

Implementasi Algoritma K-Nearest Neighbor(K-NN) Dalam Klasifikasi Kredit Motor

Efori Bu'ulolo*, Irma Suryani Tampubolon, Christin Vebiola Nababan, Lulu Nurhidayanti Nasution

Prodi Teknik Informatika, Universitas Budi Dharma, Medan, Indonesia

Email: buuloloeofori21@gmail.com

Abstrak—Penentuan kelayakan pengajuan kredit motor pada sebuah perusahaan leasing adalah hal penting, mengingat jika terjadi kesalahan pengambilan keputusan maka akan berdampak pada kerugian perusahaan FIF Group. Oleh karena itu penulis membuat Sistem Penunjang Keputusan dengan metode K-nearest Neighbor untuk menentukan kelayakan pengajuan kredit sepeda motor. Algoritma K-Nearest Neighbor (K-NN) merupakan salah satu algoritma dalam data mining untuk mengklasifikasikan objek yang baru berdasarkan mayoritas dari kategori tetangga yang terdekat. Algoritma clustering K-NN dengan menggunakan data penghasilan, pekerjaan, jumlah tanggungan dan kepemilikan rumah dapat mengelompokkan calon Pengkredit baru sehingga lebih memudahkan staf untuk menentukan penerimaan calon pengkredit sepeda motor yang baru.

Kata Kunci: K-Nearest Neighbor; Kredit; Motor; Klasifikasi

Abstract—Determining the eligibility of applying for a motorbike loan to a leasing company is important, considering that if an error in decision making occurs it will have an impact on the loss of the FIF Group company. Therefore the authors created a Decision Support System using the K-nearest Neighbor method to determine the feasibility of applying for a motorcycle loan. The K-Nearest Neighbor (K-NN) algorithm is an algorithm in data mining to classify new objects based on the majority of the nearest neighbor categories. The K-NN clustering algorithm using data on income, employment, number of dependents and home ownership can group prospective new creditors to make it easier for staff to determine acceptance of prospective new motorcycle creditors.

Keywords: K-Nearest Neighbor; Credit; Motorcycle; Classification

1. PENDAHULUAN

Banyak masyarakat yang ingin memiliki sepeda motor tetapi tidak memiliki dana yang cukup untuk membeli sepeda motor secara tunai. Tentu membutuhkan suatu cara untuk memiliki motor tetapi dengan cara sistem pembayarannya secara kredit. Para produsen motor memanfaatkan peluang ini untuk bersaing memberikan kredit yang mudah sehingga diminati banyak konsumen. Secara umum pemberian kredit pada sebuah perusahaan produsen didasarkan pada kegiatan analisa kredit. Untuk mendapatkan kredit motor konsumen harus mengikuti prosedur yang berlaku diperusahaan yang bersangkutan. Analisa yang dilakukan dengan cara interview, pengisian proposal dan kelengkapan data serta dilakukan survei lapangan. Kemudian produsen motor melakukan kegiatan administrasi kredit. Sejauh ini untuk dapat menentukan konsumen yang berhak mendapatkan kredit dengan administrasi yang meliputi kegiatan analisis secara kuantitatif dan kualitatif memerlukan waktu yang cukup lama karena data yang disajikan tidak berbentuk angka dan membutuhkan pertimbangan yang cukup lama.

Untuk membantu mengatasi masalah tersebut, maka penulis memanfaatkan data lama dalam kredit sepeda motor, agar pengguna yang dalam hal ini adalah manager, tidak kesulitan untuk menentukan memberikan kredit atau tidak. Serta mempersingkat pengerjaan waktu dalam analisa kredit. Beberapa masalah juga sering terjadi pada lembaga pengkreditan misalkan tunggakan konsumen yang sebelumnya dianggap layak menerima kredit, namun macet dalam pembayaran kredit. Munculnya masalah tersebut diakibatkan kurangnya pertimbangan atau kemantapan analisis kredit dalam menentukan kelayakan kredit. Oleh karena itu perlu dilakukan analisis kredit sehingga dapat mengetahui kelayakan dari suatu permasalahan kredit, melalui hasil analisis kreditnya, dapat diketahui apakah nasabah layak atau tidak untuk mendapatkan kredit. Salah satu solusinya adalah memanfaatkan data-data sebelumnya dan mencari informasi dari data tersebut dengan teknik data mining[1][2]. Salah teknik dalam data mining adalah algoritma k-nn. Algoritma k-nn adalah algoritma untuk mengklasifikasi objek berdasarkan mayoritas tetangga terdekat[3].

2. METODOLOGI PENELITIAN

Adapun metodologi penelitian yang dilakukan penelitian ini adalah sebagai berikut:

2.1 Dataset

Data set yang digunakan diperoleh dari FIF Group. Data yang digunakan data tahun 2019 dan data yang digunakan dibagi menjadi 2(dua) bagian yaitu data lama dan data baru. Data lama merupakan data nasabah yang sudah yang menggunakan jasa FIF Group dalam pengkreditan sepeda motor. Data baru merupakan data nasabah yang akan mengajukan pengkreditan sepeda motor. Antara data lama dan data baru mempunyai persamaan kriteria, perbedaannya hanya pada attribute keputusan, dimana dalam lama sudah mempunyai attribute keputusan sedangkan data baru belum punya keputusan.

2.2 Algoritma K-Nearest Neighbor(K-NN)

Algoritma K-Nearest Neighbor(K-NN) merupakan salah satu teknik data mining untuk mengklasifikasikan atau mengelompokkan objek yang baru berdasarkan mayoritas atau paling banyak dari kategori tetangga yang terdekat atau solusinya berdasarkan perbandingan data lama dengan data baru. Pada pembentukan klasifikasi berdasarkan attribute kriteria data lama dan data baru[4][5][6].



Gambar 1. Klasifikasi data dengan algoritma K-NN

Untuk perhitungan jarak menggunakan persamaan *Euclidean Distance* sebagai berikut:

$$d_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2} \quad (1)$$

Keterangan:

d_{ij} = Jarak dari data ke i ke pusat cluster j

x_{ki} = Data dari ke – i pada *attribute* data ke – k

x_{kj} = Data dari ke – j pada *attribute* data ke – k

3. HASIL DAN PEMBAHASAN

Berikut adalah sampel data lama yang digunakan dalam klasifikasi kredit sepeda motor:

Tabel 1. Data lama

No	Nama	Penda Patan	Status	Pendidikan	Pekerjaan	Tanggungan	Rumah	Keterangan
1	MAN	2 jt	Belum menikah	SLTA	Karyawan	0 orang	Kontrak	Ditolak
2	ESN	4 jt	Sudah menikah	S1	Pns	3 orang	Milik Instansi	Diterima
3	KN	3 jt	Sudah menikah	SLTA	Karyawan	1 orang	Milik Sendiri	Diterima
4	NT	4 jt	Belum menikah	S1	Wiraswasta	0 orang	Kontrak	Diterima
5	PT	2,5 jt	Sudah menikah	SLTA	Wiraswasta	3 orang	Milik Sendiri	Ditolak
6	EM	3 jt	Belum menikah	SLTA	Bumn	0 orang	Kontrak	Diterima
7	LS	3 jt	Sudah menikah	D3	Wiraswasta	1 orang	Keluarga	Diterima
8	ES	2 jt	Belum menikah	S1	Wiraswasta	5 orang	Milik Sendiri	Ditolak
9	MS	2,5 jt	Belum menikah	SLTA	Karyawan	3 orang	Milik Sendiri	Ditolak
10	DL	2 jt	Belum menikah	SLTA	Karyawan	0 orang	Milik Sendiri	Ditolak

3.1 Penetapan Nilai Kriteria

Algoritma K-NN tidak dapat memproses data dalam bentuk huruf oleh karena itu data dalam bentuk huruf dikonversi ke angka.

Tabel 2. Nilai variabel status

Status	Variabel Angka
Menikah	1
Belum Menikah	2

Tabel 3. Nilai variabel pendidikan

Pendidikan	Variabel Angka
S2	1
S1	2
D3	3
SLTA	4

Tabel 4. Nilai variabel pekerjaan

Pekerjaan	Variabel Angka
PNS	1
BUMN	2
Wiraswasta	3
karyawan	4

Tabel 5. Nilai variabel Tanggungan

Tanggungan	Variabel angka
0 orang	1
1-2 orang	2
3-4 orang	3
5-6 orang	4
>7 orang	5

Tabel 6. Nilai variabel rumah

Rumah	Variabel angka
Milik sendiri	1
keluarga	2
instansi	3
Kontrakan/Sewa	4

Tabel 7. Data Lama Setelah Nilai Attribute Dikonversi

No	Nama	Pendapatan	Status	Pendidikan	Pekerjaan	Tanggungan	Rumah	Keterangan
1	A	2000	2	4	4	0	4	Ditolak
2	B	4000	1	2	1	3	3	Diterima
3	C	3000	1	4	4	1	1	Diterima
4	D	4000	2	2	3	0	4	Diterima
5	E	2500	1	4	3	3	1	Ditolak
6	F	3000	2	4	2	0	4	Diterima
7	G	3000	1	3	3	1	2	Diterima
8	H	2000	2	2	3	5	1	Ditolak
9	I	2500	2	4	4	3	1	Ditolak
10	J	2000	2	4	1	0	1	Ditolak

Data baru merupakan data yang masih belum memiliki attribute keputusan, jadi melalui algoritma k-nn attribute keputusan data baru dicari.

Tabel 8. Data Baru

No	Nama	Pendapatan	Status	Pendidikan	Pekerjaan	Tanggungan	Rumah	Keterangan
1	MKS	5 jt	Menikah	S1	Karyawan	1 orang	Milik Sendiri	?

Data pada Tabel 8 attribute kriteria dalam bentuk huruf dikonversi keangka.

Tabel 9. Data Baru setelah di Konversi ke angka

No	Nama	Pendapatan	Status	Pendidikan	Pekerjaan	Tanggung ngan	Rumah	Keterangan
1	Z	5000 jt	1	2	4	1	1	?

Pencarian jarak antara data

$$d(x1,c) = \sqrt{(a1 - Ca)^2 + (b1 - Cb)^2 + (c1 - Cc)^2 + (d1 - Cd)^2 + (e1 - Ce)^2 + (f1 - Cf)^2}$$

$$= \sqrt{(2000 - 5000)^2 + (2 - 1)^2 + (4 - 2)^2 + (4 - 4)^2 + (0 - 1)^2 + (4 - 1)^2} = 3000,002$$

$$d(x2,c) = \sqrt{(a2 - Ca)^2 + (b2 - Cb)^2 + (c2 - Cc)^2 + (d2 - Cd)^2 + (e2 - Ce)^2 + (f2 - Cf)^2}$$

$$= \sqrt{(4000 - 5000)^2 + (1 - 1)^2 + (2 - 2)^2 + (1 - 4)^2 + (3 - 1)^2 + (3 - 1)^2} = 1000,008$$

$$d(x3,c) = \sqrt{(a3 - Ca)^2 + (b3 - Cb)^2 + (c3 - Cc)^2 + (d3 - Cd)^2 + (e3 - Ce)^2 + (f3 - Cf)^2}$$

$$= \sqrt{(3000 - 5000)^2 + (1 - 1)^2 + (4 - 2)^2 + (4 - 4)^2 + (1 - 1)^2 + (1 - 1)^2} = 2000,001$$

$$d(x4,c) = \sqrt{(a4 - Ca)^2 + (b4 - Cb)^2 + (c4 - Cc)^2 + (d4 - Cd)^2 + (e4 - Ce)^2 + (f4 - Cf)^2}$$

$$= \sqrt{(4000 - 5000)^2 + (2 - 1)^2 + (2 - 2)^2 + (3 - 4)^2 + (0 - 1)^2 + (4 - 1)^2} = 1000,001$$

$$d(x5,c) = \sqrt{(a5 - Ca)^2 + (b5 - Cb)^2 + (c5 - Cc)^2 + (d5 - Cd)^2 + (e5 - Ce)^2 + (f5 - Cf)^2}$$

$$= \sqrt{(2500 - 5000)^2 + (1 - 1)^2 + (4 - 2)^2 + (3 - 4)^2 + (3 - 1)^2 + (1 - 1)^2} = 2500,002$$

$$d(x6,c) = \sqrt{(a6 - Ca)^2 + (b6 - Cb)^2 + (c6 - Cc)^2 + (d6 - Cd)^2 + (e6 - Ce)^2 + (f6 - Cf)^2}$$

$$= \sqrt{(3000 - 5000)^2 + (2 - 1)^2 + (4 - 2)^2 + (2 - 4)^2 + (0 - 1)^2 + (4 - 1)^2} = 2000,005$$

$$d(x7,c) = \sqrt{(a7 - Ca)^2 + (b7 - Cb)^2 + (c7 - Cc)^2 + (d7 - Cd)^2 + (e7 - Ce)^2 + (f7 - Cf)^2}$$

$$= \sqrt{(3000 - 5000)^2 + (1 - 1)^2 + (3 - 2)^2 + (3 - 4)^2 + (1 - 1)^2 + (2 - 1)^2} = 2000,001$$

$$d(x8,c) = \sqrt{(a8 - Ca)^2 + (b8 - Cb)^2 + (c8 - Cc)^2 + (d8 - Cd)^2 + (e8 - Ce)^2 + (f8 - Cf)^2}$$

$$= \sqrt{(2000 - 5000)^2 + (2 - 1)^2 + (2 - 2)^2 + (3 - 4)^2 + (5 - 1)^2 + (1 - 1)^2} = 3000,003$$

$$d(x9,c) = \sqrt{(a9 - Ca)^2 + (b9 - Cb)^2 + (c9 - Cc)^2 + (d9 - Cd)^2 + (e9 - Ce)^2 + (f9 - Cf)^2}$$

$$= \sqrt{(2500 - 5000)^2 + (2 - 1)^2 + (4 - 2)^2 + (4 - 4)^2 + (3 - 1)^2 + (1 - 1)^2} = 2500,002$$

$$d(x10,c) = \sqrt{(a10 - Ca)^2 + (b10 - Cb)^2 + (c10 - Cc)^2 + (d10 - Cd)^2 + (e10 - Ce)^2 + (f10 - Cf)^2}$$

$$= \sqrt{(2000 - 5000)^2 + (2 - 1)^2 + (4 - 2)^2 + (4 - 4)^2 + (0 - 1)^2 + (1 - 1)^2} = 3000,009$$

Setelah dihitung jaraknya, selanjutnya urutkan nilai dari yang terkecil ke nilai yang terbesar seperti berikut:

Tabel 10. Data Jarak

No	Nama	Pendap atan	Status	Pendidikan	Pekerjaan	Tanggu ngan	Rumah	Ket	Jarak
4	D	4000	2	2	3	0	4	Diterima	1000,001
2	B	4000	1	2	1	3	3	Diterima	1000,008
3	C	3000	1	4	4	1	1	Diterima	2000,001
6	F	3000	2	4	2	0	4	Diterima	2000,001
7	G	3000	1	3	3	1	2	Diterima	2000,001
5	E	2500	1	4	3	3	1	Ditolak	2500,002
9	I	2500	2	4	4	3	1	Ditolak	2500,002
1	A	2000	2	4	4	0	4	Ditolak	3000,002
9	I	2500	2	4	4	3	1	Ditolak	2500,002
10	J	2000	2	4	1	0	1	Ditolak	3000,009

Dari perhitungan pada tabel 10, ternyata 4 (empat) tetangga terdekat dari calon pengkredit baru, Keterangannya menyatakan “Diterima”, maka dapat disimpulkan calon nasabah baru yang bernama “MKS / Z”, Pengajuan kredit yang diajukan Diterima.

4. KESIMPULAN

Berdasarkan uraian diatas, maka kesimpulan dari penelitian ini adalah metode K-Nearest Neighbor telah dapat digunakan untuk menentukan kelayakan kredit sepeda motor. Algoritma K-Nearest Neighbor(K-NN) adalah salah satu algoritma dalam data mining untuk mengklasifikasikan objek yang baru berdasarkan mayoritas dari kategori tetangga yang terdekat.

REFERENCES

- [1] A. P. Windarto, "Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method," *Int. J. Artif. Intell. Res.*, vol. 1, no. 2, p. 26, 2017, doi: 10.29099/ijair.v1i2.17.
- [2] E. Buulolo and R. Syahputra, "Implementasi Algoritma Clustering K-Means Untuk Mengelompokkan Mahasiswa Baru Yang Berpotensi (Studi Kasus : Stmik Budi Darma)," vol. 2, no. September, pp. 17–24, 2019.
- [3] I. Listiowarni and N. Puspa Dewi, "Pemanfaatan Klasifikasi Soal Biologi Cognitive Domain Bloom's Taxonomy Menggunakan KNN Chi-Square Sebagai Penyusunan Naskah Soal," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 11, no. 2, pp. 186–197, 2020, doi: 10.31849/digitalzone.v11i2.4798.
- [4] D. Nofriansyah and G. Widi Nurcahyo, *Buku Algoritma Data Mining Dan Pengujian*. Yogyakarta: deepublish, 2017.
- [5] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Yogyakarta: deepublish, 2020.
- [6] M. M. K-means, A. Solichin, and K. Khairunnisa, "Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta," vol. 5, no. 2, 2020.