



**STATISTICS ESSAY COMPETITION (SEC)
SATRIA DATA 2024
Telkom University**



**FinPredictor: Sistem Rekomendasi Kelayakan berdasarkan
Kondisi Sosial Ekonomi untuk Pelanggan Pinjaman *Online* Guna
Meningkatkan Keberlanjutan Finansial dan Kesejahteraan Pelanggan**

Sosial-Ekonomi

SEC_ SD2024020000326

1. Pendahuluan

Indonesia merupakan negara dengan populasi lebih dari 270 juta jiwa, dengan kondisi sosial ekonomi yang sangat beragam. Berdasarkan data Badan Pusat Statistik (BPS), pada tahun 2021, tingkat kemiskinan di Indonesia mencapai 10,14% atau sekitar 27,54 juta orang (Badan Pusat Statistik, 2021). Selain itu, tingkat pengangguran terbuka mencapai 7,07% pada Agustus 2020, meningkat dari 5,23% pada periode yang sama tahun sebelumnya akibat pandemi COVID-19. Kesulitan ekonomi yang dialami masyarakat ini mendorong banyak individu mencari alternatif pembiayaan untuk memenuhi kebutuhan sehari-hari.

Menjawab "alternatif" yang diinginkan masyarakat munculah pinjaman *online* sebagai Perkembangan teknologi informasi telah membawa perubahan signifikan di industri keuangan. Layanan *financial technology* (*fintech*) lending menjadi salah satu inovasi yang memungkinkan individu untuk mendapatkan dana secara daring dengan lebih mudah melalui aplikasi *fintech*.

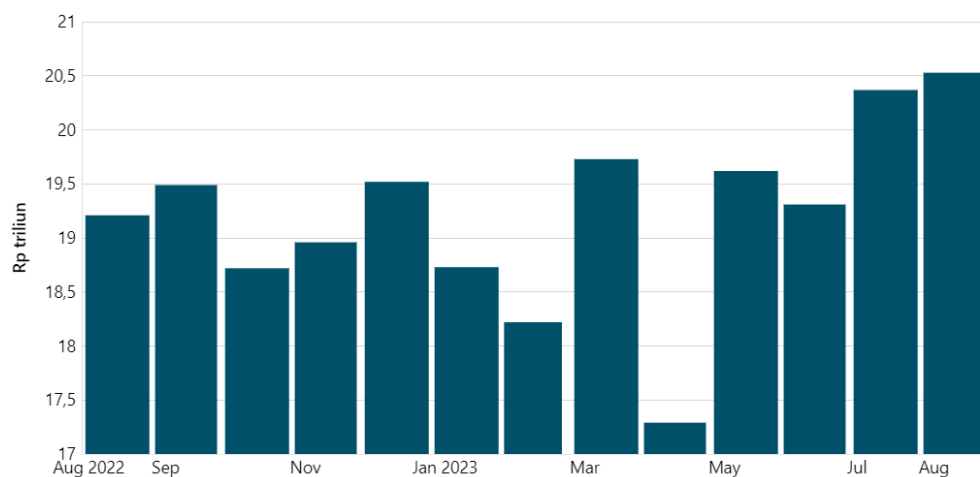
Namun, maraknya pengajuan pinjaman *online* juga membuka celah bagi praktik kejahatan. Sejumlah aplikasi *fintech* lending tidak memiliki izin resmi dari Otoritas Jasa Keuangan (OJK). Pada laporan Kinerja OJK 2019, disebutkan bahwa sekitar 1.400

fintech lending ilegal telah dihentikan, sementara jumlah *fintech* berizin hanya sekitar 144 perusahaan. Keberadaan *fintech* ilegal ini sering kali menimbulkan masalah serius bagi nasabah (Sidik, 2021).

Meskipun demikian, minat masyarakat terhadap pinjol terus meningkat. Pada Februari 2021, total pinjaman yang tersalurkan mencapai sekitar Rp 169 triliun, meningkat menjadi lebih dari Rp 207 triliun dalam tiga bulan berikutnya. Meskipun terdapat berita negatif seputar pinjol, kredibilitas layanan *fintech* tidak sepenuhnya menurun di mata masyarakat (Sidik, 2021).

Berdasarkan data terbaru dari OJK, penyaluran pinjaman *online* mencapai Rp20,53 triliun pada Agustus 2023, meningkat 6,87% dibandingkan Agustus 2022 (Annur, 2023).

Nilai Penyaluran Pinjaman Online di Indonesia (Agustus 2022-Agustus 2023)



Gambar 1. Nilai Penyaluran Pinjaman *Online* di Indonesia.

Beberapa kondisi seperti kemudahan akses, keamanan transaksi, dan responsif terhadap kebutuhan mendesak seperti kesehatan, pendidikan, serta kebutuhan gaya hidup, memberikan pemahaman yang mendalam tentang faktor-faktor internal dan eksternal yang mendorong pelanggan untuk menggunakan pinjaman *online* (pinjol). Melalui proses yang lebih mudah dibandingkan dengan alternatif tradisional dan didukung oleh promosi aktif melalui media sosial, pinjol menjadi solusi yang cepat dan relevan untuk

memenuhi kebutuhan finansial sehari-hari atau mendesak. Ini tidak hanya mencerminkan preferensi akan kenyamanan dan keamanan dalam transaksi keuangan, tetapi juga menunjukkan pergeseran perilaku konsumen terhadap inovasi teknologi finansial yang dapat mendukung keberlanjutan finansial dan kesejahteraan individu (Rahmadyanto, 2023).

Sayangnya, banyak pelanggan yang tidak menyadari dampak dikemudian hari sehingga terjebak dalam siklus utang akibat bunga yang tinggi dan biaya tersembunyi dari pinjol. Ketika mereka tidak mampu membayar hutang pada saat jatuh tempo, penagihan seringkali dialihkan kepada *debt collector* yang menggunakan metode intimidasi dan paksaan, seperti datang langsung ke rumah atau kantor dengan ancaman verbal, pelecehan seksual, dan *cyber bullying*. Debt collector bahkan memperoleh akses ke data pribadi konsumen melalui ponsel, termasuk foto di galeri, media sosial, dan aplikasi *online* lainnya. Situasi ini tidak hanya mengganggu kehidupan pribadi dan profesional konsumen, tetapi juga bisa menyebabkan trauma psikologis serius, seperti stres berat, depresi, dan bahkan berpotensi menyebabkan kehilangan pekerjaan atau bahaya bunuh diri. Ancaman penagihan yang melibatkan keluarga dan rekan kerja juga dapat merusak hubungan sosial dan menyebabkan hilangnya kepercayaan diri (Arvante, 2022).

Oleh karena itu kami mengembangkan “FinPredictor”, sistem rekomendasi AI untuk menilai kelayakan pelanggan pinjaman *online* berdasarkan kondisi sosial ekonomi. Menggunakan dataset dan algoritma *machine learning*, sistem ini akan menghasilkan rekomendasi akurat yang mengurangi risiko gagal bayar dan dampak negatif bagi pengguna pinjol di Indonesia. Teknologi ini diharapkan menciptakan ekosistem pinjaman yang lebih sehat dan berkelanjutan serta melindungi konsumen dari risiko finansial dan psikologis.

2. Pembahasan

2.1. Studi Literatur

Studi literatur mengenai fenomena pinjaman *online* (pinjol) dan dampaknya terhadap kondisi sosial ekonomi di Indonesia mengungkap beberapa aspek kunci. Pertama,

perkembangan teknologi finansial (*fintech*) seperti pinjaman *online* telah signifikan dalam mempermudah akses keuangan masyarakat dengan pelayanan yang lebih efisien dan transparan (Gomber et al., 2017). *Fintech* diakui dapat mengatasi keterbatasan akses keuangan, khususnya di negara-negara berkembang termasuk Indonesia.

Penelitian terkait sistem rekomendasi kelayakan kredit dengan menggunakan pembelajaran mesin menunjukkan bahwa teknik *deep learning* dan *gradient boosting* mampu memberikan akurasi prediksi yang tinggi, mencapai sekitar 85% (Schmitt et al., 2022; Kriebel & Stitz, 2022). Hal ini menegaskan potensi besar teknologi pembelajaran mesin dalam meningkatkan efisiensi dan akurasi proses penilaian kelayakan kredit di sektor *fintech*.

Faktor sosial ekonomi juga memiliki pengaruh signifikan dalam penilaian kelayakan kredit. Studi Gunnarsson et al. (2021) menunjukkan bahwa peminjam dengan latar belakang pendidikan tinggi dan pekerjaan tetap cenderung memiliki risiko gagal bayar yang lebih rendah, menggarisbawahi pentingnya mempertimbangkan variabel ini dalam model prediktif kelayakan kredit.

Implementasi *artificial intelligence* (AI) dalam penilaian kredit di negara berkembang seperti yang dikaji oleh Hjelkrem et al. (2022) juga menunjukkan peningkatan signifikan dalam akurasi prediksi, menegaskan bahwa AI dapat menjadi alat yang kuat dalam meningkatkan efektivitas sistem penilaian kredit.

Analisis komparatif terhadap efektivitas model prediksi kelayakan kredit di berbagai negara menunjukkan bahwa penggunaan data lokal dapat meningkatkan performa model secara signifikan (Hjelkrem et al., 2022). Ini menyoroti pentingnya konteks lokal dalam pengembangan dan implementasi model pembelajaran mesin untuk penilaian kredit yang lebih akurat dan relevan dengan kondisi pasar lokal.

2.2. Pengumpulan Data

Untuk mengembangkan FinPredictor, langkah pertama adalah mengumpulkan data yang relevan dan komprehensif. Data ini diambil dari *Kaggle*, "*Loan Default Prediction Challenge*," yang diselenggarakan oleh *Coursera*. Dataset ini terdiri dari 255,347 baris dan 18 kolom, memberikan informasi rinci tentang pinjaman dan profil peminjam.

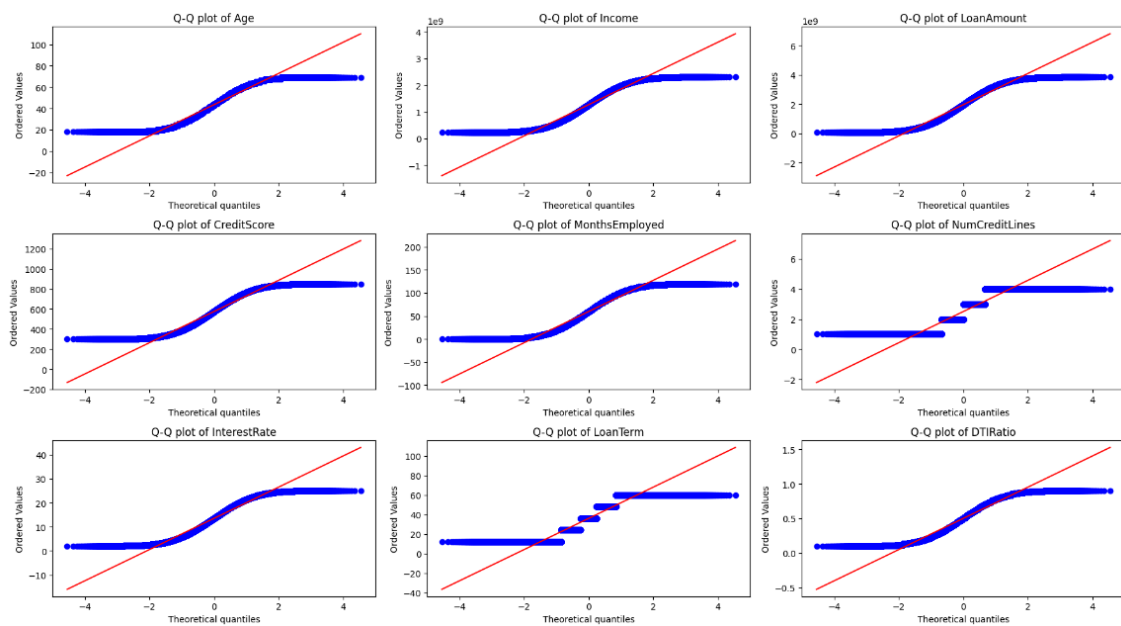
Setiap pinjaman memiliki identifikasi unik (*LoanID*) dan mencakup berbagai fitur penting seperti usia peminjam (*Age*), pendapatan (*Income*), jumlah pinjaman (*LoanAmount*), dan skor kredit (*CreditScore*). Informasi lain termasuk lama bekerja peminjam (*MonthsEmployed*), jumlah jalur kredit yang dimiliki (*NumCreditLines*), suku bunga pinjaman (*Interest Rate*), dan jangka waktu pinjaman (*LoanTerm*). Rasio utang terhadap pendapatan (*DTIRatio*) serta tingkat pendidikan (*Education*) seperti *PhD*, *Master*, *Bachelor*, dan *High School*. Jenis pekerjaan (*EmploymentType*) termasuk *Full-time*, *Part-time*, *Self-employed*, dan *Unemployed*, serta status pernikahan (*MaritalStatus*) seperti *Single*, *Married*, dan *Divorced* juga disertakan. Dataset ini mencatat apakah peminjam memiliki hipotek (*HasMortgage*) atau tanggungan (*HasDependents*), tujuan pinjaman (*LoanPurpose*) seperti untuk *Home*, *Auto*, *Education*, *Business*, atau *Other*, dan apakah memiliki penjamin (*HasCoSigner*). Variabel target biner (*Default*) menunjukkan apakah pinjaman mengalami *default* (1) atau tidak (0). Informasi ini digunakan untuk membangun model prediksi kelayakan kredit yang akurat dan berbasis kondisi sosial ekonomi.

2.3. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) merupakan tahap awal yang penting dalam proses pengembangan sistem rekomendasi kelayakan pinjaman *online*. Tujuan utama dari EDA adalah untuk memahami karakteristik dataset, pola, dan hubungan antar variabel sebelum membangun model prediktif.

2.3.1. Analisis Normalitas

Plot Q-Q (*Quantile-Quantile*) digunakan untuk memeriksa apakah distribusi dari suatu variabel mengikuti distribusi normal.



Gambar 2. Plot Q-Q Distribusi Variabel Numerik dalam Dataset

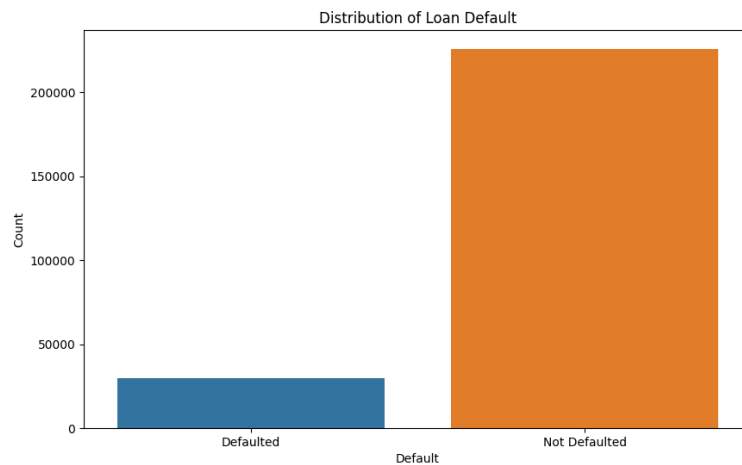
Analisis plot Q-Q menunjukkan bahwa sebagian besar variabel numerik dalam dataset memiliki distribusi mendekati normal dengan beberapa penyimpangan. Usia, skor kredit, dan lama bekerja mendekati garis diagonal di tengah, tetapi terdapat outlier di ekor. Pendapatan, jumlah pinjaman, jumlah jalur kredit, dan jangka waktu pinjaman menunjukkan penyimpangan signifikan dari garis diagonal, mengindikasikan distribusi yang jauh dari normal. Suku bunga dan rasio utang terhadap pendapatan juga mendekati garis diagonal di tengah tetapi memiliki penyimpangan di ekor. Secara keseluruhan, tidak ada variabel yang sepenuhnya normal, sehingga scaling diperlukan pada tahap *pre-processing* untuk memperbaiki distribusi data.

2.3.2. Analisis Statistik Deskriptif

Analisis statistik deskriptif memberikan gambaran umum tentang distribusi data dan statistik deskriptif dari setiap variabel dalam dataset. Misalnya, usia rata-rata pemohon adalah 43,5 tahun dengan standar deviasi 15 tahun, menunjukkan variasi yang signifikan dalam distribusi usia. Pendapatan rata-rata adalah 82.499 dengan standar deviasi 38.963, mengindikasikan adanya variasi pendapatan yang besar di antara pemohon. Jumlah pinjaman yang diminta memiliki rentang dari 5.000 hingga 249.999, dengan rata-rata 127.579 dan standar deviasi 70.841.

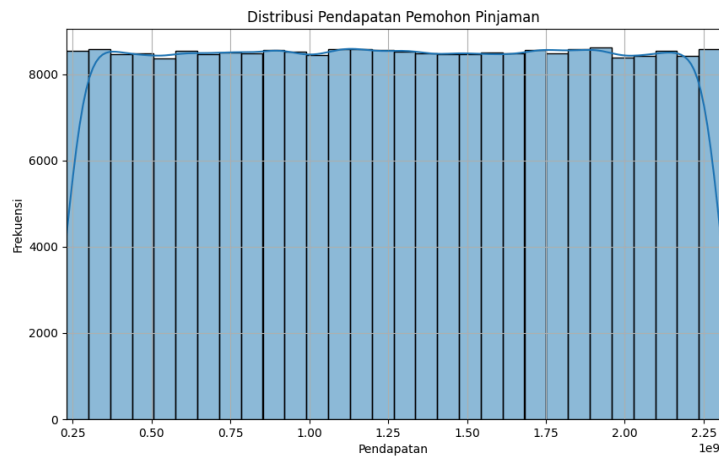
2.3.3. Analisis *Univariate*

Dalam analisis univariate pertama pada dataset kelayakan pinjaman online, fokus diberikan pada distribusi variabel secara terpisah untuk memahami karakteristik dan sebaran data. Sebuah histogram menunjukkan bahwa terdapat ketidakseimbangan yang signifikan antara jumlah data yang gagal bayar (kurang dari 50.000 entri) dan yang berhasil bayar (lebih dari 200.000 entri), yang dapat mempengaruhi pembelajaran model *machine learning* dengan memprioritaskan data mayoritas dan potensial menyebabkan bias.



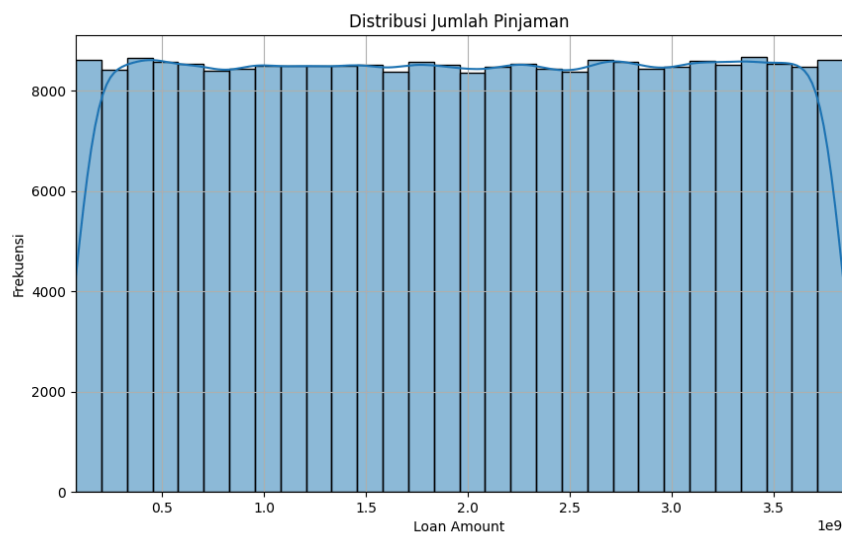
Gambar 3. Distribusi Nilai *Default* pada Dataset.

Histogram pendapatan pemohon menunjukkan sebaran pendapatan tahunan pemohon pinjaman, terpusat di kisaran 0.25 miliar hingga 2.25 miliar rupiah, menunjukkan bahwa banyak pemohon memiliki pendapatan dalam kisaran ini.



Gambar 4. Distribusi Nilai Pendapatan Pemohon Pinjaman.

Grafik garis yang menampilkan distribusi jumlah pinjaman dalam dataset memperlihatkan bahwa sebagian besar pinjaman memiliki jumlah yang relatif besar, dengan frekuensi tertinggi terjadi di berbagai kisaran pinjaman.

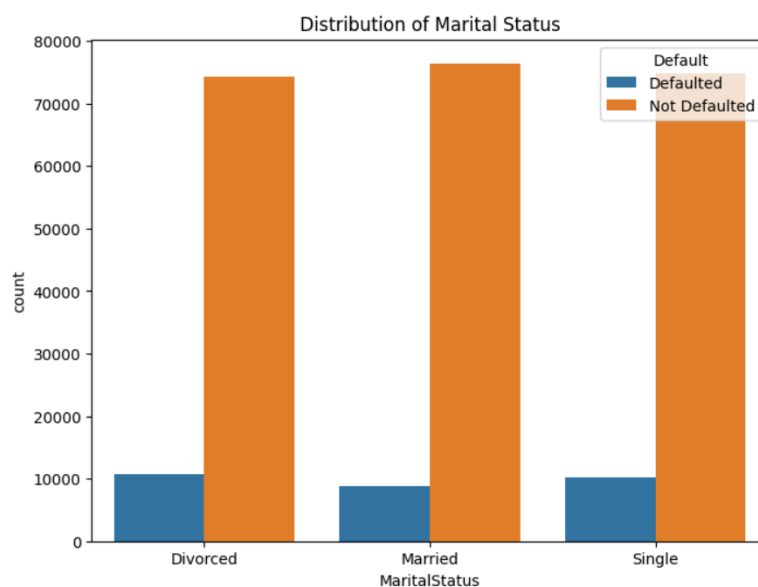


Gambar 5. Distribusi Jumlah Pinjaman

Analisis ini memberikan pandangan yang mendalam tentang karakteristik dataset kelayakan pinjaman *online*, yang penting untuk memahami profil pemohon dan menyesuaikan strategi penentuan kelayakan pinjaman.

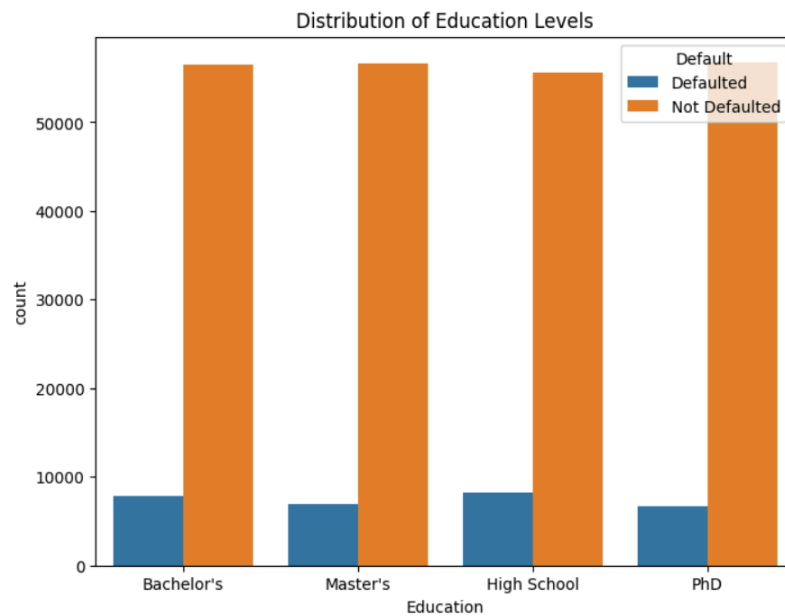
2.3.4. Analisis *Bivariate*

Analisis *bivariate* dilakukan untuk memahami pola dan hubungan antara fitur sosial ekonomi peminjam dan status *default* pinjaman. Data menunjukkan bahwa jumlah peminjam yang mengalami default di setiap kategori status pernikahan (bercerai, menikah, lajang) adalah sekitar 10.000 orang. Peminjam yang tidak mengalami default masing-masing berjumlah sekitar 70.000 (bercerai) dan 75.000 (menikah dan lajang). Ini mengindikasikan bahwa status pernikahan tidak signifikan mempengaruhi kecenderungan *default*.



Gambar 6. Distribusi Jumlah Peminjam yang Mengalami *Default* di Setiap Kategori Status Pernikahan.

Peminjam dengan berbagai tingkat pendidikan (Bachelor's, Master's, High School, PhD) yang mengalami default berjumlah sekitar 10.000 di tiap kategori, sementara yang tidak default sekitar 55.000. Hal ini menunjukkan bahwa tingkat pendidikan tidak memiliki pengaruh signifikan terhadap risiko default.

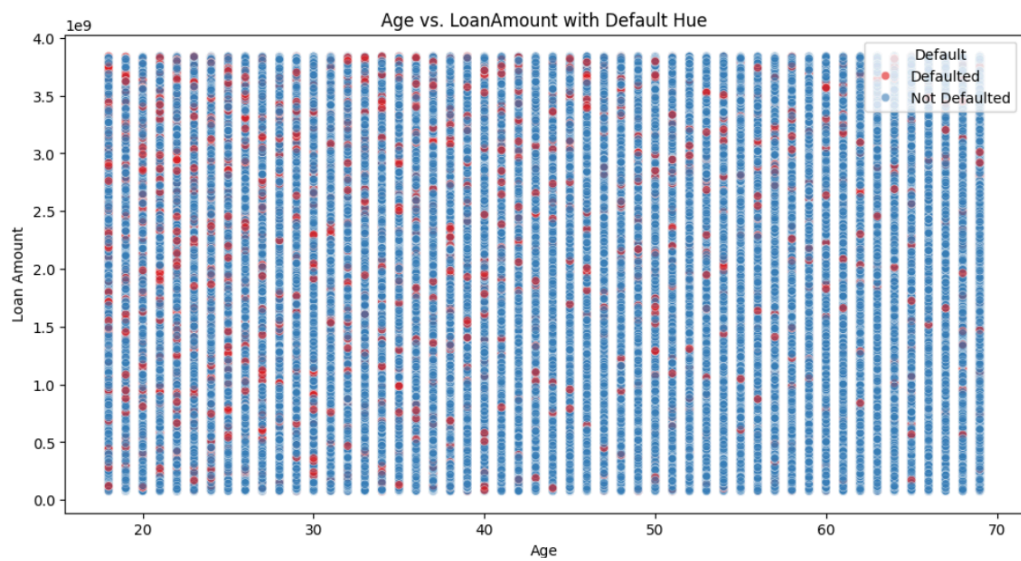


Gambar 7. Distribusi Peminjam dengan Berbagai Tingkat Pendidikan yang Mengalami *Default*.

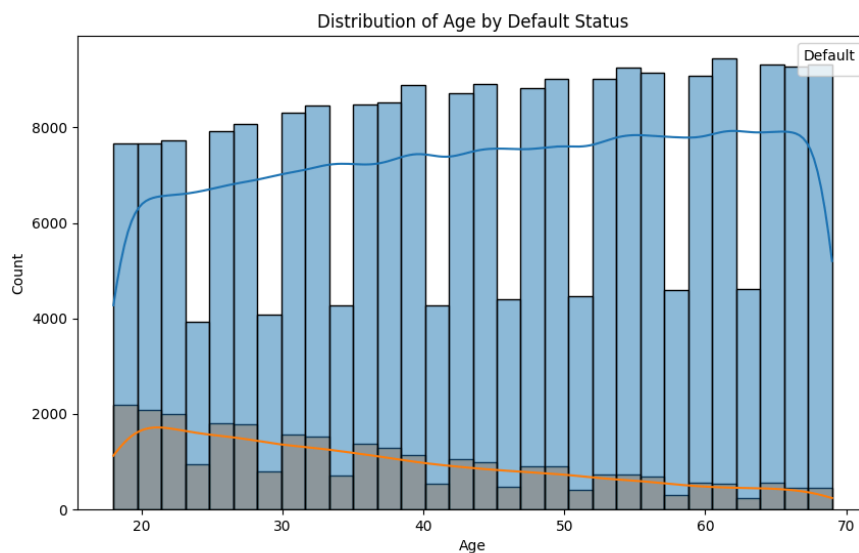
Dari analisis ini, dapat disimpulkan bahwa status pernikahan, tingkat pendidikan, jenis pekerjaan, kepemilikan hipotek, kepemilikan tanggungan, tujuan pinjaman, dan kepemilikan penandatanganan bersama tidak secara signifikan mempengaruhi risiko default. Kebijakan penilaian risiko kredit sebaiknya lebih fokus pada faktor lain seperti pendapatan, jumlah pinjaman, atau skor kredit untuk memberikan rekomendasi yang lebih akurat dan efektif.

2.3.5. Analisis *Multivariate*

Analisis *multivariate* menggunakan scatter plot dan histogram menunjukkan bahwa usia dan jumlah pinjaman tidak menunjukkan pola yang jelas dalam hubungannya dengan kemungkinan default. Namun, histogram mengungkapkan bahwa peminjam muda (20-30 tahun) dan tua (lebih dari 60 tahun) memiliki angka default lebih tinggi dibandingkan kelompok usia menengah.

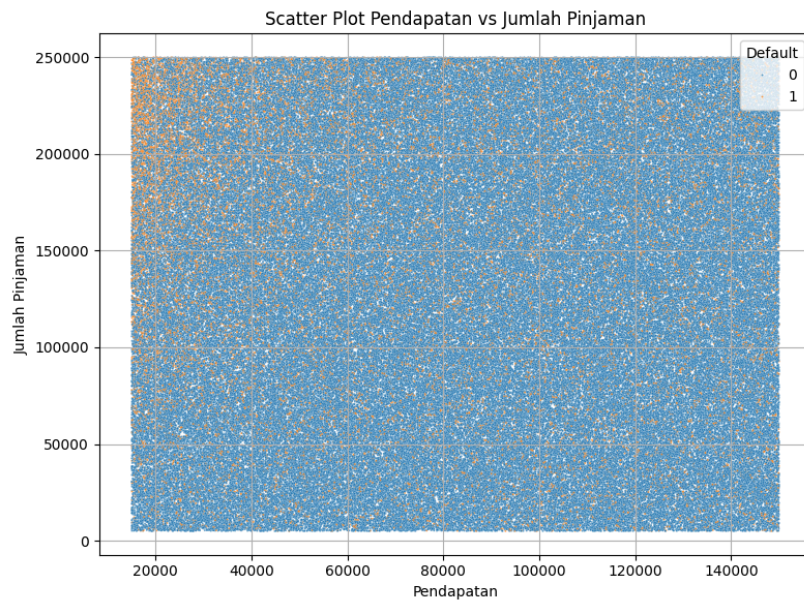


Gambar 8. Distribusi Usia dan Jumlah Pinjaman berdasarkan Nilai *Default*



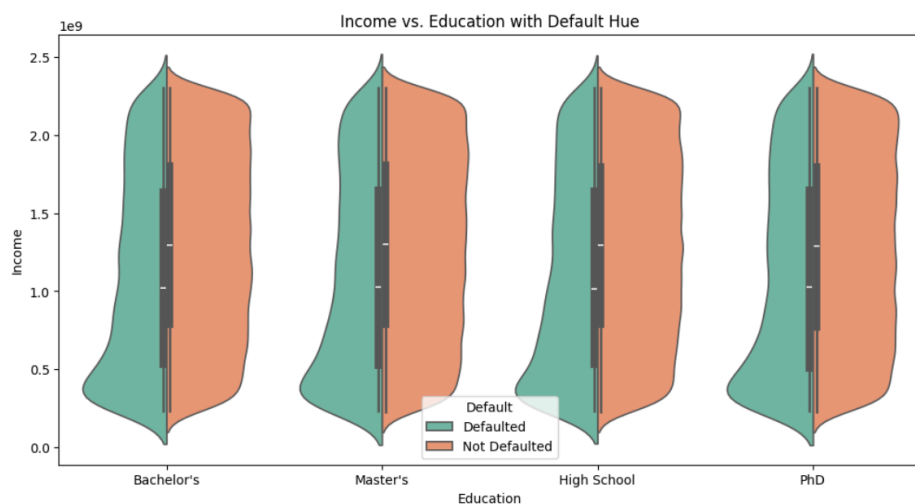
Gambar 9. Distribusi Usia berdasarkan Nilai *Default*

Scatter plot antara pendapatan dan jumlah pinjaman memperlihatkan bahwa kedua variabel tersebut saja tidak cukup untuk memprediksi default. Meskipun peminjam dengan pendapatan dan jumlah pinjaman lebih rendah cenderung lebih sering mengalami default, sebagian besar peminjam tetap tidak mengalami default, terlepas dari pendapatan dan jumlah pinjaman mereka.



Gambar 10. Distribusi Pendapatan dan Jumlah Pinjaman berdasarkan Nilai *Default*

Violin plot menunjukkan distribusi pendapatan untuk peminjam yang default dan tidak default di setiap tingkat pendidikan. Distribusi pendapatan serupa di semua kategori pendidikan, dengan median sekitar 1,5 miliar, mengindikasikan bahwa pendapatan bukanlah faktor penentu utama status *default*.



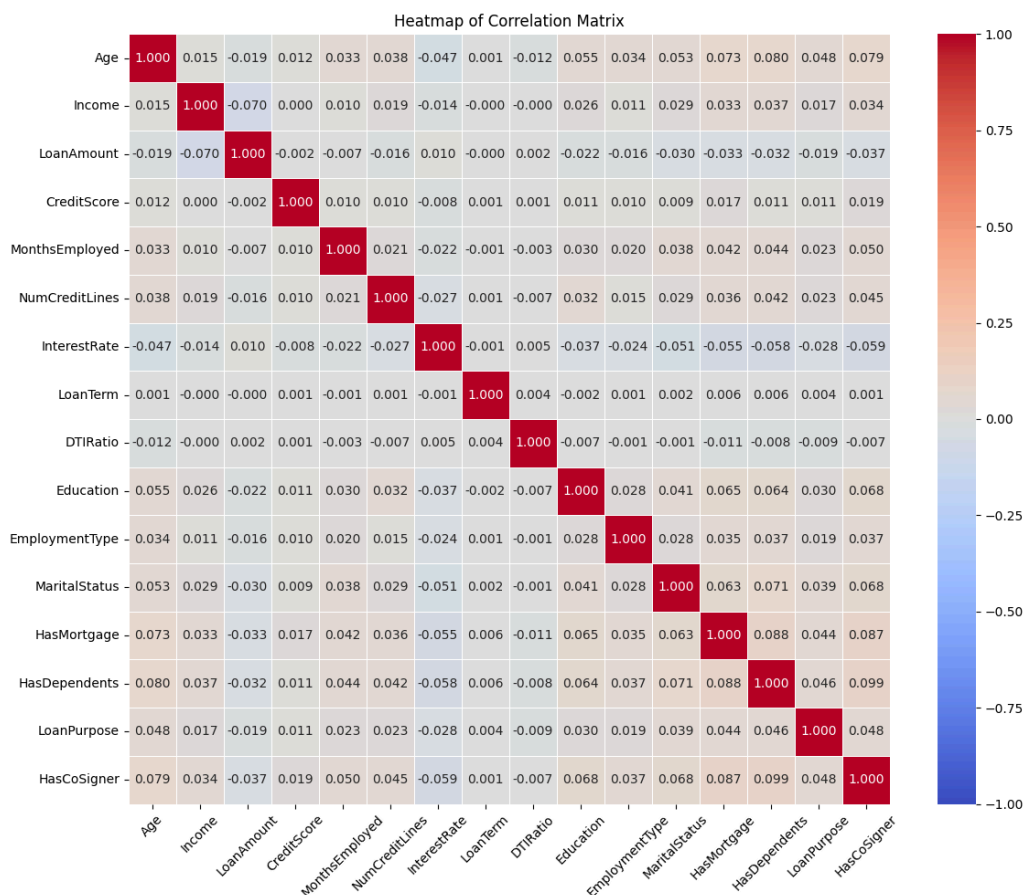
Gambar 11. Distribusi Pendapatan dan Tingkat Pendidikan berdasarkan Nilai *Default*

Kesimpulannya, meskipun faktor usia dan jumlah pinjaman memiliki pengaruh terhadap risiko *default*, pendapatan dan tingkat pendidikan saja tidak cukup untuk memprediksi *default*. Model prediksi kelayakan kredit harus mempertimbangkan banyak variabel dan

menggunakan teknik machine learning untuk menangani interaksi non-linear antar variabel, guna mengidentifikasi pola risiko *default* yang lebih kompleks dan memastikan kebijakan kredit yang adil serta tidak bias.

2.3.6. Analisis Korelasi

Heatmap korelasi pada dataset kelayakan pinjaman menunjukkan sebagian besar variabel numerik memiliki korelasi mendekati nol, menunjukkan hubungan linier yang lemah atau tidak ada. Namun, terdapat beberapa temuan yang menarik, seperti korelasi positif antara lamanya bekerja dan pendapatan, serta antara jumlah jalur kredit dan jumlah pinjaman. Di sisi lain, terlihat korelasi negatif antara rasio utang terhadap pendapatan dan pendapatan, serta antara suku bunga dan skor kredit. Temuan ini menyoroti kompleksitas hubungan variabel dalam mengevaluasi kelayakan pinjaman, mendorong perlunya pendekatan machine learning yang lebih canggih untuk mempertimbangkan interaksi non-linear dengan lebih baik.



Gambar 12. Matriks Korelasi *Heatmap*

Dalam konteks spesifik, heatmap juga mengungkapkan korelasi moderat antara variabel seperti suku bunga dan jangka waktu pinjaman (0.003), serta antara rasio hutang terhadap pendapatan (DTI) dan jumlah garis kredit (0.003). Temuan ini memberikan wawasan penting dalam memilih fitur untuk pengembangan model prediktif, serta dalam mengidentifikasi aspek-aspek yang memerlukan analisis lebih lanjut terkait pengelolaan risiko dan strategi keputusan bisnis

2.4. *Pre-processing* Data

Pre-processing data adalah tahap krusial dalam pengembangan sistem rekomendasi kelayakan pinjaman online untuk memastikan data siap digunakan dan sesuai dengan kebutuhan analisis. Berikut adalah langkah-langkah *pre-processing* yang diterapkan:

2.4.1. Konversi Mata Uang

Untuk melakukan konversi nilai dalam dataset dari USD ke IDR, digunakan fungsi `convert_to_idr` dengan parameter `amount_usd` dan `exchange_rate`, di mana nilai tukar default adalah 15.358 namun dapat disesuaikan. Kolom-kolom seperti `Income` dan `LoanAmount` yang awalnya dalam USD dikalikan dengan nilai tukar untuk mengonversi menjadi IDR. Pendekatan ini memastikan konsistensi data dalam mata uang yang sama selama analisis dan pemrosesan selanjutnya.

2.4.2. Melakukan Encode Fitur Kategorikal

Untuk memproses dataset kelayakan pinjaman, fitur kategorikal seperti `Education`, `EmploymentType`, `MaritalStatus`, `HasMortgage`, `HasDependents`, `LoanPurpose`, dan `HasCoSigner` di-encode menggunakan `LabelEncoder`. Ini mengubah nilai kategorikal menjadi bilangan bulat unik, memungkinkan algoritma machine learning memproses data lebih efisien. Proses ini penting karena kebanyakan algoritma hanya menerima input numerik, sehingga mempersiapkan data untuk analisis lanjutan dan prediksi risiko default.

2.4.3. Penghapusan Fitur yang Tidak Diperlukan

Dalam preprocessing sistem rekomendasi kelayakan pinjaman online, kolom `LoanID` dihapus karena tidak relevan, dan baris dengan nilai kosong dihapus untuk memastikan

kebersihan dataset. Langkah-langkah ini meningkatkan efisiensi analisis dan fokus pada fitur yang relevan dalam menilai kelayakan peminjam.

2.4.4. Penanganan Ketidakseimbangan Data dengan Teknik SMOTE

Penggunaan *Synthetic Minority Over-sampling Technique* (SMOTE) dalam preprocessing mengatasi ketidakseimbangan data pada sistem rekomendasi kelayakan pinjaman online. Setelah menghapus fitur yang tidak diperlukan dan menangani nilai hilang, SMOTE menyeimbangkan proporsi kelas pada variabel target dengan membuat sampel sintetis dari kelas minoritas (*defaulted*). Ini menghasilkan dataset seimbang yang siap untuk pelatihan model machine learning, meningkatkan kemampuan model dalam memprediksi kelayakan pinjaman dengan lebih akurat dan konsisten.

2.4.5. Pembagian Data Latih dan Data Uji

Pembagian data menjadi data latih dan uji menggunakan `train_test_split` dari scikit-learn adalah langkah penting dalam pengembangan sistem rekomendasi kelayakan pinjaman online. Setelah SMOTE menyeimbangkan kelas, data dibagi menjadi 80% latih dan 20% uji dengan `test_size=0.2` dan `random_state=42` untuk reproduksi hasil. Model dilatih pada data latih dan diuji pada data uji untuk memastikan generalisasi dan keakuratan prediksi kelayakan pinjaman saat diimplementasikan di dunia nyata.

2.4.6. Penskalaan Fitur menggunakan *Standard Scaler*

Penskalaan fitur dengan `StandardScaler` adalah langkah krusial dalam preprocessing data untuk sistem rekomendasi kelayakan pinjaman online. Fitur numerik seperti usia, pendapatan, jumlah pinjaman, skor kredit, masa kerja, jumlah garis kredit, suku bunga, jangka waktu pinjaman, dan rasio hutang terhadap pendapatan diubah skala untuk memastikan mean nol dan variansi satu. Proses ini menggunakan `fit_transform` pada data latih dan `transform` pada data uji untuk konsistensi. Penskalaan ini memastikan model machine learning dapat mempelajari pola data secara konsisten dan menghasilkan prediksi yang dapat diandalkan, mendukung pengembangan sistem rekomendasi yang akurat untuk evaluasi kelayakan pinjaman.

Preprocessing data dalam sistem rekomendasi kelayakan pinjaman online meliputi konversi mata uang ke IDR, *encoding* fitur kategorikal menjadi numerik, penghapusan fitur tidak relevan seperti *LoanID*, penanganan ketidakseimbangan data dengan SMOTE, dan pembagian data menjadi latih dan uji. Penskalaan fitur menggunakan *StandardScaler* diterapkan untuk konsistensi skala numerik sebelum melatih model seperti *Random Forest* atau *XGBoost*. Setelah preprocessing, model dilatih, parameter disetel, dan performa dievaluasi menggunakan data uji. Proses ini bertujuan menghasilkan sistem rekomendasi yang akurat dan mendukung keputusan pinjaman online.

2.5. Pemodelan

Setelah proses preprocessing data, langkah selanjutnya adalah membangun dan mengevaluasi tiga model untuk sistem rekomendasi kelayakan pinjaman *online*: *Decision Tree*, *Random Forest*, dan *XGBoost*.

Pertama, *Decision Tree*, model ini disetel dengan parameter terbaik yang telah didapatkan dari hasil *GridSearch* sebelumnya, seperti *min_samples_split*, *min_samples_leaf*, *max_features*, *max_depth*, *criterion*, dan *class_weight*. Setelah dilatih dengan data latih yang diproses, model dievaluasi menggunakan data uji. Hasil evaluasi, termasuk *precision*, *recall*, dan *f1-score*, dianalisis dalam *classification_report* untuk mengevaluasi kinerja model ini.

Kedua, *Random Forest*, model *Random Forest* digunakan tanpa penyetelan tambahan setelah proses *GridSearch* sebelumnya. Seperti *Decision Tree*, model ini dilatih dengan data latih yang sama dan dievaluasi dengan menggunakan data uji. Matrik evaluasi seperti *precision*, *recall*, dan *f1-score* dievaluasi dalam *classification_report* untuk memahami kinerja model memprediksi kelayakan pinjaman.

Ketiga, *XGBoost*, model *XGBoost* menggunakan parameter terbaik yang telah dipilih dari hasil *GridSearch* sebelumnya, seperti *n_estimators*, *learning_rate*, *max_depth*, *min_samples_split*, *min_samples_leaf*, *subsample*, dan *random_state*. Setelah dilatih dengan data latih yang telah diproses, model dievaluasi menggunakan data uji. Hasil evaluasi kinerja model juga ditampilkan dalam *classification_report* untuk memberikan wawasan mendalam tentang kemampuan prediktifnya.

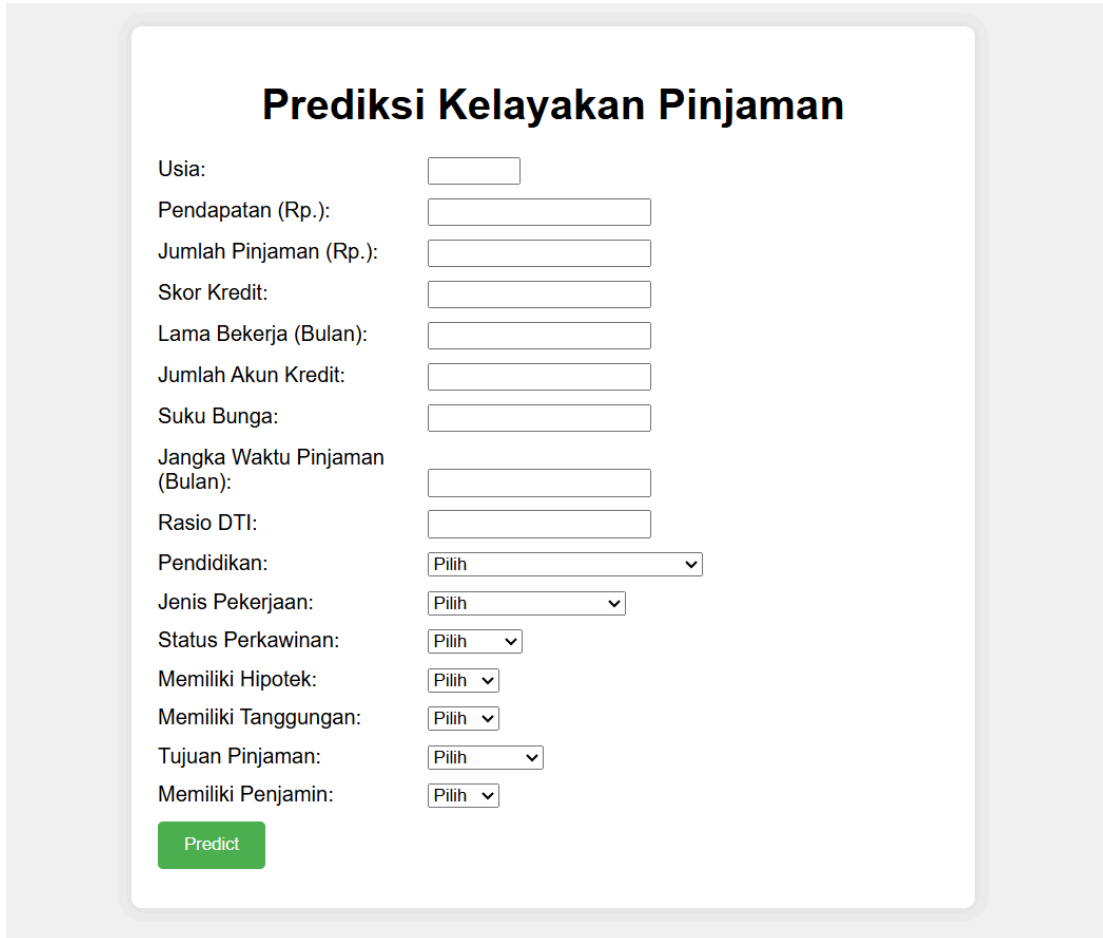
Proses evaluasi ini bertujuan untuk membandingkan kinerja ketiga model dalam konteks rekomendasi kelayakan pinjaman. Hasil evaluasi akan membantu dalam memilih model yang paling sesuai untuk implementasi di platform *online*, sehingga dapat memberikan rekomendasi kelayakan pinjaman yang akurat kepada pengguna.

Tabel 1. Hasil Evaluasi Model

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	0.83	0.85	0.81	0.82
Random Forest	0.88	0.90	0.85	0.88
XGBoost	0.88	0.91	0.83	0.87

Dalam evaluasi kinerja model untuk rekomendasi kelayakan pinjaman, fokus utama adalah pada kelas minoritas *Default 1* (gagal bayar). *Precision* mengukur akurasi prediksi positif, *recall* mengukur kemampuan model mengidentifikasi kasus sebenarnya yang positif, dan *F1-score* memberikan gambaran keseluruhan tentang kinerja model dalam mengklasifikasikan kelas minoritas. Hasil evaluasi ini menjadi dasar untuk memilih model terbaik untuk implementasi di platform *online* guna memberikan rekomendasi kelayakan pinjaman yang akurat kepada pengguna, dengan model-model yang disimpan dalam format *.pkl* untuk penggunaan masa depan dalam pengembangan produk *website*.

2.6. Produk Akhir (*Website*)



The screenshot displays a web form titled "Prediksi Kelayakan Pinjaman" (Loan Eligibility Prediction). The form is set against a light gray background with a white central panel. It contains 15 input fields: 11 text boxes for numerical or categorical data (Usia, Pendapatan (Rp.), Jumlah Pinjaman (Rp.), Skor Kredit, Lama Bekerja (Bulan), Jumlah Akun Kredit, Suku Bunga, Jangka Waktu Pinjaman (Bulan), Rasio DTI, Pendidikan, Jenis Pekerjaan, Status Perkawinan, Memiliki Hipotek, Memiliki Tanggungan, Tujuan Pinjaman, and Memiliki Penjamin) and 4 dropdown menus (Pendidikan, Jenis Pekerjaan, Status Perkawinan, and Tujuan Pinjaman). A green "Predict" button is located at the bottom left of the form area.

Gambar 13. Tampilan *Website* FinPredictor.

Website ini sebagai produk akhir memanfaatkan model *pickle* yang telah dihasilkan dari pengolahan dan pemrosesan data untuk melakukan prediksi kelayakan pinjaman secara *real-time*. Pengguna dapat mengisi formulir dengan data seperti usia, pendapatan, dan lainnya, dan kemudian mendapatkan hasil prediksi langsung di halaman web. Model *XGBoost* yang digunakan telah dilatih sebelumnya untuk memberikan rekomendasi apakah pelanggan layak atau tidak layak mendapatkan pinjaman berdasarkan karakteristik yang dimasukkan. Berikut adalah video demo penggunaan website: bit.ly/DemoFinPredictor.

3. Penutup

Dari hasil analisis EDA, dataset kelayakan pinjaman online menunjukkan variasi yang signifikan di berbagai variabel seperti usia, pendapatan, jumlah pinjaman, dan skor kredit. Proses preprocessing yang meliputi konversi mata uang, *encoding* fitur kategorikal, dan penanganan ketidakseimbangan data menggunakan SMOTE, telah meningkatkan kualitas data. Evaluasi terhadap tiga model *machine learning* menunjukkan bahwa *Random Forest* dan *XGBoost* masing-masing memiliki akurasi sekitar 0.88, sedangkan Decision Tree memiliki akurasi 0.83. Namun, untuk kelas minoritas (*Default 1*), *Random Forest* menunjukkan *precision*, *recall*, dan *F1-score* yang lebih tinggi dibandingkan model lainnya.

Secara praktis, hasil ini mengindikasikan bahwa implementasi *Random Forest* atau *XGBoost* dapat memberikan rekomendasi kelayakan pinjaman *online* yang lebih akurat, mengurangi risiko gagal bayar, dan meningkatkan pengalaman pengguna. Secara teoritis, penelitian ini menegaskan pentingnya EDA yang komprehensif dan *pre-processing* yang cermat dalam membangun model yang efektif. Namun, keterbatasan dataset seperti jumlah data yang terbatas dan ketidakseimbangan kelas memerlukan penelitian lanjutan dengan fokus pada pengumpulan data yang lebih luas, penggunaan metode *ensemble*, penanganan ketidakseimbangan data yang lebih maju, dan penyetelan parameter yang lebih mendalam untuk meningkatkan kinerja model dan mengurangi *overfitting*.

Langkah-langkah ini akan mendukung pengembangan sistem rekomendasi kelayakan pinjaman online yang lebih robust dan efisien, menghadapi berbagai tantangan dalam industri finansial *modern*.

4. Daftar Pustaka

Annur, C. M. (2023, Oktober 17). Penyaluran Pinjaman Online Meningkat pada Agustus 2023. databoks. Retrieved Juni 17, 2024, from <https://databoks.katadata.co.id/datapublish/2023/10/17/penyaluran-pinjaman-online-meningkat-pada-agustus-2023>

- Arvante, J. Z. Y. (2022, Januari 13). Dampak Permasalahan Pinjaman Online dan Perlindungan Hukum Bagi Konsumen Pinjaman Online. *Ikatan Penulis Mahasiswa Hukum Indonesia Law Journal*, 2(1), 74-86. <https://doi.org/10.15294/ipmhi.v2i1.53736>
- Badan Pusat Statistik. (2021, September 17). LAPORAN PEREKONOMIAN INDONESIA 2021. Badan Pusat Statistik. Retrieved Juni 18, 2024, from <https://www.bps.go.id/id/publication/2021/09/17/f3ece7157064514772b18335/laporan-perekonomian-indonesia-2021.html>
- Gomber, P., Kauffman, R. J., Parker, C., & Weber, B. W. (2017). On the Fintech Revolution: Interpreting the Forces of Innovation, Disruption, and Transformation in Financial Services. *Journal of Management Information Systems*, 35(1), 220-265.
- Gunnarsson, B. R., vanden Broucke, S., Baesens, B., Óskarsdóttir, M., & Lemahieu, W. (2021). Deep learning for credit scoring: do or don't?. *European Journal of Operational Research*, 295(1), 292–305. <https://doi.org/10.1016/j.ejor.2021.03.006>
- Hjelkrem, L. O., de Lange, P., & Nettet, E. (2022). An end-to-end deep learning approach to credit scoring using CNN + XGBoost on transaction data. *Journal of Risk Model Validation*, 16(2), 1-25. <https://doi.org/10.21314/JRMV.2022.018>
- Hjelkrem, L. O., et al. (2022). The Value of Open Banking Data for Application Credit Scoring: Case Study of a Norwegian Bank. *Journal of Risk and Financial Management*. 15(12), 597. <https://doi.org/10.3390/jrfm15120597>
- Kriebel, J., & Stitz, L. (2022). Credit default prediction from user-generated text in peer-to-peer lending using deep learning. *European Journal of Operational Research*, 302(1), 309-323. <https://doi.org/10.1016/j.ejor.2021.12.024>
- Nikhil. (2024, September). Loan Default Prediction Dataset, Version 1. Retrieved Mei 19, 2024 from <https://www.kaggle.com/datasets/nikhille9/loan-default>.
- Rahmadyanto, B. P. (2023, Februari 2). TREN PINJAMAN ONLINE DALAM MILENIAL: TELAHAH KONTRIBUTOR INTERNAL DAN EKSTERNAL.

JOURNAL OF DEVELOPMENT ECONOMIC AND SOCIAL STUDIES, 2(2), 249-258. <https://dx.doi.org/10.21776/jdess.2023.02.2.02>

Schmitt, M., et al. (2022). Deep Learning vs. Gradient Boosting: Benchmarking state-of-the-art machine learning algorithms for credit scoring. *International Journal of Information Management*. arXiv:2205.10535.

Sidik, B. (2021, Juli 4). Pinjaman "Online" Kian Menjadi Alternatif. Kompas. <https://www.kompas.id/baca/metro/2021/07/04/pinjaman-online-kian-menjadi-alternatif>