences, 2023. doi: 10.1051/e3sconf/202344802037.
- [30] H. Sharma and S. Agarwal, "The impact of decentralized finance (Defi) on traditional financial systems: Opportunities, challenges, and regulatory implications," in *Studies in Systems, Decision and Control*, vol. 525, Department of Commerce, Graphic Era (Deemed to be University), Uttarakhand, Dehradun, India: Springer Science and Business Media Deutschland GmbH, 2024, pp. 211–218. doi: 10.1007/978-3-031-54383-8_17.
- [31] Y. Cao, M. Hizam-Hanafiah, M. Ghazali, R. Ab Razak, and Y. Zheng, "Estimating the Impact of Government Green Subsidies on Corporate ESG Performance: Double Machine Learning for Causal Inference," *Sustain.*, vol. 18, no. 1, 2026, doi: 10.3390/su18010281.
- [32] X. Zhang and Y. Wei, "The Impact Mechanism of AI Technology on Enterprise Innovation Resilience," *Sustain.*, vol. 17, no. 11, 2025, doi: 10.3390/su17115169.
- [33] A. C. Alves-Noreña, M.-J. Rodríguez-Conde, J. P. Hernández-Ramos, and J. W. Castro-Salgado, "Technology and Emotions: AI-Driven Software Prototyping for the Analysis of Emotional States and Early Detection of Risky Behaviors in University Students," *Educ. Sci.*, vol. 15, no. 3, 2025, doi: 10.3390/educsci15030350.
- [34] S. Liu *et al.*, "Beyond the eye: A relational model for early dementia detection using retinal OCTA images," *Med. Image Anal.*, vol. 102, 2025, doi: 10.1016/j.media.2025.103513.
- [35] S. Phodhiran *et al.*, "A Competency-Based Curriculum for Fostering Artificial Intelligence Skills in Thai Children and Youth," *Int. J. Eng. Trends Technol.*, vol. 72, no. 10, pp. 357–372, 2024, doi: 10.14445/22315381/IJETT-V72I10P135.
- [36] B. Teixeira, L. Carvalhais, T. Pinto, and Z. Vale, "Explainable AI framework for reliable and transparent automated energy management in buildings," *Energy Build.*, vol. 347, 2025, doi: 10.1016/j.enbuild.2025.116246.
- [37] T. Hua, "Machine Learning Approaches to Creditworthiness Classification," in *BDAIE 2025 - Proceedings of 2025 International Conference on Big Data, Artificial Intelligence and Digital Economy*, University of Nottingham, Nottingham, United Kingdom: Association for Computing Machinery, Inc, 2025, pp. 127–134. doi: 10.1145/3767052.3767072.
- [38] M. Neshat, M. Phipps, N. Jha, D. Khojasteh, M. Tong, and A. H. Gandomi, "Effective Predictive Modelling for Emergency Department Visits and Evaluating Exogenous Variables Impact: Using Explainable Meta-Learning Gradient Boosting," *ACM Trans. Comput. Healthc.*, vol. 7, no. 1, 2026, doi: 10.1145/3768317.
- [39] W. Dossche, S. Vansteenkiste, B. Baesens, and W. Lemahieu, "Anticipating delays in recruitment: Explainable machine learning for the prediction of hard-to-fill online job vacancies," *Eur. J. Oper. Res.*, vol. 328, no. 2, pp. 680–693, 2026, doi: 10.1016/j.ejor.2025.06.027.
- [40] K. Danash, H. Harb, H. Issa, and L. Saker, "An explainable data-driven optimization framework for industrial predictive maintenance scheduling," *Results*

- Eng.*, vol. 29, 2026, doi: 10.1016/j.rineng.2026.109022.
- [41] I. Purwanto and R. Isnanto, "A Fuzzy Logic Model for Loan Recommendations in Online Lending Systems Using the California Psychological Inventory," *Ing. des Syst. d'Information*, vol. 30, no. 4, pp. 923–932, 2025, doi: 10.18280/isi.300409.
- [42] Z. H. Haghghat, A. Islam, H. Karimipour, and B. M. Fard, "An explainable big transfer learning approach for IoT-based safety management in smart factories," *Internet Things (The Netherlands)*, vol. 31, 2025, doi: 10.1016/j.iot.2025.101600.
- [43] L. B. Albate, "Recent Advancements and Emerging Trends in the Use of Artificial Intelligence for Risk Assessment in Internal Auditing," *Int. J. Comput. Digit. Syst.*, vol. 18, no. 1, 2025, doi: 10.12785/ijcds/1571163965.
- [44] F. Gutiérrez-Castillo, M. Guo, N. Zhang, X. Bai, and P. Jiao, "Comparing the Effects of Instructor Manual Feedback and ChatGPT Intelligent Feedback on Collaborative Programming in China's Higher Education," *IEEE Trans. Learn. Technol.*, vol. 17, pp. 2227–2239, 2024, doi: 10.1109/TLT.2024.3486749.
- [45] D. Kalatzis, A. Nega, and Y. Kiouvrekis, "Raman Spectra Classification of Pharmaceutical Compounds: A Benchmark of Machine Learning Models with SHAP-Based Explainability," *Eng.*, vol. 6, no. 7, 2025, doi: 10.3390/eng6070145.
- [46] Y. Shen, "Exploring the effectiveness of natural language processing technologies in automatically generating English teaching resource content," *J. Comput. Methods Sci. Eng.*, vol. 25, no. 5, pp. 4817–4832, 2025, doi: 10.1177/14727978251337956.
- [47] A. Forootani, D. E. Esmaili Aliabadi, and D. Thraen, "Bio-Eng-LLM AI Assist: A modular chatbot platform for interdisciplinary research and education," *SoftwareX*, vol. 31, 2025, doi: 10.1016/j.softx.2025.102260.
- [48] Z. Bouzidi and A. Boudries, "Managing emergency crises using secure information through educational awareness: COVID-19 case study," *Comput. Biol. Med.*, vol. 186, 2025, doi: 10.1016/j.combiomed.2024.109620.
- [49] G. F. Bati, "AIFloos: a balanced dataset for the sixth issue of the Saudi Arabian currency banknote*," *J. Umm Al-Qura Univ. Eng. Archit.*, vol. 15, no. 4, pp. 446–454, 2024, doi: 10.1007/s43995-024-00067-z.