



The Application of K-Nearest Neighbors Algorithm in Creditworthiness Evaluation: A Case Study on Bank ABC

Penerapan Algoritma K-Nearest Neighbors dalam Evaluasi Kelayakan Kredit: Studi Kasus pada Bank ABC

Hamzah Naufal Zuhdi^{1*}, Budi Prasetyo²

¹System Information, Faculty of Mathematics and Science, Universitas Negeri Semarang, Indonesia

²System Information, Faculty of Mathematics and Science, Universitas Negeri Semarang, Indonesia

E-Mail: ¹hamzahnaufal@students.unnes.ac.id, ²bprasetyo@maill.unnes.ac.id

Makalah: Diterima 18 Februari 2024; Diperbaiki 30 Februari 2024; Disetujui 20 Maret 2024
Corresponding Author: Hamzah Naufal Zuhdi

Abstrak

Dalam era digital yang terus berkembang, teknologi telah menjadi tulang punggung bagi banyak sektor, termasuk sektor perbankan. Salah satu tantangan utama yang dihadapi oleh sektor perbankan adalah evaluasi kelayakan kredit. Evaluasi ini merupakan proses penting yang menentukan apakah seorang nasabah layak menerima kredit atau tidak. Salah satu metode yang dapat digunakan untuk membantu proses ini adalah algoritma K-Nearest Neighbors (KNN). Artikel ini membahas penerapan algoritma K-Nearest Neighbors dalam evaluasi kelayakan kredit, dengan studi kasus pada Bank ABC. Studi ini bertujuan untuk menelaah penerapan algoritma KNN dalam evaluasi kelayakan kredit, dengan fokus pada penggunaan teknologi untuk meningkatkan proses evaluasi kredit. Dengan memahami tantangan yang dihadapi oleh sektor perbankan dan potensi solusi teknologi seperti algoritma KNN, penelitian ini diharapkan dapat memberikan wawasan berharga bagi praktisi industri perbankan dan para peneliti di bidang data science. Dari hasil pengujian dengan variasi nilai K, didapatkan bahwa model memiliki akurasi sekitar 93,33% - 95,00%. terdapat sedikit perbedaan dalam akurasi antara nilai K yang berbeda, namun secara umum model KNN menunjukkan kinerja yang baik dalam mengklasifikasikan nasabah berdasarkan kelayakan kredit

Keyword: K-Nearest Neighbors, evaluasi, kelayakan kredit, sektor perbankan, Confusion matrix.

Abstract

In the continuously evolving digital era, technology has become the backbone of many sectors, including the banking sector. One of the primary obstacles encountered by the banking industry involves assessing the creditworthiness of individuals. This assessment constitutes a critical procedure that establishes whether a client qualifies for credit facilities or not. One of the methods that can assist in this process is the K-Nearest Neighbors (KNN) algorithm. The article explores the utilization of the K-Nearest Neighbors algorithm for assessing creditworthiness, exemplified through a case study conducted at Bank ABC. It delves into the application of the KNN algorithm within creditworthiness evaluation, emphasizing the utilization of technology to augment the credit assessment procedure. Through comprehending the obstacles encountered by the banking industry and potential technological remedies like the KNN algorithm, this study anticipates offering valuable perspectives for professionals within the banking sector and researchers specializing in data science.

Keyword: K-Nearest Neighbors, evaluation, creditworthiness, banking sector, Confusion matrix.

1. PENDAHULUAN

Dalam era digital yang terus berkembang, teknologi telah menjadi tulang punggung bagi banyak sektor, termasuk sektor perbankan. Salah satu tantangan utama yang dihadapi oleh sektor perbankan adalah evaluasi kelayakan kredit. Evaluasi ini merupakan proses penting yang menentukan apakah seorang nasabah layak menerima kredit atau tidak. Salah satu metode yang dapat digunakan untuk membantu proses ini adalah algoritma K-Nearest Neighbors (KNN) [1]. Algoritma KNN adalah metode yang digunakan untuk klasifikasi dan regresi. Dalam konteks evaluasi kelayakan kredit, algoritma ini dapat digunakan untuk mengklasifikasikan

nasabah menjadi dua kelas, yaitu layak menerima kredit dan tidak layak menerima kredit [1]. Algoritma ini telah terbukti efektif dalam berbagai studi kasus, termasuk dalam penentuan risiko kredit [2].

Artikel ini akan membahas penerapan algoritma K-Nearest Neighbors dalam evaluasi kelayakan kredit dengan studi kasus pada Bank ABC. Studi ini bertujuan untuk meneliti penerapan algoritma KNN dalam evaluasi kelayakan kredit, dengan studi kasus yang dilakukan di Bank ABC. Dengan fokus pada penggunaan teknologi untuk meningkatkan proses evaluasi kredit, penelitian ini memiliki tujuan ganda: pertama, untuk mengevaluasi efektivitas algoritma KNN dalam mengklasifikasikan nasabah berdasarkan kelayakan kredit, dan kedua, untuk memberikan pemahaman yang lebih dalam tentang potensi teknologi dalam mendukung pengambilan keputusan kredit di lingkungan perbankan.

Dengan memahami tantangan yang dihadapi oleh sektor perbankan dan potensi solusi teknologi seperti algoritma KNN, penelitian ini diharapkan dapat memberikan pemahaman yang berharga bagi para profesional di industri perbankan serta bagi para akademisi yang berfokus pada ilmu data science. Melalui penerapan konsep-konsep ini, diharapkan bank dapat meningkatkan efisiensi operasional, mengurangi risiko kredit, dan secara keseluruhan, meningkatkan pengalaman nasabah dalam layanan perbankan.

2. MATERIALS AND METHOD

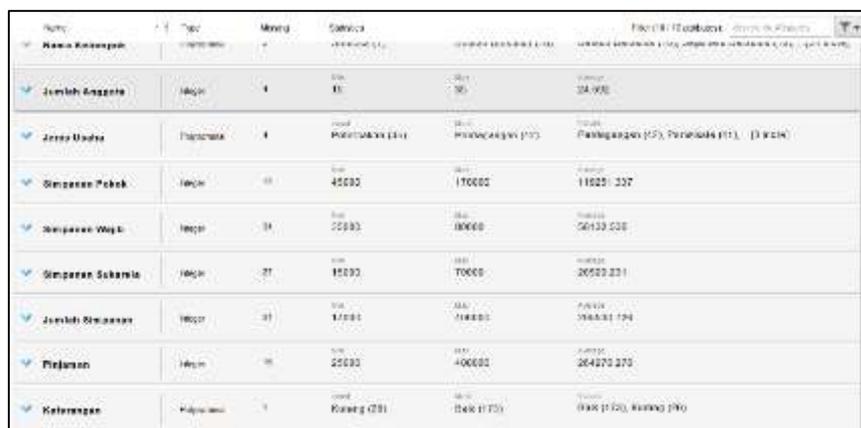
Artikel ini menerapkan metode penelitian eksperimental dengan menerapkan kerangka kerja Cross-Industry Standard Process for Data Mining (CRISP-DM). Pendekatan ini mencakup enam langkah proses yang terdiri dari:

2.1 Business Understanding

Selanjutnya, dilakukan eksplorasi untuk memperoleh pemahaman mendalam serta menetapkan tujuan dari penelitian. Dalam konteks penelitian ini, tujuannya adalah untuk menerapkan algoritma K-Nearest Neighbors dalam evaluasi kelayakan kredit di Bank ABC. Ini melibatkan pemahaman tentang bagaimana proses evaluasi kredit saat ini berjalan di bank, apa saja tantangan dan hambatan yang ada, dan bagaimana teknologi dan data science dapat membantu dalam mengatasi tantangan tersebut.

2.2 Data Understanding

Tahapan ini berfokus pada pengumpulan data, mengevaluasi kualitas data, dan memahami karakteristik data yang tersedia. Dalam penelitian ini, terdapat 200 data dengan sembilan atribut dan satu kelas. Kesembilan atribut ini mencakup Nomor Anggota, Nama Pemimpin Kelompok, Nama Kelompok, Jumlah Anggota Kelompok, Tipe Usaha, Simpanan Pokok, Simpanan Wajib, Simpanan Sukarela, Jumlah Simpanan, dan Pinjaman, bersama dengan satu atribut kelas yang bertindak sebagai label pada data tersebut. Dari 200 data, sebanyak 174 dengan label kredit Baik, dan 26 data dengan label kredit Kurang baik.

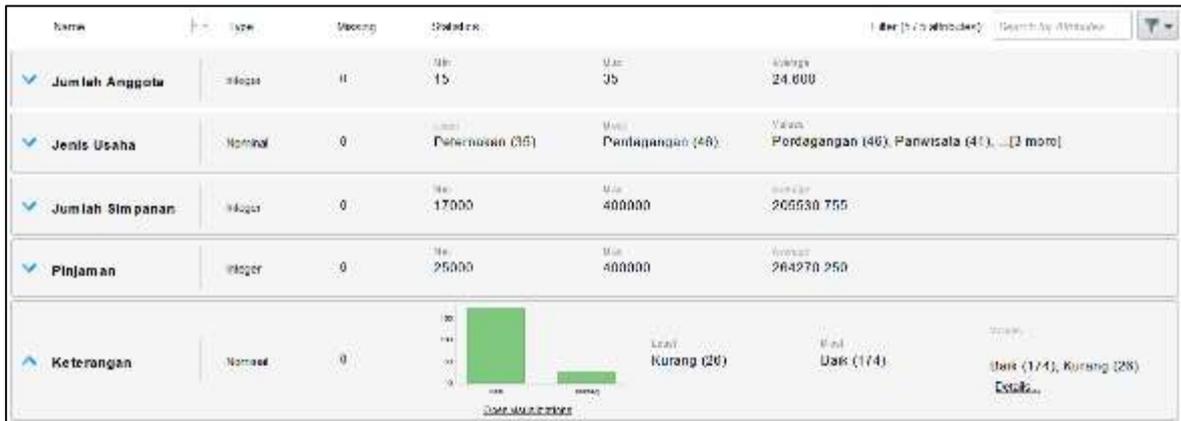


Statistik	Tipe	Min	Max	Q1	Q3	Rata-rata
Jumlah Anggota	Integer	4	30	10	20	14.999
Jenis Usaha	Kategori	1	3	1	2	1.999
Simpanan Pokok	Integer	45000	110000	60000	90000	75000
Simpanan Wajb.	Integer	25000	80000	40000	60000	50000
Simpanan Sukarela	Integer	0	70000	0	30000	20000
Jumlah Riungan	Integer	10	30	15	25	19.999
Pinjaman	Integer	25000	100000	40000	70000	55000
Keterangan	Kategori	1	2	1	2	1.999

Gambar 1. Statistik awal dataset

2.3 Data Preparation

Tahap ini melibatkan proses membersihkan, mengintegrasikan, dan mentransformasi data. Tujuannya adalah untuk mempersiapkan data secara optimal agar dapat digunakan dalam analisis menggunakan algoritma KNN. Pada langkah ini, dilakukan pembersihan data dengan menghapus entri yang duplikat serta pengisian atribut yang memiliki nilai yang kosong atau hilang. Selain itu, dilakukan juga seleksi atribut yang tidak berpengaruh terhadap analisis.



Gambar 2. Statistik Setelah dilakukan data preparation

Untuk melakukan pemodelan dengan algoritma K-Nearest Neighbors (K-NN), diperlukan data yang memiliki kelas atau target yang direpresentasikan dalam bentuk nilai numerik. Setiap entri data harus memiliki label numerik yang menjadi dasar untuk menentukan tetangga terdekat dalam proses pengklasifikasian. Detail lebih lanjut dapat ditemukan dalam Tabel 1, yang mencatat data dan hasil pembersihan yang telah dilakukan.

Table 1. Contoh Data Hasil Pembersihan

Jumlah Anggota	Jenis Usaha	Jumlah Simpanan	Pinjaman	Keterangan
20	Pertanian	170000	300000	Baik
15	Kerajinan	135000	200000	Baik
25	Peternakan	205000	250000	Kurang
30	Perdagangan	250000	350000	Baik
18	Pariwisata	153000	180000	Baik
22	Pertanian	187000	280000	Baik
16	Kerajinan	127000	150000	Kurang
28	Peternakan	224000	320000	Baik
35	Perdagangan	285000	400000	Baik
20	Pariwisata	170000	200000	Baik
25	Pertanian	205000	250000	Baik
30	Kerajinan	250000	350000	Baik
18	Peternakan	153000	180000	Baik
22	Perdagangan	187000	280000	Baik
16	Pariwisata	127000	150000	Kurang
28	Pertanian	224000	320000	Baik
35	Kerajinan	285000	400000	Baik
20	Peternakan	170000	200000	Baik
25	Perdagangan	205000	250000	Baik
30	Pariwisata	250000	350000	Baik

2.4 Modelling Measurement

Confusion matrix akan digunakan untuk menguji model yang dihasilkan dari data pelatihan. Confusion matrix atau Error Matrix adalah representasi visual dari performa model dalam memprediksi kelas atau label data uji. Detail mengenai struktur dan isi confusion matrix atau error matrix dapat ditemukan dalam tabel berikut ini.

Table 2 Confusion Matrix (Error Matrix)
Actual Values

		Actual Values	
		Positive	Negative
Predicted Values	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

2.5 Evaluation

Dalam penelitian ini, evaluasi dilakukan dengan membagi dataset menjadi dua bagian, yakni data training dan data testing, dengan rasio perbandingan 7:3. Data set awal digunakan sebagai bahan pelatihan untuk model KNN, dimanfaatkan untuk menguji performa dari model yang telah dilatih. Tindakan ini memastikan bahwa evaluasi model dilakukan secara objektif dengan menggunakan dataset yang belum pernah digunakan sebelumnya, sehingga hasil evaluasi dapat diandalkan untuk menilai kinerja model. Evaluasi dilakukan dengan memanfaatkan beberapa metrik berikut:

Sens (Sensitivitas) $= \frac{TP}{TP+FN}$ Equation (1)

NPV (Negative Predictive Value) $= \frac{TN}{TN+FN}$ Equation (2)

PPV (Positive Predictive Value) $= \frac{TP}{TP+FP}$ Equation (3)

Spec (Specificity) $= \frac{TN}{TN+FP}$ Equation (4)

Acc (Accuracy) $= \frac{TP+TN}{TP+FP+TN+FN}$ Equation (5)

Selanjutnya, untuk mengilustrasikan tingkat akurasi diskriminatif dari suatu pengujian, kita dapat menggunakan Area Under Curve (AUC). Berikut adalah klasifikasi berdasarkan nilai AUC:

0.90 – 1.00 = Sangat Unggul

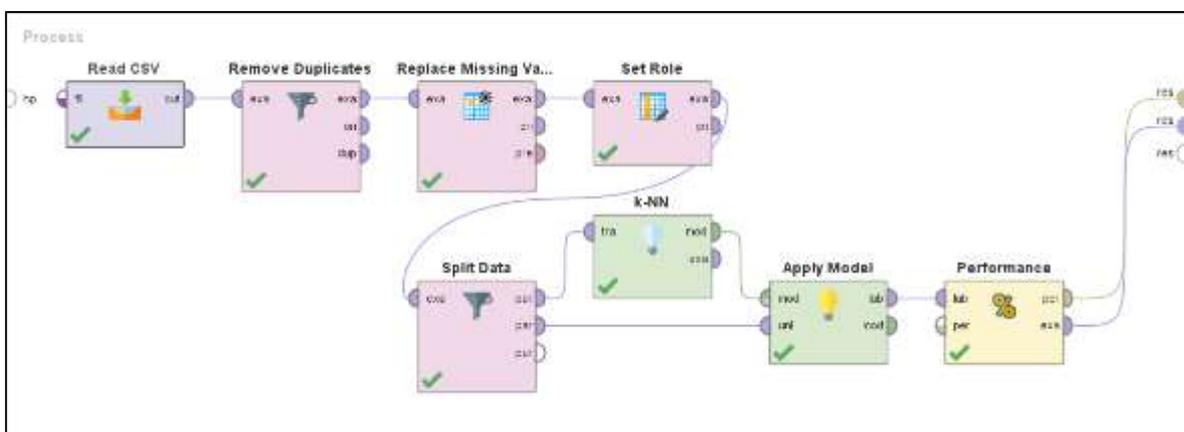
0.80 – 0.90 = Unggul

0.70 – 0.80 = Memadai

0.60 – 0.70 = Kurang

0.50 – 0.60 = Tidak Lulus

AUC memberikan indikasi seberapa efektif model dalam membedakan antara kelas positif dan negatif. Semakin tinggi nilai AUC, semakin efektif model dalam melakukan klasifikasi dengan akurat.



Gambar 3. Desain Model Pengujian menggunakan Rapidminer

3. ANALISA DAN HASIL

Model K-NN dalam penelitian ini diuji dengan berbagai variasi nilai K, diantaranya K=1, K=3, K=5, K=7, dan K=9.

Dengan nilai K=1, model mencapai tingkat akurasi sebesar 95,00%. Dalam hal ini, model berhasil mengidentifikasi 52 data sebagai True Positive, 3 data sebagai False Positive, 5 data sebagai True Negative, dan 0 data sebagai False Negative.

Table 3. *Confusion Matrix K-NN K=1*

	Positive	Negative
Positive	52	3
Negative	0	5

Untuk nilai K=3, model menunjukkan akurasi sebesar 95,00 %. Dalam hal ini, model berhasil mengidentifikasi 52 data sebagai True Positive, 3 data sebagai False Positive, 5 data sebagai True Negative, dan 0 data sebagai False Negative.

Table 4. *Confusion Matrix K-NN K=3*

	Positive	Negative
Positive	52	3
Negative	0	5

Dengan menggunakan nilai K=5, model mencapai tingkat akurasi sebesar 93,33%. Dalam kasus ini, model berhasil mengidentifikasi 51 data sebagai True Positive, 3 data sebagai False Positive, 5 data sebagai True Negative, dan 1 data sebagai False Negative.

Table 5. *Confusion Matrix K-NN K=5*

	Positive	Negative
Positive	51	3
Negative	1	5

Untuk nilai K=7, model menunjukkan akurasi sebesar 93,33 %. Dalam hal ini, model berhasil mengidentifikasi 51 data sebagai True Positive, 3 data sebagai False Positive, 5 data sebagai True Negative, dan 1 data sebagai False Negative.

Table 6. *Confusion Matrix K-NN K=7*

	Positive	Negative
Positive	51	3
Negative	1	5

Untuk nilai K=9, model menunjukkan akurasi sebesar 93,33 %. Dalam hal ini, model berhasil mengidentifikasi 51 data sebagai True Positive, 3 data sebagai False Positive, 5 data sebagai True Negative, dan 1 data sebagai False Negative.

Table 7. *Confusion Matrix K-NN K=9*

	Positive	Negative
Positive	51	3
Negative	1	5

4. KESIMPULAN

Kesimpulan dari penelitian ini menunjukkan bahwa penerapan algoritma K-Nearest Neighbors (KNN) dalam evaluasi kelayakan kredit di Bank ABC telah menghasilkan model yang memiliki tingkat akurasi yang konsisten dan cukup tinggi. Dari hasil pengujian dengan variasi nilai K, didapatkan bahwa model memiliki akurasi sekitar 93,33% - 95,00%. Meskipun terdapat sedikit perbedaan dalam akurasi antara nilai K yang berbeda, namun secara umum model KNN menunjukkan kinerja yang baik dalam mengklasifikasikan nasabah berdasarkan kelayakan kredit.

Melalui penggunaan confusion matrix, kita dapat melihat bahwa model memiliki tingkat True Positive yang cukup tinggi, yang menunjukkan kemampuan model dalam mengidentifikasi nasabah yang layak menerima kredit dengan benar. Selain itu, False Positive dan False Negative yang rendah juga menandakan bahwa model cenderung melakukan sedikit kesalahan dalam mengklasifikasikan nasabah.

Dalam konteks pengambilan keputusan kredit, hasil ini menunjukkan bahwa algoritma KNN bisa menjadi alat yang efektif dalam mendukung penilaian kelayakan kredit di Bank ABC. Namun, penting untuk diingat bahwa penggunaan algoritma ini tidak bersifat mutlak, dan harus dipertimbangkan bersama dengan faktor-faktor lain seperti kebijakan internal bank, regulasi, dan analisis risiko secara menyeluruh.

Saran untuk penelitian selanjutnya adalah untuk melakukan pengujian model dengan dataset yang lebih besar dan lebih beragam, serta mempertimbangkan penggunaan fitur atau atribut tambahan yang dapat meningkatkan kinerja model. Selain itu, memperluas cakupan penelitian untuk melibatkan beberapa bank atau lembaga keuangan lainnya juga dapat memberikan pemahaman yang lebih komprehensif tentang keefektifan algoritma KNN dalam menilai kelayakan kredit di sektor perbankan secara keseluruhan.

REFERENCES

- [1] S. Harlina, Suryani and M. O. Kadang, "Penerapan Algoritma K-Nearest Neighbor untuk Klasifikasi Kelayakan Calon Nasabah Kredit Berbasis Web," *Prosiding Sintaks*, pp. 1-10, 2020.
- [2] T. T. Muryono and Irwansyah, "Implementasi Data Mining untuk Menentukan Kelayakan Pemberian Kredit dengan Menggunakan Algoritma K-Nearest Neighbors (K-NN)," *Journal of Technology Information*, pp. 1-10, 2020.
- [3] A. Tripathy, A. Agrawal and S. K. Rath, "Classification of Sentimental Reviews Using Machine Learning Techniques," *Procedia Computer Science*, pp. 821-829, 2015.
- [4] M. W. Ohsaki, K. K. Matsuda, H. Watanabe and A. Ralescu, "Confussion Matrix Based Kernel Logistic Regression for Imbalanced Data Classification," *IEEE Transaction on Knowledge and Data Engineering*, pp. 1806-1819, 2017.
- [5] T. Wahyono and A.D. Cahyono, "Berbasis Komputasional untuk Meningkatkan Kemampuan," *Jurnal Komputer*, pp. 11-18, 2015.
- [6] P. Sri Astuti, M. N. Hayati and R. Goejantoro, "ANALISIS CREDIT SCORING TERHADAP STATUS PEMBAYARAN BARANG ELEKTRONIK DAN FURNITURE MENGGUNAKAN BOOTSTRAP AGGREGATING K-NEAREST NEIGHBOR," *BAREKENG JURNAL ILMU MATEMATIKA DAN TERAPAN*, 2021.
- [7] Ida, S. H. Baharuddin, M. Faisal, N. Ramadhan and Darniati, "IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR TERHADAP PENENTUAN RISIKO KREDIT USAHA MIKRO KECIL DAN MENENGAH," *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, 2023.
- [8] M. Goldstein, "Kn -nearest Neighbor Classification," *IEEE Transactions on Information Theory*, 1972.
- [9] E. A. P. A. B. Gustavo and D. F. Silva, "How k-nearest neighbor parameters affect its performance," 2009.
- [10] M. Faisal, W. S. Utami and S. Parmica, "Implementasi Algoritma K-Nearest Neighbor (KNN)," *Journal Sensi*, pp. 11-23, 2023.

- [11] J. Nasir, R. Saputra, G. Efendi, A. Zahmi and Y. L. Setiawan, "K-Nearest Neighbor untuk Frasa Guna Mendukung Keputusan," *Jurnal Ilmu Komputer Agri-Informatika*, pp. 19-22, 2022.
- [12] R. Wajhillah, I. H. Ubaidallah and S. Bahri, "Analisis Kelayakan Kredit Berbasis Algoritma K-Nearest Neighbor (Studi Kasus:)," *Jurnal Nasional Informatika dan*, pp. 121-125, 2019.
- [13] I. Jaya, A. Hizriadi and E. S. Purba, "Klasifikasi Surat Laporan Kehilangan Kepolisian Menggunakan Algoritma K-Nearest Neighbor," *TECHSI*, pp. 120-128, 2018.
- [14] M. K. M. & M. N. S. T. Arhami, *Data Mining-Algoritma dan Implementasi*, Penerbit Andi, 2020.
- [15] M. F. Amin, "Confusion Matrix in Binary Classification Problems: A Step-by-Step Tutorial," *Journal of Engineering Research*, 2022.