

Prediksi Limit Kredit Menggunakan Metode Regresi Linear

Ahmad Yusuf, Henny Leidiyana, Imam Budiawan

Program Studi Informatika Kampus Kaliabang, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

ARTICLE INFO

Article history:

Received August 20, 2025

Revised 25 August 2025

Accepted 25 August 2025

Available online 17 Sept. 2025

Keywords:

Regresi Linier, Limit Kredit, Google Colab, Prediksi

Keywords:

Linear Regression, Credit Limit, Google Colab, Prediction



This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.
Copyright © 2025 by Author. Published by Yayasan Daarul Huda

ABSTRACT

Penentuan limit kredit yang tepat merupakan hal penting bagi lembaga keuangan dalam mengelola risiko kredit sekaligus mengoptimalkan pendapatan. Penelitian ini bertujuan untuk membangun model prediksi limit kredit menggunakan metode regresi linear dengan fitur-fitur utama seperti Rating, Income, dan Balance. Dataset yang digunakan terdiri dari 400 data nasabah kartu kredit yang memiliki 11 variabel berupa data numerik dan kategorikal. Tahapan penelitian mengikuti metodologi Cross-Industry Standard Process for Data Mining (CRISP-DM), yang mencakup tahap pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan deployment. Proses analisis data menggunakan Aplikasi Google Colab. pemeriksaan kualitas data, pengolahan fitur kategorikal dengan metode label encoding, serta normalisasi data menggunakan MinMaxScaler. Hasil analisis korelasi menunjukkan bahwa variabel Rating, Income, dan Balance memiliki korelasi yang sangat kuat dengan variabel Limit Kredit, sehingga ketiga variabel ini digunakan sebagai prediktor utama dalam pemodelan.

ABSTRACT

Determining appropriate credit limits is essential for financial institutions to manage credit risk effectively while optimizing revenue. This study aims to develop a predictive model for credit limits using linear regression, incorporating primary features such as Rating, Income, and Balance. The dataset consists of 400 credit card customer records with 11 variables, comprising both numerical and categorical data. The research follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, covering stages including business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Data analysis was conducted using Google Colab, involving quality assessment, categorical feature encoding through label encoding, and data normalization utilizing MinMaxScaler. Correlation analysis results indicated that Rating, Income, and Balance have strong correlations with Credit Limit, hence these three variables were chosen as primary predictors for the modeling process.

PENDAHULUAN

Bank adalah organisasi internasional yang bertanggung jawab untuk menerima simpanan dari pelanggan mereka atau yang dikenal dengan nasabah dan meminjamkannya kepada pelanggan atau nasabah lain yang sedang membutuhkan dana pinjaman. Dalam hal simpanan masyarakat, maka bank memberikan pinjaman atau kredit dalam bentuk bunga. (Indah Sucianty & Suria Manda, 2022)

Kredit adalah pemberian kredit yang didasarkan atas keyakinan bank akan kemampuan dan kesanggupan nasabah untuk membayar utangnya. (Bambang Riono, 2022). Pemberian fasilitas kredit salah satu layanan utama lembaga keuangan, termasuk perbankan dan perusahaan pembiayaan. Penetapan limit kredit yang tepat memiliki peran penting dalam mengelola risiko gagal bayar sekaligus memaksimalkan pendapatan bunga. Limit risiko kredit adalah batas atau kuota yang ditentukan oleh perbankan untuk membatasi jumlah kredit yang dapat diberikan kepada satu nasabah atau grup nasabah. Ini dapat digunakan untuk memastikan bahwa perbankan tidak terlalu terpapar pada risiko dari satu nasabah atau grup nasabah. (Rohimah et al., 2023). Jika limit kredit terlalu rendah, potensi pendapatan perusahaan menjadi terbatas; sebaliknya, jika limit terlalu tinggi, risiko kredit macet dapat meningkat.

Perkembangan kecerdasan buatan dan *machine learning* membuka peluang bagi perusahaan pembiayaan untuk menerapkan model prediktif dalam penentuan limit kredit. Metode *Regresi Linier* merupakan salah satu teknik sederhana namun efektif untuk memodelkan hubungan antara karakteristik nasabah seperti pendapatan bulanan, *rating* SLIK, dan riwayat gagal bayar dengan limit kredit yang optimal.

Di Indonesia, pencatatan riwayat kredit nasabah dilakukan melalui sistem SLIK (Sistem Layanan Informasi Keuangan) yang dioperasikan oleh Otoritas Jasa Keuangan (OJK). Data SLIK memuat informasi status pembayaran, jumlah pinjaman, dan nilai tunggakan, yang menjadi dasar perhitungan *credit score* oleh berbagai lembaga keuangan. proses *BI Checking* yang menggunakan Sistem Informasi Debitor (*SID*) sudah digantikan dengan Sistem Layanan Informasi Keuangan (SLIK) yang dikelola oleh lembaga Otoritas Jasa Keuangan (OJK). Sistem Layanan Informasi Keuangan Otoritas Jasa Keuangan (SLIK OJK) yang dulunya dikenal dengan nama *BI Checking/SID* adalah sistem informasi yang pengelolaannya bertujuan untuk melaksanakan tugas pengawasan dan pelayanan informasi keuangan, salah satunya berupa penyediaan informasi debitor. SLIK. (Simanjuntak, 2023)

METODE PENELITIAN

Penulis melakukan proses melakukan metode CRISP-DM, Dalam tahap CRISP-DM ini diawali pada *Business Understanding*, proses dimulai dengan memahami tujuan bisnis dan merumuskan masalah yang dapat diselesaikan dengan data mining. Selanjutnya, tahap *Data Understanding* dilakukan dengan mengumpulkan, mengeksplorasi, dan mengevaluasi data yang akan digunakan. Setelah itu, tahap *Data Preparation* melibatkan proses pembersihan, seleksi, dan transformasi data agar siap untuk analisis lebih lanjut. Pada tahap *Modeling*, algoritma data mining diterapkan untuk menemukan pola dalam data, seperti teknik *clustering* atau *classification*. Setelah model dibangun, tahap *Evaluation* dilakukan untuk mengukur performa model guna memastikan hasilnya sesuai dengan tujuan awal. Terakhir, tahap *Deployment* adalah proses implementasi hasil analisis dalam bentuk laporan atau sistem yang dapat digunakan untuk pengambilan keputusan (Rawan et al., 2025)

Business Understanding

Pada tahap ini, tujuan bisnis utama adalah membangun model prediktif untuk menentukan limit kredit nasabah berdasarkan karakteristik demografis dan historis. Permasalahan utamanya adalah ketergantungan bank pada metode konvensional yang subjektif dalam menetapkan limit kredit, yang berpotensi menyebabkan kerugian akibat limit terlalu longgar atau kehilangan nasabah karena limit terlalu ketat. Solusi yang diusulkan adalah mengembangkan model regresi linear yang memanfaatkan variabel-variabel prediktif seperti pendapatan, rating kredit, jumlah kartu, usia, dan saldo.

Data Understanding

Dataset: 400 sampel nasabah dengan 11 fitur, termasuk *Income* (pendapatan), *Rating* (skor kredit), *Cards* (jumlah kartu kredit), *Age*, *Education*, *Balance*, serta variabel kategorikal seperti *Gender*, *Student*, *Married*, dan *Ethnicity*.

Data Preparation

Tahap ini melibatkan transformasi dataset menjadi format yang siap dimodelkan. Variabel kategorikal seperti *Gender*, *Student*, *Married*, *Ethnicity* di ubah menjadi numerik dengan metode *Label Encoding* dan Normalisasi data *MinMaxScaler* pada dataset tersebut,

Modeling

Pada tahap Modeling dalam kerangka CRISP-DM, pertama-tama diputuskan bahwa *regresi linear* berganda menjadi metode utama karena hubungan linier yang kuat antara beberapa fitur finansial (seperti *Rating*, *Income*, dan *Balance*) dengan variabel target Limit Kredit.

Evaluation

Pada tahap Evaluation dalam metodologi CRISP-DM, fokus utama adalah menilai seberapa baik model regresi linear yang telah dibangun memenuhi tujuan bisnis dan kriteria keberhasilan yang telah ditetapkan. Pertama-tama, model diuji pada data yang belum pernah dilihat sebelumnya (data uji) untuk mengukur kemampuan generalisasinya. Metrik yang digunakan meliputi *Mean Squared Error (MSE)*, yang menunjukkan rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual, serta *R-squared (R^2)*, yang menunjukkan proporsi variansi variabel target (limit kredit) yang berhasil dijelaskan oleh model. Dengan MSE rendah dan R^2 mendekati 1, model regresi linear dinyatakan cukup akurat dalam memprediksi limit kredit sesuai target awal (misalnya $R^2 \geq 0,90$).

Deployment

Deployment adalah tahapan di mana model atau solusi *data mining* yang telah dibangun dan divalidasi dipindahkan dari lingkungan eksperimen (*development*) ke lingkungan operasional (*production*) sehingga bisa digunakan secara nyata untuk mendukung keputusan bisnis.

Namun Penulis Hanya melakukan pengujian Evaluasi tidak sampai *Deployment*

Metode Pengolahan dan Analisis Data

Regresi Linear

Regresi Linear atau merupakan metode statistik yang bertujuan untuk membentuk sebuah model antara variabel dependen (Y) dengan variabel independen (X). *Regresi Linear* yang memiliki satu variabel

bebas disebut dengan Regresi Linear Sederhana, sedangkan Regresi Berganda diperuntukkan apabila memiliki lebih dari satu variabel bebas. Regresi Linear menggunakan garis kecenderungan apabila pola data menunjukkan suatu kecenderungan, baik berpola turun atau naik. *Regresi Linear* merupakan bagian regresi yang mencakup hubungan linear satu peubah tak bebas Y dengan satu peubah bebas X. Analisis regresi adalah metode untuk menentukan hubungan sebab-akibat antara satu variabel dengan variabel-variabel yang lain.

$$\hat{Y} = a + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

Dimana:

$\hat{Y}$: nilai variabel Y hasil prediksi

Y : variabel tak bebas

X : variabel bebas

a : konstanta regresi

b : koefisien regresi

Sumber : Galton dan Pearson

Variabel “penyebab” disebut dengan bermacam-macam istilah yaitu, variabel penjelas, variabel eksplanatorik, variabel independen, atau secara bebas, variabel X (karena seringkali digambarkan dalam grafik sebagai absis atau sumbu X). Variabel terkena akibat dikenal sebagai variabel yang dipengaruhi, variabel dependen, variabel terikat, atau variabel Y. Kedua variabel ini dapat merupakan variabel acak (*random*), namun variabel yang dipengaruhi harus selalu variabel acak. (Panggabean et al., 2020)

Data Set

Dataset (<https://www.kaggle.com/datasets/syedasimalishah/credit-card-limit-prediction>), didapatkan pada Kaggle Dataset: 400 sampel nasabah dengan 11 fitur, termasuk *Income* (pendapatan), *Rating* (skor kredit), *Cards* (jumlah kartu kredit), *Age*, *Education*, *Balance*, serta variabel kategorikal seperti *Gender*, *Student*, *Married*, dan *Ethnicity*. Analisis awal berupa ringkasan statistik (*mean*, *median*, *range*) dan visualisasi distribusi masing-masing fitur mengungkapkan bahwa fitur numerik relatif bersih tanpa missing value, namun beberapa variabel seperti *Balance* dan *Income* memiliki sebaran yang *skewed*.

Analisis Data

Dataset ini berisi 400 entri data terkait pemegang kartu kredit dengan 11 variabel yang menggambarkan profil demografi, penggunaan kartu kredit, dan kondisi keuangan.

Kolom	Tipe	Deskripsi
Income	float64	Pendapatan tahunan (dalam ribuan atau jutaan, tergantung unit).
Limit	int64	Limit kredit yang diberikan (target prediksi).
Rating	int64	Skor kredit resmi atau dihasilkan (biasanya 200–850).
Cards	int64	Jumlah kartu kredit yang dimiliki nasabah.
Age	int64	Usia nasabah (dalam tahun).
Education	int64	Tahun pendidikan formal atau level pendidikan (dalam skala tertentu, misalnya 11–20).
Gender	object	Jenis kelamin ("Male"/"Female").
Student	object	Status mahasiswa ("Yes"/"No").
Married	object	Status pernikahan ("Yes"/"No").
Ethnicity	object	Etnisitas nasabah ("Caucasian", "Asian", "African American").
Balance	int64	Saldo rata-rata kartu kredit (jumlah tagihan yang belum dibayar).

Gambar 2 Dataset

Income: rata-rata sekitar 100 (unit), dengan rentang dari 15 hingga 190.

Limit: rata-rata 6200, berkisar dari 1000 hingga 15000.

Rating: rata-rata 500, rentang 250–850.

Cards: mayoritas nasabah memiliki 1–6 kartu.

Age: tersebar antara 20–80 tahun, rata-rata 45 tahun.

Education: berkisar antara 9–20, rata-rata 14.

Balance: nilai rata-rata saldo sekitar 400, dengan rentang 0–2000.

1. Skewed Fitur

Income, *Limit*, dan *Balance* cenderung miring kanan (*skewed*), artinya sebagian kecil nasabah memiliki pendapatan, limit, atau saldo sangat tinggi dibanding mayoritas.

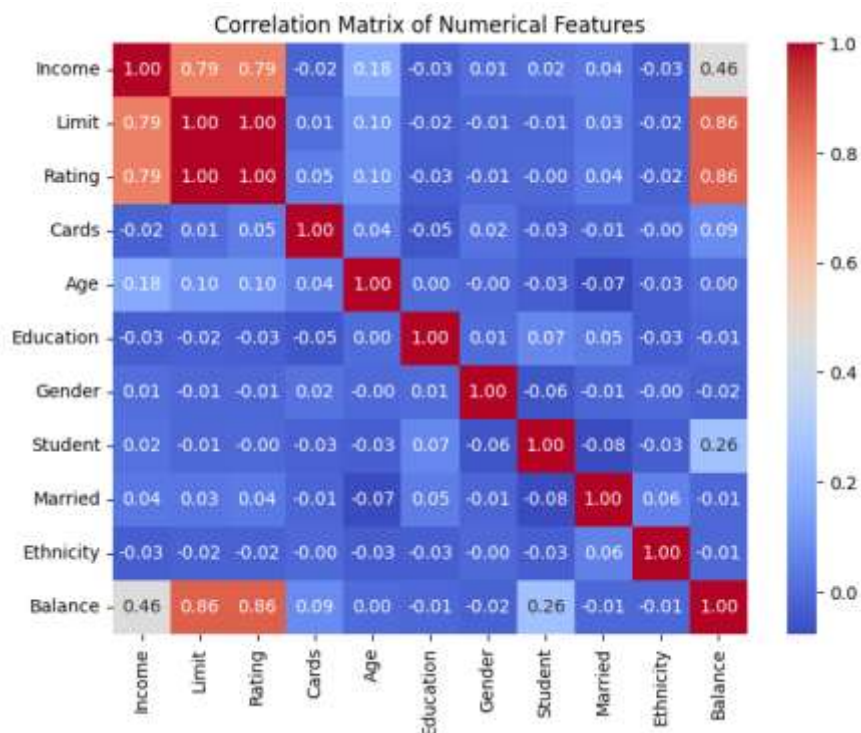
2. Variasi Demografis
Proporsi *Gender*, *Student*, dan *Married* relatif seimbang, sehingga tidak ada dominasi ekstrem satu kategori.
3. Kesenambungan Fitur Finansial
Rating dan *Balance* biasanya naik seiring *Income* nasabah dengan pendapatan tinggi cenderung mendapatkan skor kredit lebih baik dan saldo tagihan lebih besar.
Mengubah variabel kategorikal menjadi numerik (*Label Encoding*). Menilai korelasi antar fitur.
Membangun model regresi linier untuk memprediksi Limit dan mengevaluasinya dengan metrik RMSE dan R^2 .

HASIL DAN PEMBAHASAN

Hasil Penelitian

1. Data Set

Dataset yang digunakan berisi 400 sampel nasabah kartu kredit dengan 11 kolom yang mencakup berbagai aspek finansial dan demografis. Kolom *Income* menunjukkan pendapatan tahunan nasabah (dalam ribuan), sedangkan Limit adalah batas kredit yang diberikan dan berfungsi sebagai target yang ingin diprediksi. *Rating* menggambarkan skor kredit masing-masing nasabah, dan *Balance* adalah rata-rata saldo tagihan kartu kredit. Fitur *Cards*, *Age*, dan *Education* mencerminkan jumlah kartu yang dimiliki, usia, serta tingkat pendidikan nasabah. Empat kolom kategorikal *Gender*, *Student*, *Married*, dan *Ethnicity* memberi informasi jenis kelamin, status mahasiswa, status pernikahan, serta latar belakang etnis.



Gambar 3 Dataset

Seluruh kolom numerik tidak memiliki nilai kosong, dan sebarannya bervariasi; misalnya *Income* dan *Balance* cenderung memiliki rentang yang luas dan sedikit miring ke kanan. Fitur kategorikal perlu diubah ke format numerik (*Label Encoding*) sebelum pemodelan.

2. Kolerasi Antar Fitur

Matriks korelasi fitur numerik dari dataset yang digunakan untuk prediksi limit kredit. Korelasi menunjukkan seberapa erat hubungan linear antar variabel.

```

Korelasi Fitur dengan Variabel Target ('Limit'):
Limit      1.000000
Rating     0.996880
Balance    0.861697
Income     0.792088
Age        0.100888
Married    0.031155
Cards      0.010231
Student    -0.006015
Gender     -0.009397
Ethnicity  -0.020837
Education  -0.023549
Name: Limit, dtype: float64

Fitur dengan Korelasi Absolut > 0.5 dengan 'Limit':
Limit
Rating 0.996880
Balance 0.861697
Income 0.792088

dtype: float64

```

Gambar 3 Kolerasi Matriks

	Income	Limit	Rating	Cards	Age	Education	Gender	Student	Married	Ethnicity	Balance
0	14.891	3606	283	2	34	11	Male	No	Yes	Caucasian	333
1	106.025	6645	483	3	82	15	Female	Yes	Yes	Asian	903
2	104.593	7075	514	4	71	11	Male	No	No	Asian	580
3	148.924	9504	681	3	36	11	Female	No	No	Asian	964
4	55.882	4897	357	2	68	16	Male	No	Yes	Caucasian	331
...	...	...	...	...	...	...	...	...	...	...	...
395	12.096	4100	307	3	32	13	Male	No	Yes	Caucasian	560
396	13.364	3838	296	5	65	17	Male	No	No	African American	480
397	57.872	4171	321	5	67	12	Female	No	Yes	Caucasian	138
398	37.728	2525	192	1	44	13	Male	No	Yes	Caucasian	0
399	18.701	5524	415	5	64	7	Female	No	No	Asian	966

400 rows x 11 columns

Gambar 5 Kolerasi antar Variabel

- Limit memiliki korelasi positif yang sangat kuat dengan *Rating* (0.9968) dan *Balance* (0.8616), serta korelasi kuat dengan *Income* (0.7920). Hal ini berarti semakin tinggi *rating*, saldo, dan penghasilan seseorang, maka semakin tinggi pula limit kredit yang diberikan.
- Income* juga menunjukkan korelasi tinggi dengan *Rating* (0.79) dan cukup tinggi, mengindikasikan bahwa semakin tinggi pendapatan seseorang, cenderung semakin baik pula skor kredit dan saldo tagihan mereka.
- Fitur seperti *Cards*, *Age*, *Education*, *Gender*, *Student*, *Married*, dan *Ethnicity* memiliki nilai korelasi yang rendah terhadap limit kredit, yang artinya pengaruh fitur tersebut secara linier terhadap penentuan limit kredit tidak signifikan.

3. Periksa Missing Value

Hasil analisa pada gambar tersebut menunjukkan bahwa tidak terdapat *missing value* (nilai yang hilang) di seluruh kolom pada dataset yang digunakan. Hal ini ditunjukkan oleh hasil pengecekan `df.isnull().sum()` yang semuanya bernilai 0 pada setiap fitur, seperti *Income*, *Limit*, *Rating*, *Cards*, *Age*, *Education*, *Gender*, *Student*, *Married*, *Ethnicity*, dan *Balance*.

```
# Cek missing values
print(df.isnull().sum())
```

```
Income      0
Limit       0
Rating      0
Cards       0
Age         0
Education   0
Gender      0
Student     0
Married     0
Ethnicity   0
Balance     0
dtype: int64
```

Gambar 6 Missing Value

Kondisi data yang lengkap seperti ini sangat baik karena proses analisis dan pembuatan model prediksi dapat dilakukan secara optimal tanpa perlu penanganan khusus terhadap data yang hilang, sehingga hasil analisis akan lebih valid dan akurat.

4. Cek Duplikasi Pada Dataset

Pengecekan data duplikat pada dataset menggunakan perintah `df.duplicated().sum()`. Hasil yang muncul adalah 0,

```
# Cek duplikasi
print(df.duplicated().sum())
```

```
0
```

Gambar 7 Cek Duplikasi pada Dataset

yang berarti tidak ada baris data yang duplikat di dalam dataset. Ini merupakan kondisi yang sangat baik, karena data yang bebas dari duplikasi akan memastikan hasil analisis dan model prediksi yang lebih akurat, serta menghindari bias akibat data yang berulang.

5. Label Encoding Pada Data

Label Encoding pada beberapa kolom kategorikal, yaitu Gender, Student, Married, dan Ethnicity.

	Income	Limit	Rating	Cards	Age	Education	Gender	Student	Married	\
0	14.891	3606	283	2	34	11	1	0	1	
1	106.025	6645	483	3	82	15	0	1	1	
2	104.593	7075	514	4	71	11	1	0	0	
3	148.924	9504	681	3	36	11	0	0	0	
4	55.882	4897	357	2	68	16	1	0	1	

	Ethnicity	Balance
0	2	333
1	1	903
2	1	580
3	1	964
4	2	331

Gambar 8 Label Encoding pada Data

Proses ini dilakukan menggunakan *LabelEncoder* dari pustaka *scikit-learn*, sehingga setiap nilai kategori diubah menjadi angka.

6. Normalisasi Data

Penerapan *MinMaxScaler* pada seluruh kolom dataset baik variabel asli seperti *Income*, *Limit*, *Rating*, *Cards*, *Age*, *Education*, *Balance*, maupun kolom hasil *label encoding* seperti *Gender*, *Student*, *Married*, dan *Ethnicity* di mana setiap nilai diubah agar berada dalam rentang 0 hingga 1.

	Income	Limit	Rating	Cards	Age	Education	Gender	Student	\
0	0.025737	0.210675	0.213723	0.125	0.146667	0.400000	1.0	0.0	
1	0.542722	0.443406	0.438695	0.250	0.786667	0.666667	0.0	1.0	
2	0.534598	0.476336	0.473566	0.375	0.640000	0.400000	1.0	0.0	
3	0.786079	0.662353	0.661417	0.250	0.173333	0.400000	0.0	0.0	
4	0.258271	0.309542	0.296963	0.125	0.600000	0.733333	1.0	0.0	

	Married	Ethnicity	Balance
0	1.0	1.0	0.166583
1	1.0	0.5	0.451726
2	0.0	0.5	0.290145
3	0.0	0.5	0.482241
4	1.0	1.0	0.165583



Gambar 9 Normalisasi Data

Dari keluaran *describe()* terlihat bahwa nilai minimum untuk semua kolom adalah 0 dan nilai maksimum adalah 1, sedangkan nilai *mean* dan standar deviasi mencerminkan distribusi relatif setiap fitur setelah diskalakan. Normalisasi semacam ini penting agar variabel dengan skala besar (misalnya *Income* dan *Balance*) tidak mendominasi variabel lain dalam proses pelatihan model regresi linear, sehingga setiap fitur dapat berkontribusi secara seimbang terhadap prediksi limit kredit

7. Regresi Linear

Regresi Linear menggunakan seluruh fitur (X_{train}) untuk memprediksi limit kredit (y_{train}).

```
Mean Squared Error: 0.00016175912035471681
R-squared: 0.9923846114340685
Coefficients: [0.96958677 0.02126985 0.0213184 ]
Intercept: 0.0014799141115311998
```

Gambar 10 Regresi Linear

$$Y = 0.01479 + 0.96958677 X_1 + 0.02126985 X_2 + 0.0213184 X_3$$

Model tersebut untuk memprediksi nilai limit pada data uji (X_{test}) dan mengevaluasi kinerjanya dengan dua metrik: *Mean Squared Error (MSE)* sebesar 0.0001617591203547172, yang menunjukkan bahwa rata-rata kuadrat selisih antara prediksi dan nilai aktual sangat kecil (artinya prediksi sangat akurat secara rata-rata), serta *R-squared* (R^2) sebesar 0.9923846114340684 (99.23%), yang berarti 99.23% variasi limit kredit dalam data uji berhasil dijelaskan oleh model regresi linear ini. Nilai R^2 yang mendekati 1 dan MSE yang nyaris mendekati 0 mengindikasikan bahwa model sangat baik dalam menangkap hubungan linier antara fitur-fitur yang digunakan dengan target prediksi limit kredit. Hasil Intercept 0.0014799141115311998 dan Hasil pada Koefisien Rating sebesar 0.96958677, Income sebesar 0.02126985, Balance sebesar 0.0213184.

8. Visualisasi Hubungan Rating dan Limit

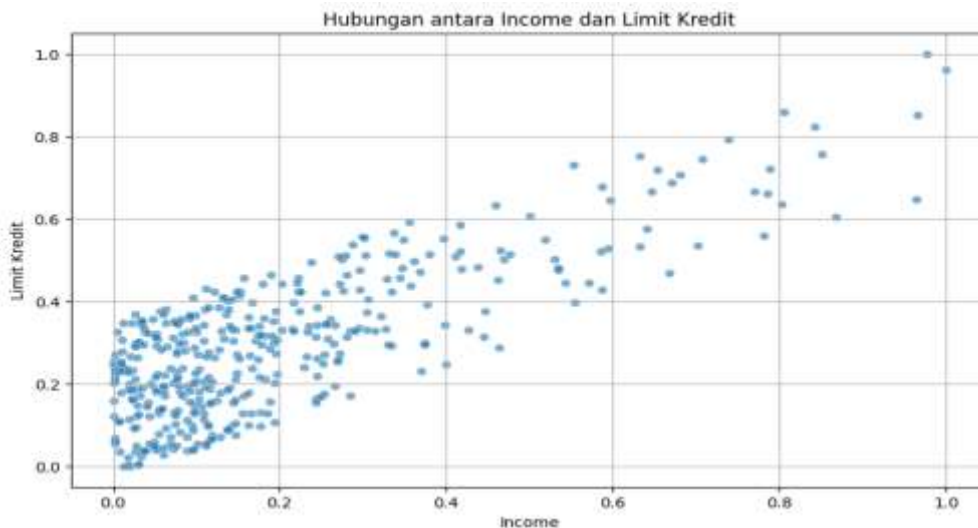
Scatter plot antara variabel *Rating* (sumbu X) dan Limit Kredit (sumbu Y) setelah keduanya dinormalisasi ke rentang 0–1. Setiap titik mewakili satu nasabah. Pola sebaran titik membentuk garis diagonal yang sangat rapat dan hampir lurus, menunjukkan hubungan linier positif yang sangat kuat: semakin tinggi nilai *Rating*, semakin tinggi pula Limit Kredit yang diberikan. Kerapatan titik di dekat garis *imajiner* ($y = x \times \text{skala linier}$) memperlihatkan bahwa variasi Limit Kredit dapat dijelaskan secara dominan oleh perubahan *Rating*. Hampir tidak terlihat penyimpangan sistematis atau lengkungan (*curvature*), sehingga asumsi linearitas untuk *regresi linear* pada fitur ini sangat mendukung.

Gambar I1 Hubungan antara Rating dan Limit

Selain itu, sebaran titik relatif homogen di sepanjang rentang menengah (sekitar *Rating* 0,2–0,6) dan hanya beberapa titik di ujung atas (*Rating* > 0,8) yang berfungsi sebagai pengamatan bernilai tinggi (*high leverage*). Hal ini mengindikasikan bahwa model yang menggunakan *Rating* sebagai salah satu prediktor utama akan stabil. Visual ini juga memperkuat hasil analisis korelasi sebelumnya yang menempatkan *Rating* sebagai fitur dengan korelasi tertinggi terhadap Limit Kredit.

9. Hubungan antara *Income* dan Limit Kredit

Grafik *scatter plot* ini menampilkan hubungan antara *Income* (pendapatan nasabah) pada sumbu X dan Limit Kredit pada sumbu Y. Keduanya telah dinormalisasi ke rentang 0–1 sehingga perbandingan bersifat proporsional, bukan nilai asli. Terlihat pola korelasi positif: secara umum, titik-titik dengan *Income* lebih tinggi cenderung memiliki Limit Kredit lebih tinggi. Namun, berbeda dengan hubungan *Rating Limit* yang hampir membentuk garis lurus, sebaran *Income Limit* lebih menyebar (*dispersed*) dan menampilkan *variability* yang lebih besar pada level *Income* menengah.

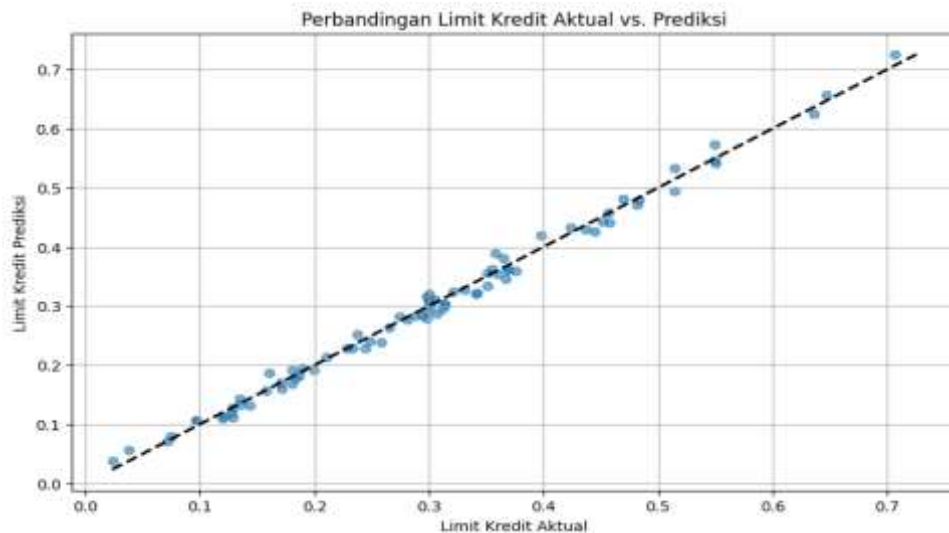


Gambar I2 Hubungan antara *Income* dan Limit Kredit

Pada rentang *Income* rendah hingga menengah ($\approx 0,00$ – $0,30$), titik-titik cukup padat dan Limit Kredit tersebar dari sangat rendah hingga menengah. Ini menandakan bahwa nasabah dengan pendapatan relatif mirip bisa memperoleh limit yang berbeda artinya keputusan limit tidak hanya ditentukan oleh pendapatan, tetapi juga faktor lain (misalnya *Rating* dan *Balance*). Setelah *Income* melewati kisaran menengah ($\approx 0,40$ ke atas), tren kenaikan Limit Kredit menjadi lebih jelas: titik-titik cenderung bergerak ke kanan-atas, dan hampir tidak ada *Limit* yang sangat rendah pada *Income* tinggi. Keberadaan beberapa titik di *Income* tinggi dengan Limit sangat tinggi (mendekati 1) berfungsi sebagai observasi pendorong (*leverage points*) yang mempertegas bahwa kelompok berpendapatan tinggi memiliki peluang menerima limit lebih besar.

10. Hubungan antara *Balance* dan Limit Kredit

Scatter plot ini menggambarkan hubungan antara *Balance* (saldo yang masih terutang / pemakaian kartu kredit saat ini) pada sumbu X dan Limit Kredit pada sumbu Y, keduanya telah dinormalisasi ke rentang 0–1. Secara umum tampak tren korelasi positif: semakin besar *Balance*, cenderung semakin tinggi Limit Kredit. Namun, dibandingkan hubungan *Rating Limit* yang hampir membentuk garis lurus, pola di sini lebih menyebar (*varians* vertikal lebih besar) sehingga kekuatan linearitasnya.

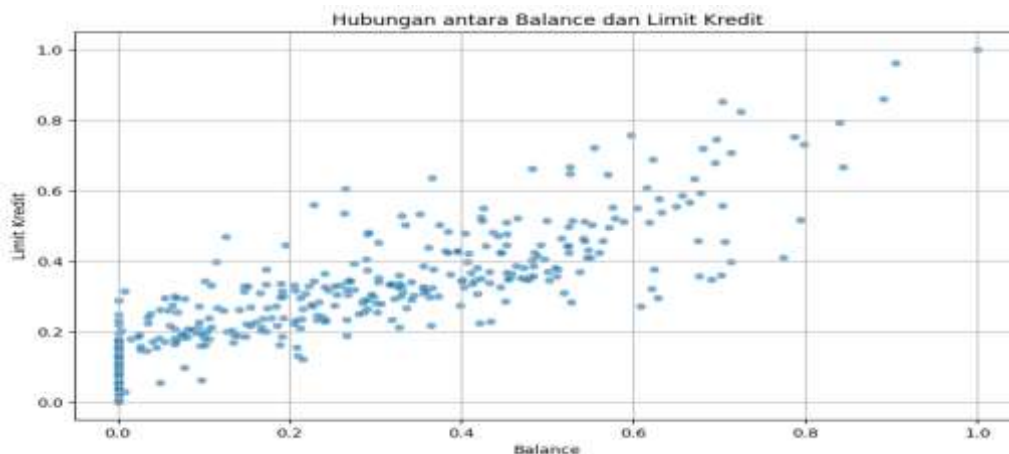


Gambar I3 Hubungan antara Balance dan Limit

Terlihat kluster padat di dekat *Balance* yang sangat rendah (mendekati 0). Kelompok ini menunjukkan banyak nasabah dengan limit yang bervariasi meskipun saldo pemakaian mereka minim—artinya adanya berbagai profil limit awal (*baseline limit*) yang tidak semata-mata ditentukan oleh pemakaian berjalan, melainkan oleh faktor lain seperti *Rating* dan *Income*. Seiring kenaikan *Balance* (0,20 – 0,60), titik-titik bergerak ke arah kanan-atas, tetapi sebaran vertikal masih cukup luas; ini mengindikasikan bahwa nasabah dengan tingkat pemakaian serupa bisa memiliki limit berbeda (misalnya karena perbedaan skor kredit atau pendapatan).

11. Perbandingan Limit Kredit Aktual vs. Prediksi

Grafik *scatter plot* ini membandingkan Limit Kredit Aktual (sumbu X) dengan Limit Kredit Prediksi (sumbu Y) pada data uji, keduanya dalam skala hasil normalisasi (0–1). Setiap titik mewakili satu observasi nasabah. Garis putus-putus hitam adalah garis identitas ($y = x$) yang merepresentasikan kondisi prediksi sempurna: setiap titik yang tepat berada di garis ini berarti nilai prediksi sama persis dengan nilai aktual.



Gambar I4 Perbandingan Limit Kredit Aktual vs Prediksi

Sebagian besar titik pada plot terlihat sangat dekat dengan garis identitas, menunjukkan bahwa model regresi linear menghasilkan prediksi yang akurat. Tidak tampak pola sistematis penyimpangan—misalnya kurva melengkung atau kluster jauh di atas/bawah garis sehingga tidak ada indikasi kuat bias linier atau pelanggaran asumsi bentuk hubungan.

Hasil Pengujian

Karakteristik Data

Dataset terdiri dari 400 nasabah dengan 11 variabel (7 numerik: *Income*, *Rating*, *Cards*, *Age*, *Education*, *Balance*, *Limit*; 4 kategori: *Gender*, *Student*, *Married*, *Ethnicity*), tanpa missing value maupun duplikasi.

Proses *Label Encoding* pada variabel kategori dan Normalisasi (*MinMaxScaler*) pada semua fitur memastikan data siap diolah model dan menghindari dominasi fitur berdimensi besar.

Pemilihan Fitur

Analisis *Correlation Matrix* mengidentifikasi bahwa *Rating*, *Balance*, dan *Income* memiliki korelasi paling kuat dengan Limit kredit (masing-masing *Rating* = 0.996880, *Income* = 0.861697, *Balance* = 0.792088). Ketiga fitur ini menjadi pilar utama dalam membangun model *regresi linear*.

Fitur lain (*Cards*, *Age*, *Education*, *Gender*, *Student*, *Married*, *Ethnicity*) memiliki pengaruh lebih kecil, tetapi tetap memperkaya model dengan informasi demografis.

Performa Model

Model *regresi linear*, dengan data dibagi 80% *train* dan 20% *test*, menghasilkan:

- a. *Mean Squared Error* (MSE) = 0.0001617591203547172
- b. *R-squared* (R^2) = 0.9923846114340684

Angka MSE yang sangat kecil dan R^2 mendekati 1 menunjukkan model mampu menjelaskan 99,44% variasi Limit kredit dan memberikan prediksi yang sangat akurat.

SIMPULAN

Model *regresi linear* yang dibangun menggunakan variabel (*Rating*, *Income*, *Balance*) mampu memprediksi limit kredit dengan sangat baik.

- c. Evaluasi model pada data uji menunjukkan MSE sebesar 0.0001617591203547172 dan nilai R^2 sebesar 0.9923846114340684, yang berarti 99,23% variasi limit kredit berhasil dijelaskan oleh model.
- d. *Rating* muncul sebagai fitur paling dominan (koefisien tertinggi dan korelasi = 0,99), diikuti *Income* (koefisien positif tinggi, korelasi = 0,86) dan *Balance* (koefisien positif sedang, korelasi = 0,79).
- e. Visualisasi memperlihatkan titik-titik data sangat rapat di sekitar garis ideal, membuktikan akurasi dan linearitas hubungan.

SARAN

Dari hasil penelitian yang telah dilakukan, penulis dapat mengusulkan beberapa saran yang dapat diambil agar dapat bermanfaat untuk kemajuan dalam penelitian tersebut

1. Eksplorasi Model Lanjutan Uji model non-linier seperti *Random Forest*, *XGBoost*, atau *Neural Network* untuk membandingkan akurasi prediksi.
2. Melakukan pengecekan mengenai dataset tersebut dengan *Outlier* dan *Noise*

REFERENSI

- Amansyah, I., Indra, J., Nurlaelasari, E., & Juwita, A. R. (2024). Prediksi Penjualan Kendaraan Menggunakan Regresi Linear: Studi Kasus pada Industri Otomotif di Indonesia. *Journal Of Social Science Research*, 4, 1199–1216. <https://j-innovative.org/index.php/Innovative>
- Anggrawan, A., Hairani, H., & Azmi, N. (2022). Prediksi Penjualan Produk Unilever Menggunakan Metode Regresi Linear. *Jurnal Bumigora Information Technology (BITE)*, 4(2), 123–132. <https://doi.org/10.30812/bite.v4i2.2416>
- Arinal, V., Azhari, M., & Pendahuluan, I. (2023). Penerapan Regresi Linear Untuk Prediksi Harga Beras Di Indonesia. 5(1), 341–346.
- Bambang Riono, S. (2022). Analisis Strategi Pemasaran Kredit, Prosedur Pemberian Kredit, Dan Kualitas Pelayanan Terhadap Kepuasan Nasabah Koperasi Simpan Pinjam. *Jurnal Ilmiah Manajemen Dan Kewirausahaan*, 1(3), 375–386. <https://doi.org/10.55606/jimak.v1i3.508>
- Farid, M., Wahyudi, D., Chrisvo, A. R., Hidayatullah, M. F., & Parhusip, J. (2024). Analisis Distribusi Selisih Tingkat Pengangguran Terbuka (Tpt) Dan Tingkat Partisipasi Angkatan Kerja (Tpak) Menggunakan Google Colab. *Jurnal Teknik Informatika Dan Multimedia*, 4(2), 39–45. <https://journal.politeknik-pratama.ac.id/index.php/JTIM/article/view/828/755>
- Indah Sucianty, F., & Suria Manda, G. (2022). Analisis Manajemen Risiko Kredit Usaha (KUR) Untuk Meminimalkan. *Faradila Indah Sucianty1, Gusganda Suria Manda2*, 2022(24), 18–25. <https://doi.org/10.5281/zenodo.7460372>
- Mudassir, A. F., Modding, B., & Hajering, H. (2020). Analisis Kebijakan Pemberian Kredit di Bank Perkreditan Rakyat (BPR). *Journal of Accounting and Finance (JAF)*, 1(1), 1–16. <https://doi.org/10.52103/jaf.v1i1.61>
- Panggabean, D. S. O., Buulolo, E., & Silalahi, N. (2020). Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda. *JURIKOM (Jurnal Riset Komputer)*, 7(1), 56. <https://doi.org/10.30865/jurikom.v7i1.1947>
- Rawan, D., Pidana, T., & Di, N. (2025). 3 1,2,3. 2023, 606–615.
- Rohimah, W., Budianto, E. W. H., & Dewi, N. D. T. (2023). Pemetaan Penelitian seputar Bank CIMB Niaga Syariah dan Konvensional: Studi Bibliometrik VOSviewer dan Literature Review. *JEMPER (Jurnal*

- Ekonomi Manajemen Perbankan*), 5(1), 30. <https://doi.org/10.32897/jemper.v5i1.2607>
- Setyoningrum, N. R., Rahimma, P. J., Teknologi, S. T., Tanjungpinang, I., & Tanjungpinang, K. (2022). Implementasi Algoritma Regresi Linear Dalam Sistem Prediksi Pendaftar Mahasiswa Baru Sekolah Tinggi Teknologi Indonesia Tanjungpinang. *Prosiding Seminar Nasional Ilmu Sosial Dan Teknologi (SNISTEK)*, 4, 13–18. <https://ejournal.upbatam.ac.id/index.php/prosiding/article/view/5200>
- Simanjuntak, J. (2023). Penggunaan Informasi Debitor Dari Sistem Layanan Informasi Keuangan Otoritas Jasa Keuangan (Slik Ojk) Sebagai Alat Bukti Permohonan Pkpu. *Jurnal Hukum To-Ra : Hukum Untuk Mengatur Dan Melindungi Masyarakat*, 9(1), 73–84. <https://doi.org/10.55809/tora.v9i1.209>
- Triono, A., Setia Budi, A., & Abdillah, R. (2023). Agus Triono, Apri Setia Budi, Rafi Abdillah, dan Wahyudi. "Implementasi Peretasan Sandi Vigenere Cipher Menggunakan Bahasa Pemrograman Python." *Jurnal JOCOTIS - Journal Science Informatica and Robotics*, vol. 1, no. 1, September 2023, pp. 1-9. E-ISSN: xxxx. *Jurnal JOCOTIS-Journal Science Informatica and Robotics E-ISSN: Xxxx-Xxxx*, 1(1), 1–9. <https://jurnal.ittc.web.id/index.php/jct/>