

Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Pemberian Kredit

Application of the K-Nearest Neighbor Algorithm for Determining Credit Grants

Retno Wahyusari^{*1}, Dwi Purnomo Putro²

^{1,2}Program Studi Informatika, Sekolah Tinggi Teknologi Ronggolawe

e-mail: *1retnowahyusari@gmail.com, dwipurnomoputro@sttrcepu.ac.id

Abstrak - Penelitian ini bertujuan untuk menerapkan algoritma K-Nearest Neighbor (K-NN) dalam menentukan pemberian kredit, yang merupakan aspek penting dalam pengelolaan risiko kredit di lembaga keuangan. Algoritma K-NN dipilih karena kemampuannya yang kuat dalam klasifikasi dan prediksi berdasarkan kemiripan data. Dalam studi ini, kami menguji kinerja K-NN dengan variasi parameter K untuk mengidentifikasi jumlah tetangga terdekat yang optimal guna mencapai akurasi tertinggi. Data yang digunakan dalam penelitian ini terdiri dari sekumpulan data historis pemohon kredit yang mencakup variabel-variabel seperti usia, jenis kelamin, pendapatan, pendidikan, jumlah anak, dan status kepemilikan rumah. Hasil eksperimen menunjukkan bahwa dengan jumlah K=3, algoritma K-NN mencapai tingkat akurasi sebesar 98,36%. Lebih lanjut, dengan jumlah K=2, akurasi meningkat hingga 100%, yang menunjukkan kemampuan algoritma dalam mengklasifikasikan data dengan presisi tinggi pada kondisi tertentu.

Kata kunci: K-Nearest Neighbor, pemberian kredit, klasifikasi, akurasi, pembelajaran mesin.

Abstract - This research aims to apply the K-Nearest Neighbor (K-NN) algorithm in determining credit provision, which is an important aspect in managing credit risk in financial institutions. The K-NN algorithm was chosen because of its strong ability in classification and prediction based on data similarity. In this study, we test the performance of K-NN with varying K parameters to identify the optimal number of nearest neighbors to achieve the highest accuracy. The data used in this research consists of a set of historical data on credit applicants which includes variables such as age, gender, income, education, marital status, number of children, and home ownership. Experimental results show that with the number K=3, the K-NN algorithm achieves an accuracy level of 98.36%. Furthermore, with the number K=2, accuracy increases to 100%, which shows the algorithm's ability to classify data with high precision under certain conditions.

Keywords: K-Nearest Neighbor, crediting, classification, accuracy, machine learning

I. PENDAHULUAN

Dalam dunia perbankan, penentuan kelayakan kredit merupakan salah satu aspek krusial yang mempengaruhi profitabilitas dan keberlanjutan bisnis. Kelayakan kredit adalah proses

evaluasi untuk menentukan apakah calon peminjam layak menerima kredit berdasarkan berbagai faktor risiko [1]. Seiring perkembangan teknologi informasi, penerapan algoritma dan teknik data mining dalam penentuan kelayakan kredit telah menjadi semakin umum. Salah satu algoritma yang banyak digunakan dalam klasifikasi dan prediksi adalah algoritma K-Nearest Neighbor (K-NN) [2].

Algoritma K-Nearest Neighbor merupakan metode yang sederhana namun efektif dalam memecahkan masalah klasifikasi. Prinsip dasar algoritma ini adalah mengklasifikasikan sebuah data berdasarkan mayoritas kategori dari k tetangga terdekatnya dalam ruang fitur [3]. Dalam konteks pemberian kredit, algoritma K-NN dapat digunakan untuk mengklasifikasikan calon peminjam ke dalam kategori layak atau tidak layak berdasarkan data historis dari peminjam sebelumnya [4].

Penerapan algoritma K-NN dalam penentuan kelayakan kredit menawarkan berbagai keuntungan. Pertama, algoritma ini tidak memerlukan model parametrik yang kompleks, sehingga implementasinya relatif mudah [5]. Kedua, K-NN memiliki fleksibilitas yang tinggi karena dapat digunakan untuk data dengan berbagai jenis atribut. Ketiga, algoritma ini mampu menangani data yang bersifat non-linear, yang seringkali ditemukan dalam dataset kelayakan kredit [6].

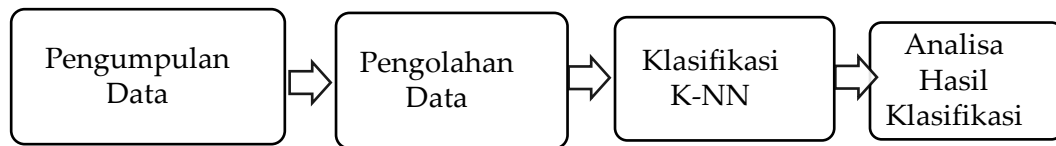
Namun, meskipun K-NN memiliki berbagai keunggulan, terdapat beberapa tantangan dalam penerapannya. Salah satu tantangan utama adalah pemilihan nilai k yang optimal. Nilai k yang terlalu kecil dapat menyebabkan model overfitting, sementara nilai k yang terlalu besar dapat menyebabkan model underfitting [7]. Selain itu, K-NN juga sensitif terhadap ukuran dataset dan keberadaan outlier [8]. Oleh karena itu, perlu dilakukan studi mendalam untuk mengevaluasi kinerja algoritma ini dalam konteks penentuan kelayakan kredit.

Seiring dengan peningkatan jumlah data yang tersedia dan kemajuan dalam teknik *big data*, penggunaan algoritma K-NN dalam penentuan kelayakan kredit menjadi semakin relevan. Dengan memanfaatkan data historis yang besar dan beragam, algoritma ini dapat membantu bank dalam membuat keputusan kredit yang lebih akurat dan efisien [9]. Selain itu, penerapan K-NN juga dapat mengurangi risiko kredit macet dengan mengidentifikasi calon peminjam yang berisiko tinggi [10].

Di sisi lain, perkembangan teknologi komputasi dan akses terhadap data besar memungkinkan penerapan K-NN pada skala yang lebih luas dan dengan kecepatan yang lebih tinggi. Hal ini membuka peluang bagi bank untuk mengintegrasikan algoritma ini dalam sistem informasi mereka secara real-time, sehingga dapat memberikan keputusan kredit secara cepat dan tepat waktu [11]. Dalam jangka panjang, hal ini dapat meningkatkan kepercayaan pelanggan dan memperkuat posisi kompetitif bank di pasar [12].

II. METODE

Penelitian dilakukan dengan beberapa tahapan yaitu pengumpulan data, pengolahan data, dilanjutkan dengan mengklasifikasikan data menggunakan algoritma K-NN, langkah terakhir menganalisa hasil klasifikasi. Gambar alur penelitian tersaji pada Gambar 1.



Gambar 1. Alur Penelitian

Proses pengolahan data dilakukan dengan mencari data pada data set Kaggel.com, data pada penelitian berjumlah 61 data dengan variabel usia, jenis kelamin, pendapatan, pendidikan, jumlah anak, dan status kepemilikan rumah. Data yang didapat kemudian dilakukan pengolahan data berupa pembersihan data yang tidak lengkap. Proses klasifikasi dilakukan menggunakan algoritma K-NN, dimana pengujian menggunakan jumlah k sebanyak 2 dan 3. Adapun algoritma K-NN, sebagai berikut:

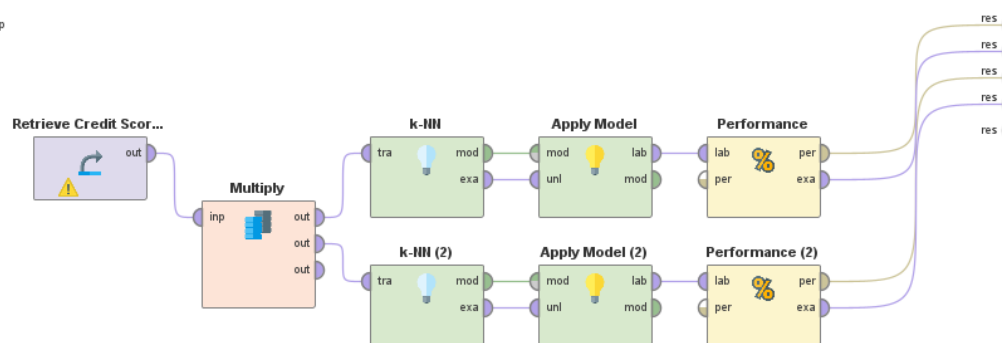
1. Menentukan jumlah K neighbor atau terdekat.
2. Hitung jarak Euclidean dari bilangan K terdekat.
3. Dapatkan K terdekat ini berdasarkan jarak Euclidean yang dihitung.

$$dis = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2 + (y_{1i} - y_{2i})^2 + \dots} \dots\dots\dots(1)$$

4. Hitung jumlah titik data di setiap kategori K titik data terdekat ini.
5. Menetapkan titik data baru ke kategori dengan jumlah tetangga terbanyak.
6. Model yang dibuat kini siap.

III. HASIL DAN PEMBAHASAN

Penelitian dilakukan menggunakan aplikasi rapidminer. Langkah pertama memasukkan data didalam inputan data. Setelah data dimasukkan lalu double klik lalu ditahan dan digeser sampai lembar kerja. Selanjutnya ditambahkan kolom multhy dan K-NN. Disusun hingga tersaji pada Gambar 2.



Gambar 2. Model Diusulkan

Penentuan jumlah K dengan memasukkan nilai K sesuai yang diinginkan, disini peneliti memasukkan nilai 3 dan 2. Konfigurasi yang telah didapat kemudian dijalankan dengan hasil K-

3 hasil akurasi yang didapat sebesar 98,36%. Sedangkan untuk K-2 tingkat akurasi sebesar 100%. Menggunakan jumlah K-3 terdapat 1 data yang salah dalam proses klasifikasi, dimana harusnya masuk dalam kelompok resiko *average*, masuk ke dalam klasifikasi resiko *high*. Tabel 1 merupakan hasil tingkat akurasi.

Tabel 1. Tingkat Akurasi Hasil Klasifikasi

Akurasi %	Jumlah K	
	3	2
Tingkat akurasi	98,36	100

IV. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma K-Nearest Neighbor (K-NN) dalam proses penentuan pemberian kredit, yang merupakan aspek penting dalam pengelolaan risiko kredit di lembaga keuangan. Algoritma K-NN dipilih karena kemampuannya yang kuat dalam melakukan klasifikasi dan prediksi berdasarkan kemiripan data. Melalui studi ini, kami menguji kinerja K-NN dengan variasi parameter K untuk menentukan jumlah tetangga terdekat yang optimal guna mencapai akurasi tertinggi. Data yang digunakan dalam penelitian ini mencakup sekumpulan data historis pemohon kredit dengan variabel-variabel seperti usia, jenis kelamin, pendapatan, pendidikan, jumlah anak, dan status kepemilikan rumah. Hasil eksperimen menunjukkan bahwa dengan jumlah K=3, algoritma K-NN mencapai tingkat akurasi sebesar 98,36%. Lebih lanjut, ketika jumlah K diatur menjadi 2, akurasi meningkat hingga 100%, yang mengindikasikan kemampuan algoritma K-NN untuk mengklasifikasikan data dengan presisi tinggi dalam kondisi tertentu.

DAFTAR PUSTAKA

- [1] S. Jones and D. A. Hensher, "Predicting firm financial distress: A mixed-methods approach," *Economic Modelling*, vol. 88, pp. 450-464, 2020.
- [2] A. Laha and S. Biswas, "Credit scoring by modeling longitudinal consumer behavior," *Journal of the Operational Research Society*, vol. 66, no. 2, pp. 287-298, 2015.
- [3] J. J. Martínez, C. J. Pérez, A. Suárez, and A. Fernández, "A proposal for credit scoring with reject inference in imbalanced datasets," *Expert Systems with Applications*, vol. 101, pp. 175-185, 2018.
- [4] T. Fitzpatrick and C. Mues, "An empirical comparison of classification algorithms for mortgage default prediction: Evidence from a distressed mortgage market," *European Journal of Operational Research*, vol. 249, no. 2, pp. 427-439, 2016.
- [5] F. Provost and T. Fawcett, *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*, Sebastopol: O'Reilly Media, 2014.
- [6] W. Zhao and J. Wang, "Prediction of credit default using dynamic information and gradient boosting decision tree," *Procedia Computer Science*, vol. 55, pp. 1210-1216, 2015.
- [7] Y. Kim and S. Y. Sohn, "Random-effects model for credit scoring using longitudinal consumer information," *Expert Systems with Applications*, vol. 62, pp. 111-119, 2016.
- [8] S. Jones and D. A. Hensher, "Predicting firm financial distress: A mixed-methods approach," *Economic Modelling*, vol. 88, pp. 450-464, 2020.
- [9] L. Nguyen and W. Wu, "Ensemble learning for credit scoring: A case study," *Journal of Risk and Financial Management*, vol. 14, no. 6, p. 265, 2021.
- [10] F. Barboza, H. Kimura, and E. Altman, "Machine learning models and bankruptcy prediction,"

- Expert Systems with Applications, vol. 83, pp. 405-417, 2017.
- [11] Y. Huang, X. Zhou, and Z. H. Zhou, "Cost-sensitive credit risk assessment with classifier ensembles," IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 6, pp. 2833-2845, 2022.
- [12] Y. Wang and Z. Y. Dong, "Machine learning-based credit scoring models: A review and empirical study," Journal of Financial Stability, vol. 57, p. 100962, 2023.