

Penerapan Algoritma Naive Bayes dan Multiple Linear Regression untuk Prediksi Status dan Plafon Kredit (Studi Kasus: Bank ABC)

Arya Prawira¹, Desi Arisandi², Tri Sutrisno³

^{1, 2, 3} Universitas Tarumanagara, Jl. Letjen S. Parman No. 1 Kota Jakarta Barat 11440 DKI Jakarta
arya.535190015@stu.untar.ac.id

Abstract

Along with changing technology, Human resources are still needed in many parts of decision making. The companies and organizations still use human to analysis data. Despite of that performance, human analysis took longer time and effort to complete. And sometimes there is always a negative factor of human resources such as unmanageable human. Therefore, it's always important to provide an excellent training source so this human resource able to reach an outcome that needed. Machine learning is one of the most common knowledge that use in decision making. There are many forms of machine learning such as regression, classification, clustering, etc. two of which is used in this application, regression and classification. Naive Bayes regression is one of classification method which rooted on Bayes theorem. Naive Bayes use historical data to predict future outcome based on the characteristic on that historical data. Multiple Linear Regression involves more than one independent variable or predictor. With machine learning and human resources, man can easily to analyse credit worthiness and determine the credit limit of one bank costumer without taking a long time and much effort.

Keywords: Machine learning, classification, Naive bayes algorithm, regression, Multi Linear Regression

Abstrak

Seiring dengan perubahan teknologi, sumber daya manusia masih dibutuhkan dalam banyak bagian pengambilan keputusan. Perusahaan dan organisasi masih menggunakan manusia untuk menganalisis data. Terlepas dari kinerja itu, analisis manusia membutuhkan waktu dan upaya lebih lama untuk diselesaikan. Dan terkadang selalu ada faktor negatif dari sumber daya manusia seperti manusia yang tidak dapat diatur. Oleh karena itu, selalu penting untuk menyediakan sumber pelatihan yang baik agar sumber daya manusia ini dapat mencapai hasil yang dibutuhkan. Pembelajaran mesin adalah salah satu pengetahuan paling umum yang digunakan dalam pengambilan keputusan. Ada banyak bentuk pembelajaran mesin seperti regresi, klasifikasi, pengelompokan, dll. Dua diantaranya digunakan dalam aplikasi ini, regresi dan klasifikasi. Regresi Naive Bayes adalah salah satu metode klasifikasi yang berakar pada teorema Bayes. Naive Bayes menggunakan data historis untuk memprediksi hasil masa depan berdasarkan karakteristik pada data historis tersebut. Regresi Linier Berganda melibatkan lebih dari satu variabel atau prediktor independen. Dengan machine learning dan sumber daya manusia, manusia dapat dengan mudah menganalisis kelayakan kredit dan menentukan batas kredit satu nasabah bank tanpa memakan waktu lama dan tenaga.

Kata kunci: Pembelajaran mesin, klasifikasi, Algoritma naive bayes, regresi, Regresi Berganda

Copyright (c) 2022 Arya Prawira, Desi Arisandi, Tri Sutrisno

Corresponding author: Arya Prawira

Email Address: arya.535190015@stu.untar.ac.id (Jl. Letjen S. Parman No. 1 Kota Jakarta Barat, DKI Jakarta)

Received 20 Desember 2022, Accepted 26 Desember 2022, Published 27 Desember 2022

PENDAHULUAN

Kredit merupakan sebuah kegiatan memberikan suatu prestasi baik itu berupa barang, jasa, ataupun uang dari suatu pihak (pemberi kredit) kepada pihak lain (penerima kredit) yang beratkan atas rasa percaya kepada peminjam prestasi bahwa dalam tenggat waktu yang diberikan peminjam tersebut akan mengembalikan kredit yang telah ia pinjam disertai dengan suatu kontra prestasi (balas jasa) yaitu dapat berupa bunga pinjaman sesuai dengan kesepakatan yang telah dicapai sebelumnya.

Dalam praktiknya, pemberi kredit (kreditur) tentunya mempunyai syarat dalam memberikan sebuah kredit baik itu dalam bentuk modal maupun hal yang lainnya. Hal ini dilakukan agar terdapat

suatu pengamanan terkait pemberian kredit tersebut. Syarat tersebut berupa pemberian jaminan kepada kreditur. Jaminan ini berfungsi untuk memberikan kepastian hukum bagi kreditur dalam pelunasan sebuah hutang. Dengan adanya pemberian pinjaman ini, pihak kreditur dapat meminimalisir risiko kerugian yang terjadi ketika seorang debitur gagal dalam memenuhi kewajibannya.

Meskipun sebuah perjanjian kredit bersifat mengikat, namun tidak jarang seorang peminjam kredit tidak dapat memenuhi kewajiban kreditnya. Hal ini tentu saja sering menimbulkan masalah karena mengakibatkan kerugian yang besar bagi para kreditur. Untuk itu diperlukan sebuah Lembaga keuangan yang mengatur proses kredit di Indonesia.

OJK atau Otoritas Jasa Keuangan adalah sebuah lembaga yang dibentuk untuk mengawasi dan mengatur keseluruhan sektor jasa keuangan yang ada di Indonesia. OJK sendiri memiliki tugas mengawasi sektor keuangan dari lembaga - lembaga keuangan seperti Bank, Pasar Modal, dan IKNB. OJK sendiri juga mengawasi historis kredit dari seluruh debitur yang ada di Indonesia sehingga data dari OJK dapat digunakan untuk memprediksi.

Pada bank ABC, setelah memperoleh data - data yang diperlukan untuk pengambilan keputusan kredit, tahap selanjutnya adalah memberikan data tersebut kepada tim analis untuk kemudian di analisis secara manual. Dikarenakan proses analisis yang masih dilakukan dengan cara manual, tentunya hal ini memakan waktu yang cukup lama dan tidak dapat diputuskan secara instan. Untuk itu diperlukan sebuah sistem yang dapat memprediksi status kredit seseorang dengan instan dan cepat sehingga proses kredit dapat berjalan lebih efisien. Untuk itu digunakan algoritma *Naive Bayes* dan *Multiple Linear Regression* untuk memprediksi kredit seorang debitur.

METODE

Dalam Data Mining, terdapat tiga metode pembelajaran berbeda yaitu *supervised learning*, *unsupervised learning*, dan *semi-supervised learning*. Pada *supervised learning* algoritma data mining bekerja dengan menggunakan label yang telah diketahui nilainya. Biasanya nilai label - label tersebut dapat berupa nilai numerik maupun nilai kategorikal. Terdapat beberapa komponen penting dalam klasifikasi yaitu:

1. Kelas (*class*)

Kelas merupakan sebuah variabel dependen yang ada di dalam sebuah model klasifikasi.

2. Prediktor (*predictor*)

Prediktor merupakan variabel independen yang ada di dalam model klasifikasi.

3. Data latih (*training data*)

Dataset latih adalah kumpulan data yang berisi nilai - nilai yang digunakan untuk melatih model.

4. Data uji (*testing data*)

Dataset uji merupakan kumpulan data baru yang akan digunakan dalam proses uji klasifikasi model.

Naïve Bayes

Algoritma *Naive Bayes* adalah metode algoritma pengklasifikasian statistik yang biasanya digunakan untuk memprediksi sebuah probabilitas keanggotaan suatu kelas. Dasar dari algoritma *Naive Bayes* adalah teorema bayes yang memiliki kemampuan klasifikasi yang mirip dengan algoritma *Decision Tree* dan *Neural Network*. Berikut adalah bentuk umum dari teorema Bayes dapat dilihat pada persamaan (1).

$$P(X) = \frac{P(X|H)(H)}{P(X)} \quad (1)$$

Keterangan :

X = Data dengan kelas yang belum diketahui

H = Hipotesis data X yang merupakan suatu kelas spesifik

$P(H|X)$ = Peluang hipotesis H berdasarkan kondisi X (*Posteriori prob*)

$P(H)$ = Peluang hipotesis H (*prior prob*)

$P(X|H)$ = peluang X berdasarkan kondisi tersebut

$P(X)$ = peluang dari X

Langkah kerja dari algoritma Naïve Bayes adalah sebagai berikut :

- a. Menghitung peluang masing - masing $P(H)$
- b. Menentukan peluang X atau kategori dari atribut
- c. Menghitung peluang terbagi menjadi dua bagian
 1. Jika data yang diambil merupakan data kategorikal, maka peluang kategori dihitung dari atribut yang sama dengan label yang sama menggunakan persamaan (1) .
 2. Jika data yang diambil merupakan data numerik maka langkah yang digunakan adalah sebagai berikut:
 - a) Menghitung nilai rata - rata (*mean*) dengan persamaan (2)

$$\underline{X} = \frac{\sum X}{n} \quad (2)$$

Keterangan:

\underline{X} = nilai rata - rata,

$\sum X$ = penjumlahan keseluruhan data,

n = banyak data

- b) Menghitung nilai standar deviasi dari masing masing kelas data numerik dengan persamaan (3).

$$\sigma = \sqrt{\frac{\sum (X_i - \underline{X})^2}{N}} \quad (3)$$

Keterangan:

σ = Standar Deviasi

X_i = Nilai X ke i

\underline{X} = Nilai rata – rata

N = banyak data

- c) Menghitung probabilitas tiap kelas
- d) Menghitung nilai *Gaussian* dari data numerik dengan menggunakan persamaan *Gaussian* (4).

$$P(H_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(h_j - \mu)^2}{2\sigma^2}\right) \quad (4)$$

- e) Mengalikan semua nilai hasil sesuai dengan data yang dicari labelnya.
- f) Menentukan nilai label pada atribut tertentu.

d. Regresi Linier Berganda

Terdapat dua jenis regresi linier yaitu regresi linier sederhana yang hanya melibatkan satu variabel pemberi pengaruh dan regresi linier berganda yang melibatkan lebih dari satu variabel berpengaruh (Santoso dan Suryadi,2010: 33).

Di dalam regresi, variabel dibedakan menjadi dua yaitu variabel bebas (*independent variable*) dan variabel terikat (*dependent variabel*). variabel bebas adalah variabel yang memiliki pengaruh terhadap variabel lain, sedangkan variabel terikat adalah variabel yang nilainya dipengaruhi oleh variabel lain.

Bentuk persamaan umum dari regresi linear berganda dapat dilihat pada rumus (5).

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_n X_n + \epsilon_i \quad (5).$$

Keterangan:

Y = Variabel Dependen

X_n = Variabel Independen

α = Konstanta

β_1 = Konstanta Variabel Independen

ϵ_i = error

Berdasarkan persamaan (5) maka dapat dibentuk sebuah matriks regresi linier berganda dengan bentuk seperti dapat dilihat pada matriks (6).

$$[Y_1 \ Y_2 \ : \ Y_n] = [1 \ X_{11} \ X_{21} \ \dots \ X_{k1} \ 1 \ X_{12} \ X_{22} \ \dots \ X_{k2} \ \vdots \ \vdots \ \vdots \ 1 \ X_{1n} \ X_{2n} \ \dots \ X_{kn}] [\beta_1 \ \beta_2 \ : \ \beta_n] + [\epsilon_1 \ \epsilon_2 \ : \ \epsilon_n]$$

Keterangan:

Y = Vektor pengamatan berukuran $n \times 1$.

X = Matriks variabel bebas ukuran $n \times k$.

β_n = Vektor parameter yang akan diprediksi berukuran $k \times 1$.

ϵ_n = Vektor random error berukuran $n \times 1$.

HASIL DAN DISKUSI

Berikut adalah kriteria yang digunakan dalam proses prediksi status kredit yang dapat Tabel

1.

Tabel 1 Kriteria Prediksi Status Kredit

Kode	Kriteria	Deskripsi	Tipe
C1	Jaminan	Jaminan Yang Diberikan Seorang Debitur	Kategorikal
C2	Pemasukan	Total Pemasukan Yang Diterima Seorang Debitur Per Tahun	Numerik
C3	Kol_Class	Kolektibilitas Ojk Yang Dimiliki Oleh Debitur	Kategorikal

Berikut adalah atribut pada kriteria jaminan yang dapat dilihat pada Tabel 2.

Tabel 2 Kriteria Jaminan

Tabel Kriteria JAMINAN
Kriteria
SHM (Surat Hak Milik)
Kendaraan Surat Lengkap
Kendaraan Surat Tidak Lengkap

Berikut adalah atribut pada kriteria kelas kolektibilitas dapat dilihat pada Tabel 3.

Tabel 3 Kriteria Kelas Kolektibilitas

Tabel Kriteria KOL_CLASS
Kriteria
KOL-5
KOL-4
KOL-3
KOL-2
KOL-1

Untuk melakukan perhitungan, diperlukan sebuah data latih. Berikut adalah data latih yang digunakan yang berjumlah 20 data dapat dilihat pada Tabel 4.

Tabel 4 Data Latih

NO	CUST_NAME	JAMINAN	PEMASUKAN	KOL_CLASS	KREDIT_STAT
1	Reza	SHM	2.400.000.000	KOL-1	1
2	Adtar	SHM	3.140.000.000	KOL-1	1
3	Alex	KENDARAAN SURAT LENGKAP	250.000.000	KOL-4	0
4	Anwar	SHM	380.000.000	KOL-2	1
5	Arif	SHM	600.000.000	KOL-1	1
6	Armen	SHM	1.500.000.000	KOL-1	1
7	Boy	KENDARAAN SURAT LENGKAP	2.000.000.000	KOL-1	1
8	Kairul	SHM	2.320.000.000	KOL-1	1
9	Andra	SHM	1.800.000.000	KOL-2	1
10	Ihsan	SHM	850.000.000	KOL-1	1
11	Multhi CV	KENDARAAN SURAT LENGKAP	1.800.000.000	KOL-1	1
12	Dutra CV	SHM	750.000.000	KOL-1	1

13	Edwin	SHM	900.000.000	KOL-2	1
14	Harna	KENDARAAN SURAT TIDAK LENGKAP	400.000.000	KOL-4	0
15	Hendra	SHM	900.000.000	KOL-1	1
16	Adi	SHM	700.000.000	KOL-1	1
17	Indah CV	SHM	2.000.000.000	KOL-1	1
18	Radi	KENDARAAN SURAT TIDAK LENGKAP	800.000.000	KOL-1	0
19	Irmantedi	SHM	450.000.000	KOL-3	1
20	Isra	SHM	700.000.000	KOL-2	1

Data latih terlebih dahulu dipisahkan menjadi dua bagian yaitu data yang berkategori numerik dan data yang berkategori kategorikal. Berikut adalah data yang telah dipisahkan berdasarkan target dan kategorinya yang dapat dilihat pada Tabel 5.

Tabel 5 Pemisahan berdasarkan target dan tipe data

Data Diterima			
Data Numerik	Data Kategorikal		
2.400.000.000	SHM	KOL-1	
3.140.000.000	SHM	KOL-1	
380.000.000	SHM	KOL-2	
600.000.000	SHM	KOL-1	
1.500.000.000	SHM	KOL-1	
2.000.000.000	KENDARAAN SURAT LENGKAP	KOL-1	
2.320.000.000	SHM	KOL-1	
1.800.000.000	SHM	KOL-2	
850.000.000	SHM	KOL-1	
1.800.000.000	KENDARAAN SURAT LENGKAP	KOL-1	
750.000.000	SHM	KOL-1	
900.000.000	SHM	KOL-2	
900.000.000	SHM	KOL-1	
700.000.000	SHM	KOL-1	
2.000.000.000	SHM	KOL-1	
450.000.000	SHM	KOL-3	
700.000.000	SHM	KOL-2	
Data Ditolak			
250.000.000		Kendaraan Surat Lengkap	KOL-4
400.000.000		Kendaraan Surat Tidak Lengkap	KOL-4
800.000.000		Kendaraan Surat Tidak Lengkap	KOL-2

Setelah dipisahkan, selanjutnya dihitung nilai rata – rata dari data dengan tipe kategorikan untuk tiap kelas berikut adalah nilai rata-rata tiap kelas dapat dilihat pada Tabel 6.

Tabel 6 Nilai Rata - Rata Tiap Kelas

Nilai Rata-Rata Tiap Kelas	
Kelas	Pemasukan
1	1.364.117.647
0	483.333.333

Kemudian dihitung nilai standar deviasi tiap kelas dengan menggunakan nilai rata – ratanya. Berikut adalah nilai standar deviasi dari tiap kelas yang dapat dilihat pada Tabel 7.

Tabel 7 Nilai Standar Deviasi Tiap Kelas

Nilai Standar Deviasi Tiap Kelas	
Kelas	Pemasukan
1	817.389.586
0	284.312.035

Langkah selanjutnya adalah menghitung probabilitas atau peluang dari tiap kelas. Berikut adalah probabilitas tiap kelas pada Tabel 8.

Tabel 8 Probabilitas Tiap Kelas

Kelas	Jumlah Data	Probabilitas
1	17	0,85
0	3	0,15
	20	1

Lalu, dihitung densitas Gauss dari masing masing kelas dengan hasil yang dapat dilihat pada Tabel 9.

Tabel 9 Densitas Gauss

Nilai Densitas Gauss Tiap Kelas	
Kelas	Pemasukan
1	9,58022e-06
0	5,57904e-06

Setelah didapatkan densitas Gauss dari data numerik, maka selanjutnya perlu dihitung nilai probabilitas dari masing – masing atribut pada kriteria. Berikut adalah hasil perhitungan dari data kategorikal yang dapat dilihat pada Tabel 10.

Tabel 10 Perhitungan Probabilitas Data Kategorikal

Tabel Kriteria JAMINAN	0	1
SHM	0	0,882352941
Kendaraan Surat Lengkap	0,333333333	0,117647059
Kendaraan Surat Tidak Lengkap	0,666666667	0
TOTAL	1	1
Tabel Kriteria KOL_CLASS	0	1

KOL-5	0	0
KOL-4	0,666666667	0
KOL-3	0	0,058823529
KOL-2	0	0,235294118
KOL-1	0,333333333	0,705882353
TOTAL	1	1

Setelah didapatkan nilai probabilitas tiap kriteria pada tiap kelas, maka selanjutnya adalah menggunakan data uji untuk menghitung nilai maksimal. Berikut adalah data uji yang digunakan pada model dapat dilihat pada Tabel 10.

Tabel 11 Data Uji

PEMASUKKAN	655000000
JAMINAN	SHM
KOL_CLASS	KOL-2

Lalu dihitung nilai maksimal dari masing – masing kelas yang dapat dilihat pada Tabel 12.

Tabel 12 Nilai Maksimal

Menghitung Nilai Maximal Dari Masing Masing Kelas				
Kelas	655.000.000	SHM	KOL-2	TARGET
1	9,58022E-06	0,882353	0,235294118	0,000001989
0	5,57904E-06	0	0	0
				1,98897E-06

Dari hasil perhitungan tersebut terlihat bahwa nilai dari kelas 1 (0,000001989) lebih besar daripada nilai dari kelas 0 (0) sehingga dapat ditarik hasil prediksi yaitu diterima. Selanjutnya menghitung plafon kredit. Untuk proses menghitung plafon kredit menggunakan algoritma Regresi Linier Berganda dengan menggunakan perhitungan matriks. Data latih yang digunakan untuk proses perhitungan ini berjumlah 10 data latih. Berikut adalah data latih yang digunakan yang dapat dilihat pada Tabel 13.

Tabel 13. Data Latih Plafon Kredit

Cust Name	Expense	Dependent	Work Status	Marital Stat	House Stat	Asset Stat	Limit Plafon Idr
Reza	1.900.000.000	1	Pekerja Tetap	Married	Owned	New	3.000.000.000
Aftar	2.600.000.000	1	Wirausaha	Single	Owned	New	4.835.000.000
Anwar	105.100.000	1	Pekerja Lepas	Single	Dinas	Used	640.000.000
Arif	504.400.000	1	Pekerja Tetap	Single	Owned	New	850.000.000
Armen	1.000.000.000	3	Wirausaha	Married	Owned	New	2.000.000.000
Boy	1.500.000.000	3	Wirausaha	Married	Owned	New	2.430.000.000
Kairul	1.500.000.000	3	Pekerja Tetap	Single	Owned	Used	2.890.000.000
Andra	700.000.000	2	Pekerja Tetap	Married	Owned	New	2.700.000.000
Ihsan	800.000.000	2	Wirausaha	Single	Owned	Used	1.600.000.000

Multhi Cv	1.450.000.000	3	Wirausaha	Single	Contract	New	2.950.000.000
-----------	---------------	---	-----------	--------	----------	-----	---------------

Dikarenakan terdapat data yang bertipe kategorikal, maka diperlukan normalisasi data ke dalam bentuk numerik terlebih dahulu. Berikut adalah normalisasi data kategorikal Work Status pada Tabel 14, Marital Status pada Tabel 15, House Status pada Tabel 16, dan Asset Status pada Tabel 17.

Tabel 14 Normalisasi Work Status

Work_Status	
Wirausaha	4
Pekerja Tetap	3
Pekerja Kontrak	2
Pekerja Lepas	1

Tabel 15 Normalisasi Marital Status

Marital_Stat	
Married	1
Single	0

Tabel 16 Normalisasi House Status

House Stat	
Owned	4
Family	3
Dinas	2
Contract	1

Tabel 17 Normalisasi Asset Status

Asset_Stat	
New	1
Used	0

Berikut adalah bentuk dari data latih setelah dinormalisasi yang dapat dilihat pada Tabel 18.

Tabel 18 Data Latih Normalisasi

	X1	X2	X3	X4	X5	X6	Y
Cust_Name	Expense	Dependent	Work_Status	Marital_Status	House Stat	Asset_Status	Limit_Plafon_Idr
REZA	1.900.000.000	1	3	1	4	1	3.000.000.000
ADTAR	2.600.000.000	1	4	0	4	1	4.835.000.000
ANWAR	105.100.000	1	1	0	2	0	640.000.000
ARIF	504.400.000	1	3	0	4	1	850.000.000
ARMEN	1.000.000.000	3	4	1	4	1	2.000.000.000
BOY	1.500.000.	3	4	1	4	1	2.430.000.000

	000						
KAIRUL	1.500.000.000	3	3	0	4	0	2.890.000.000
ANDRA	700.000.000	2	3	1	4	1	2.700.000.000
IHSAN	800.000.000	2	4	0	4	0	1.600.000.000
MULTHI CV	1.450.000.000	3	4	0	1	1	2.950.000.000

Kemudian dibentuk sebuah matriks X dari data latih yang telah dinormalisasi. Matriks yang terbentuk dapat yaitu sebagai berikut.

$$X = \begin{bmatrix} 1 & 1900000000 & 1 & 3 & 1 & 4 & 1 \\ 1 & 2600000000 & 1 & 4 & 0 & 4 & 1 \\ 1 & 105100000 & 1 & 1 & 0 & 2 & 0 \\ 1 & 504400000 & 1 & 3 & 0 & 4 & 1 \\ 1 & 1000000000 & 3 & 4 & 1 & 4 & 1 \\ 1 & 1500000000 & 3 & 4 & 1 & 4 & 1 \\ 1 & 1500000000 & 3 & 3 & 0 & 4 & 0 \\ 1 & 700000000 & 2 & 3 & 1 & 4 & 1 \\ 1 & 800000000 & 2 & 4 & 0 & 4 & 0 \\ 1 & 1450000000 & 3 & 4 & 0 & 1 & 1 \end{bmatrix}$$

Langkah selanjutnya adalah melakukan transpose terhadap matriks X. Bentuk transpose dari matriks X adalah sebagai berikut.

$$X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1900 & 26000000 & 105100 & 50440 & 1000000 & 1500000 & 1500000 & 70000 & 8000000 & 14500000 \\ 00000 & 00 & 000 & 0000 & 000 & 000 & 000 & 0000 & 00 & 00 \\ 1 & 1 & 1 & 1 & 3 & 3 & 3 & 2 & 2 & 3 \\ 3 & 4 & 1 & 3 & 4 & 4 & 3 & 3 & 4 & 4 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 4 & 4 & 2 & 4 & 4 & 4 & 4 & 4 & 4 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Lalu, dilakukan proses perkalian matriks X dengan matriks X^T sehingga diperoleh matriks X^TX. Hasil perkalian matriks tersebut adalah sebagai berikut:

$$X^T X = \begin{bmatrix} 10 & 12059500000 & 20 & 33 & 4 & 35 & 7 \\ 12059500000 & 1,9368E+19 & 24459500000 & 43318300000 & 5100000000 & 43677800000 & 9654400000 \\ 20 & 24459500000 & 48 & 70 & 9 & 69 & 14 \\ 33 & 43318300000 & 70 & 117 & 14 & 118 & 25 \\ 4 & 5100000000 & 9 & 14 & 4 & 16 & 4 \\ 35 & 43677800000 & 69 & 118 & 16 & 133 & 25 \\ 7 & 9654400000 & 14 & 25 & 4 & 25 & 7 \end{bmatrix}$$

Kemudian, matriks $X^T X$ akan di invers sehingga terbentuk matriks $(X^T X)^{-1}$. Bentuk dari invers matriks adalah sebagai berikut.

$$(X^T X)^{-1} = \begin{pmatrix} 2,728831148 & 4,71681E-11 & -0,416344831 & 0,017339005 & 0,751283838 & -0,494058049 & -0,687932763 \\ 4,71681E-11 & 3,39245E-19 & 7,01022E-11 & -1,72288E-10 & 1,28476E-11 & 2,66723E-14 & -4,73825E-11 \\ -0,416344831 & 7,01022E-11 & 0,339718999 & -0,303827694 & -0,379037098 & 0,14727392 & 0,415935099 \\ 0,017339005 & -1,72288E-10 & -0,303827694 & 0,483989391 & 0,384668407 & -0,168659626 & -0,518052345 \\ 0,751283838 & 1,28476E-11 & -0,379037098 & 0,384668407 & 1,180460971 & -0,308787588 & -0,956481068 \\ -0,494058049 & 2,66723E-14 & 0,14727392 & -0,168659626 & -0,308787588 & 0,194774349 & 0,282656607 \\ -0,687932763 & -4,73825E-11 & 0,415935099 & -0,518052345 & -0,956481068 & 0,282656607 & 1,451529304 \end{pmatrix}$$

Langkah selanjutnya adalah membentuk sebuah Matriks Y yang berasal dari variabel dependent dari data latih. Kemudian Matriks Y dikalikan dengan matriks X^T untuk membentuk matriks $X^T Y$. berikut adalah bentuk matriks $X^T Y$

$$X^T Y = \begin{pmatrix} 2389500000 \\ 3,61945E+19 \\ 48735000000 \\ 8422000000 \\ 10130000000 \\ 85450000000 \\ 18765000000 \end{pmatrix}$$

Kemudian perlu dibentuk sebuah koefisien yang berguna dalam pembentukan persamaan model dari data latih. Koefisien ini terbentuk dengan melakukan invers matriks $X^T Y$. Berikut adalah Koefisien yang terbentuk.

b0	566558553
b1	1,555535645
b2	106525199,3
b3	-103543377,4
b4	-34857834,23
b5	-12169861,69
b6	188890902,3

Dengan koefisien tersebut maka dapat dibentuk sebuah persamaan Regresi Linear. Persamaan yang terbentuk adalah sebagai berikut.

$$Y = 566558552,970736 + 1,55553564547877 X_1 + 106525199,349524 X_2 + (-103543377,385317) X_3 + (-34857834,2272606) X_4 + (-12169861,6867857) X_5 + 188890902,330379 X_6$$

Lalu digunakan sebuah data uji untuk menguji sistem dan memprediksi plafon kredit. Berikut adalah data uji yang digunakan.

- Expense = 587400000
- Dependent = 1
- Work Status = 1
- Marital Status = 1

Jenis Jaminan = SHM

House Status = 1

Data uji tersebut kemudian dimasukkan ke dalam persamaan yang telah terbentuk.

$$Y = 566558552,970736 + 1,55553564547877 (587400000) + 106525199,349524 (1) + (-103543377,385317) (1) + (-34857834,2272606) (1) + (-12169861,6867857) (3) + 188890902,330379 (1)$$

Setelah dilakukan perhitungan maka diketahui plafon yang terbentuk adalah sebesar Rp 1.637.295.081,19229. Kemudian untuk mengetahui keefektifan dari metode klasifikasi yang digunakan, dilakukan perhitungan akurasi. Akurasi dihitung dengan menggunakan rumus Cross Validation Score. Setelah dilakukan perhitungan diperoleh akurasi dari metode Naïve Bayes sebesar 89,43 %. Berikut adalah proses perhitungan akurasi yang dilakukan di Jupyter Notebook dapat dilihat pada gambar 1



```
In [50]: from sklearn.model_selection import cross_val_score
scores = cross_val_score(model,x_train,y_train,cv=6)
scores.mean()

Out[50]: 0.8943355119825788
```

Gambar 1 Akurasi Naive Bayes

Untuk keefektifan dari metode regresi dihitung dengan menggunakan library Sklearn. Setelah dihitung, diperoleh akurasi linear regresi sebesar 75,08 %. Proses perhitungan akurasi regresi dapat dilihat pada gambar 2.

KESIMPULAN

Kesimpulan yang didapatkan pada pengujian data dengan algoritma Naïve Bayes dan Regresi Linear Berganda adalah sebagai berikut, (1) Akurasi yang didapatkan untuk metode *Naïve Bayes* adalah sebesar 89,43% dan untuk metode Regresi Linear Berganda adalah sebesar 75,08% yang menandakan metode ini memiliki kecocokan yang bagus dan dapat digunakan untuk pembuatan aplikasi, (2) Dengan menggunakan kedua metode ini dapat membantu proses analisa dari tim analisis bank ABC sebagai patokan dalam menentukan kelayakan kredit dan plafon kredit yang diberikan pada saat seorang debitur mengajukan kredit.

Selain kesimpulan, tentunya ada saran – saran yang dapat membuat jurnal ini menjadi lebih baik kedepannya. Diantaranya adalah dapat meningkatkan akurasi model dengan mencari metode yang memiliki kecocokan lebih baik lagi dengan data yang digunakan. Jurnal ini akan dilanjutkan dengan pembuatan aplikasi website yang diharapkan akan digunakan oleh Bank ABC.

REFERENSI

Annur, Haditsah. "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes." *ILKOM Jurnal Ilmiah*, Vol.10, no. 2: 160-165., Tahun 2018.

- Bawono, Bonggo, and Rochdi Wasono. "Perbandingan Metode random Forest dan Naive Bayes Untuk Klasifikasi Debitur Berdasarkan Kualitas Kredit." EDUSAINTEK, Vol.3, Tahun 2019.
- Ciptohartono, Claudia Clarentia. Algoritma Klasifikasi Naive Bayes untuk Menilai Kelayakan Kredit. http://eprints.dinus.ac.id/13201/1/jurnal_13671.pdf. Tanggal Akses 9 September 2022.
- Dahri, Diasrina, Fahrul Agus, and Dyna Marisa Khairina. "Metode Naive Bayes Untuk Penentuan Penerima Beasiswa Bidikmisi Universitas Mulawarman." Jurnal Informatika Mulawarman, Vol. 11, No. 2, Tahun 2017.
- Gunawan, Rudi. "Implementasi Data Mining Menggunakan Regresi Linier Berganda dalam Memprediksi Jumlah Nasabah Kredit Macet Pada BPR Tanjung Morawa." Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer), Vol. 18, no. 1: 87-91., Tahun 2019.
- Iriadi, Nandang, and Nia Nuraeni. "kajian penerapan metode klasifikasi data mining algoritma C4. 5 untuk prediksi kelayakan kredit pada bank mayapada jakarta." Jurnal teknik komputer amik BSI ,Vol.2, no. 1: 132-137., Tahun 2016.
- Jonathan, Kevin Marcello. "Perbandingan Kinerja Algoritma Naive Bayes, dan C. 4.5 Untuk Mendeteksi Pengelabuan Uniform Resource Locator (Phishing URL)." Jurnal Ilmu Komputer dan Sistem Informasi , Vol. 8 No. 1 ,Tahun 2020.
- Kelvin, Kelvin, Bagus Mulyawan, and Tri Sutrisno. "PENERAPAN ALGORITMA C4. 5 UNTUK PENERIMAAN KELAYAKAN KREDIT PADA KOPERASI (STUDI KASUS: KOPERASI XYZ)." Jurnal Ilmu Komputer dan Sistem Informasi, Vol 8, no.1: 109-115 tahun 2020.
- Neelamegam, S., and E. Ramaraj. "Classification algorithm in data mining: An overview." International Journal of P2P Network Trends and Technology (IJPTT), Vol.4, no. 8: 369-374., Tahun 2013.
- Ningsih, Setia, and Hendra H. Dukalang. "Penerapan metode suksesif interval pada analisis regresi linier berganda." Jambura Journal of Mathematics, Vol.1, no. 1: 43-53., Tahun 2019.
- Nugraheni, Annisa, Rima Dias Ramadhani, Amalia Beladonna Arifa, and Agi Prasetiadi. "Perbandingan Performa Antara Algoritma Naive Bayes Dan K-Nearest Neighbour Pada Klasifikasi Kanker Payudara." Journal of Dinda: Data Science, Information Technology, and Data Analytics, Vol.2, no. 1: 11-20, Tahun 2022.
- Permana, Tedi, Amril Mutoi Siregar, Anis Fitri Nur Masruriyah, and Ayu Ratna Juwita. "Perbandingan Hasil Prediksi Kredit Macet Pada Koperasi Menggunakan Algoritma KNN dan C5. 0." In Conference on Innovation and Application of Science and Technology (CIASTECH), vol. 3, no. 1, pp. .737-746, Tahun 2020.
- Prihatina Ratih. Mengenal Kolektibilitas (Kol) Kredit Perbankan Kaitannya Dengan dengan Undang-Undang No 4 Tahun 1996 (UUHT).<https://www.djkn.kemenkeu.go.id/kpknl-pekalongan/baca-artikel/14713/Mengenal-Kolektibilitas-Kol-Kredit-Perbankan-Kaitannya-Dengan-dengan-Undang-Undang-No-4-Tahun-1996>

- UUHT.html#:~:text=Kolektibilitas%20(Bahasa%20Inggris%3A%20collectability),surat%20berharga%20atau%20penanaman%20lainnya.Tanggal akses 12 Oktober 2022
- Siregar, Apriliana Mart. "Penyelesaian Kredit Macet dengan Jaminan Personal Guarantee yang Meninggal Dunia Sebelum Pelunasan Kredit." *Jurnal Wawasan Yuridika*, Vol.4, no. 2: 194-212, Tahun. 2020.
- Sulistiyono, Sulistiyono, and Wiwik Sulistiyowati. "Peramalan produksi dengan metode regresi linier berganda." *PROZIMA (Productivity, Optimization and Manufacturing System Engineering)*, Vol. 1, no. 2: 82-89. Tahun.2018.
- Sucipto, Adi. "Prediksi Kredit Macet Melalui Perilaku Nasabah Pada Koperasi Simpan Pinjam Dengan Menggunakan Metode Algoritma Klasifikasi C4. 5." *Jurnal DISPROTEK*, Vol.6, no. 1, Tahun. 2015.
- Wati, Embun Fajar, and Arvin Anggoro Kusumo. "Penerapan Metode Unified Modeling Language (UML) Berbasis Dekstop pada Sistem Pengolahan Kas Kecil Studi Kasus pada PT Indo Mada Yasa Tangerang." *Jurnal Informatika*, Vol.5, no. 1: 24-36. , Tahun 2016.
- Yuliara, I. Made. Regresi Linier Berganda. https://simdos.unud.ac.id/uploads/file_pendidikan_1_dir/5f0221d2b0bb7ced1d61798fab7f4ad3.pdf .Tanggal akses 14 Oktober 2022.