

# PENENTUAN DITERIMANYA PENGAJUAN KREDIT MENGUNAKAN LOGISTIK REGRESI

**Murti Setto Hidayani**

Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jakarta

2018470055@ftumj.ac.id

## Abstrak

Pada peraturan yang di atur pada Undang – Undang untuk perbankan, ada beberapa tugas yang harus dijalankan oleh perbankan, salah satunya adalah membuat kredit. Bagi sebagian masyarakat, pemberian kredit dapat membantu ekonomi dari kehidupannya. Namun dalam pengajuan kredit, sering kali ada kredit yang tidak diterima oleh pihak bank dengan beberapa kondisi. Pada artikel ini akan dibahas tentang penentuan diterima atau tidaknya suatu ajuan kredit menggunakan metode logistik regresi. Dalam melakukan implementasi algoritma logistik regresi, penulis menggunakan program *python* dari platform *google colaboratory*.

**Kata Kunci:** *kredit, logistik regresi, google colaboratory*

## Abstract

In the regulations regulated in the Law for banking, there are several tasks that must be carried out by banks, one of which is making credit. For some people, the provision of credit can help the economy of their lives. However, in applying for credit, there are often credits that are not accepted by the bank under several conditions. In this article, we will discuss the determination of whether or not a credit application is accepted using the logistic regression method. In implementing the logistic regression algorithm, the author uses the *python* program from the *google colaboratory* platform.

**Keywords:** *credit, logistic regression, google colaboratory*

## PENDAHULUAN

Dilingkup masyarakat, kredit dianggap dapat membantu kondisi ekonomi suatu keluarga. Menurut bahasa Yunani, kredit (*creede*) artinya kepercayaan. Menurut Undang-Undang Nomor 10 Tahun 1998 tentang Perubahan Atas Undang-Undang Nomor 7 Tahun 1992 tentang Perbankan, kredit merupakan penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam meminjam antara bank

dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga.

Istilah kredit dikenal sebagai salah satu fitur dari dunia perbankan. Sebelum mendapatkan kredit, masyarakat harus melakukan pengajuan terlebih dahulu. Pengajuan ini perlu melampirkan beberapa data yang akan menjadi tolak ukur diterima atau tidaknya suatu pengajuan kredit. Data ini biasanya berupa nama, alamat, gaji, dan lain sebagainya. Data yang sudah diajukan

oleh pengaju akan diolah oleh pemberi kredit untuk tingkat kelayakannya. Tingkat kelayakan ini mempengaruhi tinggi rendahnya sebuah pinjaman yang diberikan kepada pengaju kredit. Dalam hal ini diperlukannya sebuah penalaran dalam mengambil keputusan layak atau tidaknya pengajuan tersebut. Maka dari itu metode logistik regresi bisa menjadi salah satu pilihan untuk mengolah data yang ada untuk memprediksi probabilitas kejadian dari suatu peristiwa dengan cara mencocokkan data tersebut pada suatu fungsi.

Tujuan dari penelitian ini adalah memprediksi tingkat akurasi data dari variabel – variabel penentu dengan label yang ada pada data sehingga dapat menentukan tinggi rendahnya peminjaman dari pengajuan kredit.

Data yang digunakan merupakan data record dari suatu aplikasi kredit sesuai dengan ketentuan – ketentuan yang dapat menentukan tinggi rendahnya suatu peminjaman. Data ini didapatkan dari website kaggle.com untuk menjadi data training. Data ini merupakan data mentah yang nantinya akan diproses terlebih dahulu (*cleaning*). Berikut merupakan variabel serta label yang belum melalui proses *cleaning*.

TABEL 1.  
VARIABEL DATA MENTAH

#	Column	Non-Null Count	Dtype
0	Unnamed: 0	1000 non-null	int64
1	Age	1000 non-null	int64
2	Sex	1000 non-null	object
3	Job	1000 non-null	int64
4	Housing	1000 non-null	object
5	Saving accounts	817 non-null	object
6	Checking account	606 non-null	object
7	Credit amount	1000 non-null	int64
8	Duration	1000 non-null	int64
9	Purpose	1000 non-null	object

Metode logistik regresi dapat menghasilkan suatu prediksi probabilitas

kelayakan diterima atau tidaknya suatu pengajuan yang ada dengan cara mencocokkan data data yang ada sehingga ada tolak ukur seberapa tingginya sebuah peminjaman.

## TINJAUAN PUSTAKA

### A. Kredit

Kredit merupakan penyediaan uang berupa tagihan berdasarkan kesepakatan dan persetujuan pinjam – meminjam anatrabbank dengan pihak peminjam. Pihak peminjam diwajibkan melunasi utangnya sesuai dengan jangka waktu yang ditentukan dan pemberian bunga. (Apriyanto, 2015)

### B. Logistik Regresi

Menurut Hosmer & Lemeshow, logistik regresi merupakan sebuah analisis yang menjelaskan hubungan antara respon variabel. Variabel yang digunakan diperjelas berupa data kuantitatif atau kualitatif. (Indraswari & Yuhan, 2017)

### C. Google Colaboratory

Google Colaboratory merupakan platform yang memungkinkan pengguna untuk menjalankan program *python* di browser sehingga pengguna dapat mengakses program *python* tanpa melakukan konfigurasi terlebih dahulu. (Google, t.thn.)

## METODE PENELITIAN

### A. Variabel Penelitian

	Age	Sex	Job	Housing	Saving accounts	Credit amount	JK	House	Sosial	Credit
0	22	female	2	own	little	tinggi	0	1	0	1
1	49	male	1	own	little	rendah	1	1	0	0
2	45	male	2	free	little	tinggi	1	0	0	1
3	53	male	2	free	little	tinggi	1	0	0	1
4	53	male	2	own	quite rich	rendah	1	1	2	0

Gambar 1. Data Bersih

Adapun variabel dan label yang digunakan untuk melakukan prediksi antara lain ada pada tabel 2.

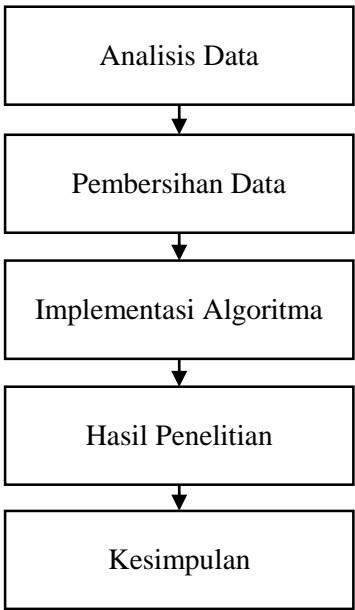
TABEL 2.  
VARIABEL PENELITIAN

#	Column	Variabel
0	Age	X1
1	JK(Sex)	X2
2	Job	X3
3	House (Housing)	X4
4	Sosial (Saving accounts)	X5
5	Credit (Credit amount)	Y

Pada tabel 2 dijelaskan bahwa untuk age, jk, job, house dan sosial merupakan variabel penguji sedangkan untuk credit merupakan variabel terikat.

B. Langkah Analisis

Berikut merupakan metode dari analisis data di gambar 1.



Gambar 2. Metode Analisis

HASIL DAN PEMBAHASAN

Data yang telah dipaparkan pada tabel 1 mengalami pembersihan data sehingga didapatkan empat variabel prediksi (X) dan satu label. Data yang type datanya berupa *object* dirubah menjadi *integer* sehingga didapatkan pada tabel 2. Berikut pada gambar 2 tampilan datanya.

Pada data bersih dapat dilihat bahwa adanya perubahan data *object* menjadi data *integer* yang dijelaskan perubahannya pada tabel 3.

TABEL 3.  
PERUBAHAN TIPE DATA

Data <i>object</i>	Data <i>integer</i>
Sex	JK
Housing	House
Saving accounts	Sosial
Credit amount	Credit

Setelah melakukan pembersihan data dan perubahan tipe data, dapat dilihat korelasi antar kolom numerik pada setiap variabel sehingga didapatkan hasil seperti pada gambar 3.

	Age	Job	JK	House	Sosial	Credit
Age	1.000000	-0.015036	0.150733	-0.314986	0.037330	-0.016902
Job	-0.015036	1.000000	0.070399	-0.103523	-0.031768	0.241557
JK	0.150733	0.070399	1.000000	-0.191171	-0.005562	0.049174
House	-0.314986	-0.103523	-0.191171	1.000000	0.041994	-0.095612
Sosial	0.037330	-0.031768	-0.005562	0.041994	1.000000	-0.058287
Credit	-0.016902	0.241557	0.049174	-0.095612	-0.058287	1.000000

Gambar 3. Korelasi untuk Setiap Variabel

Dari hasil tersebut didapatkan korelasi antara dari variabel prediksi terhadap label. Didapatkan bahwa variabel job memiliki korelasi terdekat dengan label credit. Hal tersebut dapat digunakan sebagai x dalam perhitungan logistik regresi. Namun penulis menggunakan kelima variabel untuk mendapatkan nilai akurasi yang lebih baik. Maka didapatkan variabel x dan y yang akan digunakan seperti pada gambar 4.

```
x = dataa[['Age', 'Job', 'JK',
           'House', 'Sosial']].values
y = dataa['Credit'].values
```

Gambar 4. *Syntax* variabel x dan y

Setelah menentukan nilai x dan nilai y dapat dilakukan perhitungan dengan metode logistik regresi. Berikut merupakan hasil program yang telah dibuat pada gambar 5.

```
[39] normal_x = scaler.fit_transform(x)
normal_x

array([[ -1.15747673,  0.16106733, -1.45954383, -0.14796669, -0.52548243],
       [ 1.22320198, -1.3690723 ,  0.68514558, -0.14796669, -0.52548243],
       [ 0.87050884,  0.16106733,  0.68514558, -2.03685403, -0.52548243],
       ...,
       [ 0.25329584,  0.16106733,  0.68514558, -0.14796669, -0.52548243],
       [-1.06930344,  0.16106733,  0.68514558, -2.03685403, -0.52548243],
       [-0.7166103 ,  0.16106733,  0.68514558, -0.14796669,  0.62550724]])

[40] X_train, X_test, y_train, y_test = train_test_split(normal_x, y.ravel(),
                                                         test_size=0.2, random_state=0)

[41] LR = LogisticRegression()
LR.fit(X_train, y_train)

LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
intercept_scaling=1, l1_ratio=None, max_iter=100,
multi_class='auto', n_jobs=None, penalty='l2',
random_state=None, solver='lbfgs', tol=0.0001, verbose=0,
warm_start=False)

[42] from sklearn.metrics import accuracy_score

ypred = LR.predict(x)
print("Akurasi ", accuracy_score(y, ypred))

ypredik = LR.predict(X_test)
print("Akurasi terstandarisasi ", accuracy_score(y_test, ypredik))

Akurasi 0.6634026927784578
Akurasi terstandarisasi 0.676829268292683
```

Gambar 5. Program Logistik Regresi

Dari perhitungan algoritma logistik regresi dengan data yang digunakan, didapatkan akurasi sebesar 66,3% dan akurasi untuk data yang sudah distandarisasi didapatkan akurasi sebesar 67,7%.

Kemudian dibuat percobaan kembali dengan menghitung dengan algoritma logistik regresi namun dengan nilai credit yang berupa jumlah kredit yang didapatkan. Datanya seperti pada gambar 6.

	Age	Sex	Job	Housing	Saving accounts	Credit amount	JK	House	Sosial
1	22	female	2	own	little	5951	0	1	0
2	49	male	1	own	little	2096	1	1	0
3	45	male	2	free	little	7882	1	0	0
4	53	male	2	free	little	4870	1	0	0
6	53	male	2	own	quite rich	2835	1	1	2

Gambar 6. Data dengan Nilai Credit Amount

Pada tabel 4 tersebut dibuat perhitungan menggunakan metode logistik regresi dan didapatkan nilai akurasinya sebesar 0%. Maka untuk melakukan perhitungan diperlukan adanya pelabelan terlebih dahulu untuk variabel terikatnya. Berikut merupakan hasil akurasinya pada gambar 7.

```
ypred1 = lr.predict(x_)
print("Akurasi ", accuracy_score(y_, ypred1))

Akurasi 0.0
```

Gambar 7 Hasil Akurasi dengan Nilai Credit Amount

## KESIMPULAN DAN SARAN

Pada perhitungan prediksi data penerima kredit menggunakan metode logistik regresi didapatkan bahwa dari variabel yang ada memiliki akurasi sebesar 66,3%. Angka tersebut dinilai cukup baik sehingga dapat disimpulkan bahwa variabel yang ada memiliki ketepatan dalam menentukan atau memprediksikan diterima tidaknya ajuan kredit. Namun setelah dilakukannya standarisasi, angka akurasi yang didapat justru meningkat yaitu sebesar 67,7%.

Ketika tidak melakukan pelabelan untuk variabel terikatnya, nilai akurasi yang didapatkan justru tidak baik. Maka untuk melakukan prediksi dengan menggunakan metode logistik regresi disarankan untuk melakukan pelabelan terlebih dahulu agar mendapatkan nilai yang baik.

Berdasarkan hasil analisis yang telah dilakukan, disarankan untuk memperhatikan keseimbangan data untuk mendapatkan nilai akurasi yang lebih tinggi.

## DAFTAR PUSTAKA

- Apriyanto, Y. B. (2015). PENYELESAIAN KREDIT BERMASALAH PADA BANK DKI JAKARTA CABANG SOLO MELALUI JALUR NON LITIGASI.
- Google. (n.d.). *Selamat Datang di Colaboratory*. Retrieved Juli 2021, from Google Colaboratory: <https://colab.research.google.com/notebooks/welcome.ipynb?hl=id-ID>
- Indraswari, R. R., & Yuhan, R. J. (2017). FAKTOR-FAKTOR YANG MEMENGARUHI PENUNDAAN KELAHIRAN ANAK PERTAMA DI WILAYAH PERDESAAN INDONESIA: ANALISIS DATA SDKI 2012. *JURNAL KEPENDUDUKAN INDONESIA*, 1-12.