



Analisis Kelayakan Pembiayaan Anggota Koperasi Dengan Metode Komparasi Algoritma K-Nearest Neighbors Dan Naive Bayes (Studi Kasus Pada KSP. XYZ)

*Rafi Lutfansyah¹

^{1,2,3)} Teknik Informatika, Teknik, Universitas Pamulang, Tangerang Selatan, Banten

Email: ¹rafilutfansyah@gmail.com

ABSTRACT

Savings and Loan Cooperatives often face the challenge of defaulted loans, which pose risks to financial stability and member trust. This study aims to compare the performance of the K-Nearest Neighbors (KNN) and Naive Bayes algorithms in classifying loan eligibility using the CRISP-DM (Cross-Industry Standard Process for Data Mining) approach. A case study was conducted on member financing data to identify a more accurate classification model to minimize loan defaults. The CRISP-DM methodology encompasses business understanding, data analysis, data preparation, modeling, evaluation, and deployment. The results show that KNN achieved the highest accuracy rate of 92.86%, while Naive Bayes only reached 85.71%. Additionally, KNN outperformed Naive Bayes in terms of precision and recall. Thus, KNN was selected as the optimal model to assist cooperatives in predicting loan eligibility. The implementation of this model is expected to improve financing efficiency, reduce default risks, and strengthen data-driven decision-making in cooperatives.

Keywords: K-Nearest Neighbors, Naive Bayes, CRISP-DM, classification, loan eligibility

ABSTRAK

Koperasi Simpan Pinjam sering menghadapi tantangan gagal bayar yang berisiko mengganggu stabilitas keuangan dan kepercayaan anggota. Penelitian ini bertujuan untuk membandingkan kinerja algoritma K-Nearest Neighbors (KNN) dan Naive Bayes dalam klasifikasi kelayakan pembiayaan menggunakan pendekatan CRISP-DM (Cross-Industry Standard Process for Data Mining). Studi kasus dilakukan pada data pembiayaan anggota untuk menentukan model klasifikasi yang lebih akurat guna mengurangi risiko gagal bayar. Tahapan CRISP-DM mencakup pemahaman bisnis, analisis data, persiapan data, pemodelan, evaluasi, hingga implementasi. Hasil analisis menunjukkan bahwa KNN memberikan tingkat akurasi tertinggi sebesar 92,86%, sementara Naive Bayes hanya mencapai 85,71%. Selain itu, KNN juga memiliki keunggulan dalam precision dan recall dibandingkan Naive Bayes. Dengan demikian, KNN dipilih sebagai model terbaik untuk membantu koperasi dalam memprediksi kelayakan pembiayaan. Implementasi model ini diharapkan dapat meningkatkan efisiensi proses pembiayaan, mengurangi risiko gagal bayar, dan memperkuat pengambilan keputusan berbasis data di koperasi.

Kata kunci: *K-Nearest Neighbors, Naive Bayes, CRISP-DM, klasifikasi, kelayakan pembiayaan*

1. PENDAHULUAN

Koperasi Simpan Pinjam merupakan salah satu lembaga ekonomi yang berperan penting dalam meningkatkan kesejahteraan anggotanya. Dalam era digital saat ini,

koperasi dihadapkan pada tantangan untuk mengoptimalkan layanan dan operasionalnya melalui pemanfaatan teknologi informasi.

Salah satu permasalahan yang sering muncul adalah dalam hal analisis data anggota koperasi untuk berbagai keperluan, seperti prediksi keanggotaan aktif, analisis pinjaman, dan penentuan tingkat risiko kredit. Metode analisis yang efektif dan efisien sangat dibutuhkan untuk membantu koperasi dalam mengambil keputusan strategis. Fenomena ini menarik untuk diteliti karena pentingnya pemilihan algoritma yang tepat dalam proses pengolahan data dapat berdampak signifikan terhadap kualitas keputusan yang diambil oleh koperasi.

Sejumlah penelitian sebelumnya telah dilakukan dalam bidang analisis data menggunakan berbagai algoritma pembelajaran mesin (*machine learning*). Contohnya, penelitian oleh [1] membandingkan kinerja algoritma *K-Nearest Neighbors* (KNN) dan *Naive Bayes* dalam klasifikasi data pelanggan pada industri perbankan. Hasil penelitian menunjukkan bahwa KNN memiliki tingkat akurasi yang lebih tinggi, namun *Naive Bayes* lebih efisien dalam hal waktu komputasi.

Penelitian lain oleh [2] mengaplikasikan metode CRISP-DM dalam analisis data pada sektor retail, menunjukkan bahwa pendekatan ini membantu dalam meningkatkan ketepatan hasil analisis. Berdasarkan penelitian-penelitian terdahulu tersebut, dapat disimpulkan bahwa metode CRISP-DM dan algoritma KNN serta *Naive Bayes* memiliki potensi untuk diterapkan dalam analisis data.

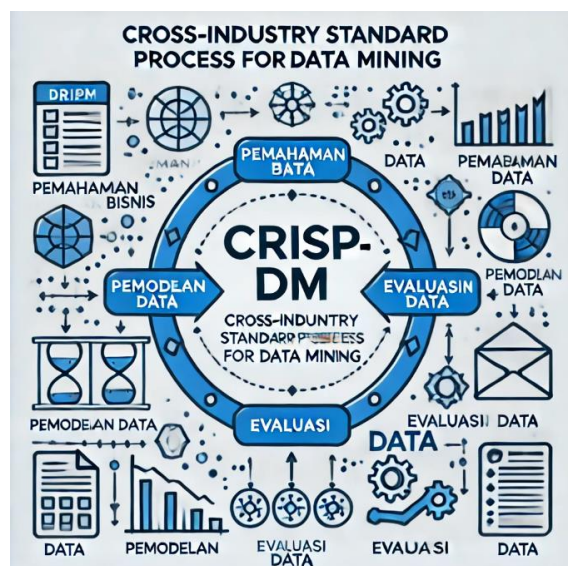
Penelitian ini akan menguraikan variabel, indikator, paradigma penelitian, jenis penelitian, populasi dan sampel, serta metode yang digunakan dengan membandingkannya terhadap penelitian terdahulu. Dalam penelitian ini, variabel yang akan dianalisis adalah data keanggotaan koperasi. Populasi penelitian ini adalah beberapa anggota koperasi, sedangkan sampelnya akan dipilih secara acak. Dengan menggunakan metode CRISP-DM, penelitian ini akan membandingkan kinerja algoritma KNN dan *Naive Bayes* dalam menganalisis data koperasi. Perbedaan utama dari penelitian ini dibandingkan penelitian terdahulu adalah fokus pada sektor koperasi, yang belum banyak diteliti sebelumnya.

Lokasi penelitian ini adalah Koperasi Simpan Pinjam. Koperasi ini dipilih karena memiliki basis data anggota yang cukup besar dan telah melakukan digitalisasi dalam pengelolaan data anggotanya. Selain itu, koperasi ini juga menghadapi berbagai permasalahan dalam pengambilan keputusan yang berbasis data, sehingga sangat relevan untuk dijadikan objek penelitian.

Berdasarkan latar belakang tersebut, peneliti akan melakukan penelitian dengan judul "Analisis Komparasi *K-Nearest Neighbors* dan *Naive Bayes* dengan Metode CRISP-DM (Studi Kasus: Klasifikasi Kelayakan Pembiayaan)".

2. METODE

Metodologi yang diterapkan dalam penelitian ini mengikuti pendekatan *Cross-Industry Standard Process for Data Mining* (CRISP-DM) [3]. Pendekatan ini meliputi tahapan-tahapan berikut: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*, sebagaimana dijelaskan dalam Gambar 1:



Gambar 1. *Cross-Industry Standard Process for Data Mining*

2.1. *Business Understanding*

Tahap *Business Understanding* merupakan bagian krusial dalam proses penelitian yang bertujuan untuk mendapatkan pemahaman mendalam mengenai kondisi, konteks, dan tujuan dari analisis data. Pada tahap ini, peneliti berusaha memahami permasalahan bisnis yang ada, termasuk masalah pembiayaan tidak layak yang sering menghambat gagal bayar yang diperlukan. Identifikasi dan pemahaman yang baik terhadap masalah ini

sangat penting untuk memastikan kelancaran aliran dana dan pengelolaan pembiayaan yang lebih efektif.

Penelitian ini fokus pada klasifikasi data historis pembiayaan menjadi dua kategori utama: layak dan tidak layak. Hasil klasifikasi ini akan menjadi dasar penting bagi pengambilan keputusan dalam proses persetujuan pembiayaan di masa mendatang, membantu mengurangi risiko pembiayaan tidak layak, dan memastikan aliran dana tetap efisien. Dengan klasifikasi yang tepat, pihak terkait dapat menentukan anggota yang layak menerima pembiayaan, sehingga tantangan terkait pembiayaan tidak layak dapat diatasi dengan lebih baik.

2.2. *Data Understanding*

Tahap *Data Understanding* merupakan langkah awal yang sangat penting dalam proses penelitian ini. Pada tahap ini, data dikumpulkan dengan seksama dari berbagai sumber yang relevan dan dianalisis secara mendalam untuk mendapatkan gambaran yang jelas mengenai situasi pembiayaan. Setelah data berhasil dikumpulkan, *dataset* tersebut akan melalui proses pengolahan lebih lanjut pada tahap-tahap berikutnya untuk memastikan akurasi dan relevansi informasi yang diperoleh.

Pada tahap berikutnya, setiap entri data diberi label "layak" atau "tidak layak" untuk mengidentifikasi status kelayakan dari pembiayaan yang dianalisis. Pemberian label ini bertujuan untuk memahami performa pembayaran serta mendeteksi kendala yang mungkin muncul dalam proses pembiayaan. Dengan demikian, label ini mempermudah analisis terhadap bagaimana pembayaran dilakukan dan apakah ada masalah yang perlu diperhatikan lebih lanjut.

2.3. *Data Preparation*

Data Preparation merupakan tahap untuk menyiapkan data yang akan digunakan dalam pemodelan. Data pembiayaan yang telah dikumpulkan kemudian diseleksi menjadi 10 kriteria. Penyeleksian kriteria ini dilakukan untuk menentukan variabel yang relevan dalam klasifikasi data. Berikut adalah Tabel 1 yang berisi data yang akan digunakan untuk pemodelan:

Tabel 1. *Data Preparation*

Atribut	Deskripsi	Tipe Data
Jenis Kelamin	Laki-laki/Perempuan	Kategorikal

Status Perkawinan	Kawin/Belum Kawin	Kategorikal
Tanggung	Jumlah orang	Kategorikal
Lulus SMA	Lulus/Tidak Lulus	Kategorikal
Pekerjaan	Wiraswasta/Wirausaha	Kategorikal
Pendapatan Anggota	Pendapatan Utama	Numerik
Pendapatan Penjamin	Pendapatan tambahan	Numerik
Jumlah Pembiayaan	Jumlah pinjaman	Numerik
Tenor Pembiayaan	Jangka Waktu	Numerik
Status Kelayakan	Layak/Tidak Layak	Kategorikal

2.4. Modelling

Tahap pemodelan memiliki peranan yang sangat penting dalam penerapan teknik *data mining*. Pada tahap ini, *dataset* pembiayaan yang terdiri dari 614 *record*, yang telah diperoleh pada tahap sebelumnya, digunakan sebagai input untuk algoritma klasifikasi. Pada penelitian ini, algoritma *K-Nearest Neighbors* (KNN) dan *Naive Bayes* diterapkan untuk melakukan klasifikasi data. Implementasi kedua algoritma tersebut dilakukan menggunakan *library Scikit-learn*, yang merupakan salah satu *library machine learning* populer untuk *Python*. Tahap ini sangat krusial karena hasil dari algoritma klasifikasi akan menentukan seberapa baik *model* dapat memprediksi data yang tersedia [4-6].

2.5. Deployment

Pada tahap *deployment*, hasil dari seluruh tahapan sebelumnya diimplementasikan dalam konteks yang nyata dan operasional. Pada fase ini, model yang telah terbukti memiliki akurasi tertinggi dipilih dan diterapkan ke dalam sistem yang ada. Tujuan utama dari penerapan *model* ini adalah untuk memprediksi data pembiayaan anggota dengan mengategorikannya ke dalam dua kelompok utama, yaitu kategori layak dan kategori tidak layak. Model ini diharapkan dapat memberikan prediksi yang benar dan membantu dalam pengambilan keputusan yang lebih tepat terkait kelayakan pembiayaan anggota [7-8]. status kelayakan tersebut. Pemisahan ini penting untuk mempersiapkan data sebelum dilakukan proses pelatihan model prediktif.

2.5.1. Normalisasi Data

Pada tahapan ini, dilakukan normalisasi data dengan memanfaatkan metode skala *Min-Max*. Tujuan utama dari normalisasi ini adalah untuk mengatur ulang rentang nilai-nilai dalam *dataset* sehingga variabel-variabel yang memiliki skala atau rentang yang berbeda bisa dibandingkan secara adil. Dengan melakukan normalisasi menggunakan skala *Min-Max*, setiap nilai dalam *dataset* diubah sedemikian rupa sehingga berkisar

antara nilai minimum dan maksimum yang telah ditentukan sebelumnya. Hal ini penting untuk memastikan bahwa data yang diolah memiliki konsistensi dan dapat menghasilkan akurasi prediksi yang lebih tinggi saat proses berlangsung. [9]

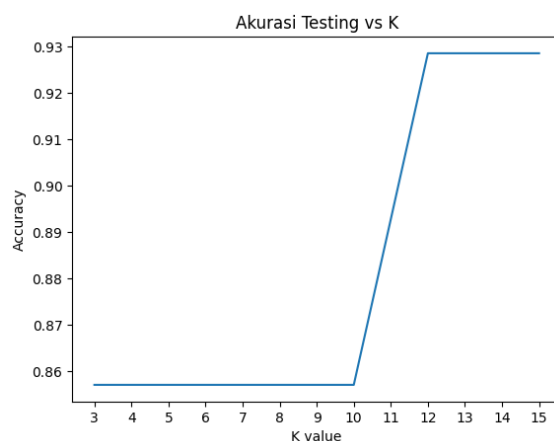
2.5.2. Pembagian Data

Pada langkah ini, dilakukan pembagian data menjadi dua set terpisah, yaitu *data training* dan *data testing*. *Data training* digunakan untuk melatih *model* atau algoritma prediktif, sementara *data testing* digunakan untuk menguji kinerja *model* yang telah dilatih. Pembagian dilakukan dengan rasio 80:20, di mana 80% dari data digunakan untuk *training* dan 20% sisanya untuk *testing* [10]. Hal ini penting untuk memastikan bahwa model yang dikembangkan dapat menggeneralisasi dengan baik terhadap data yang belum pernah dilihat sebelumnya, sehingga hasil evaluasi akhir dapat diandalkan untuk keperluan prediksi dan analisis lebih lanjut.

2.6. Modeling

2.6.1. K-Nearest Neighbors

Dalam evaluasi hasil klasifikasi, uji coba dilakukan dengan menggunakan nilai K dari 1 hingga 15. Berikut adalah hasil akurasi dari setiap percobaan dengan nilai K yang berbeda-beda. Hasil percobaan dari $K=1$ hingga $K=15$ dapat dilihat pada Gambar 4.12:



Gambar 2. Menentukan Nilai K

Nilai K dalam algoritma KNN menentukan jumlah tetangga yang diperiksa untuk klasifikasi data. Variasi nilai K dalam setiap pengujian mempengaruhi *accuracy*. Dalam penelitian ini, nilai *accuracy* tertinggi dari berbagai percobaan dengan nilai K yang berbeda akan digunakan. Setelah melakukan uji coba dengan nilai K dari 1 hingga 15,

ditemukan peningkatan akurasi secara umum, tetapi terjadi penurunan akurasi pada $K=3$. Akurasi terbaik dicapai pada $K=15$ dengan nilai 92,86%.

2.6.2. Naive Bayes

Peneliti Pada tahap ini, pelatihan dilakukan dengan menggunakan algoritma *Naive Bayes (Gaussian)*, yang menghasilkan nilai akurasi yang sangat memuaskan, yaitu sebesar 85,71%. Hasil ini menunjukkan bahwa *model* yang dihasilkan memiliki kemampuan prediksi yang tinggi dan dapat diandalkan untuk memisahkan data ke dalam kategori yang benar dengan tingkat keakuratan yang signifikan. Dengan *accuracy* hampir mencapai 86%, algoritma ini membuktikan efisiensinya dalam analisis dan pengolahan data yang bersifat kontinu dan mengikuti distribusi normal.

2.7. Evaluation

Berdasarkan Tabel 2 hasil evaluasi performa dua model klasifikasi, KNN dan *Naive Bayes*, dapat dilihat bahwa KNN memiliki nilai *accuracy*, *precision*, *recall*, dan *f1 score* yang lebih tinggi dibandingkan *Naive Bayes*.

Tabel 2. Data Preparation

Metric	K-Nearest Neighbors	Naive Bayes
Accuracy	92,86%	85,71%
Precision	92,86%	92,31%
Recall	100%	92,31%
F1 Score	96,30%	92,31%

Model KNN mencapai *accuracy* sebesar 92.86%, dengan *precision* dan *f1 score* masing-masing sebesar 92.86% dan 96.30%. Meskipun *Naive Bayes* juga menunjukkan hasil yang cukup baik dengan *accuracy* 85.71% dan nilai *precision* serta *recall* sebesar 92.31%, nilai *f1 score* sedikit lebih rendah dibandingkan KNN. Hal ini menandakan bahwa dalam kasus ini, KNN mungkin lebih cocok untuk digunakan dalam memprediksi kelas data, mengingat kemampuannya dalam menghasilkan hasil yang lebih seimbang antara *precision* dan *recall*.

2.8. Deployment

Setelah dianalisis dan ditemukan bahwa algoritma KNN menunjukkan *accuracy* yang lebih tinggi dibandingkan dengan algoritma *Naive Bayes*, langkah berikutnya adalah melakukan *deployment model*. Pada tahap *deployment* ini, model yang telah dilatih akan diekspor ke dalam satu *file* dengan format *pickle*. Proses ekspor ini bertujuan untuk

menyimpan *model* sehingga dapat digunakan kembali di masa mendatang tanpa perlu melatih ulang, memudahkan integrasi dengan aplikasi lain, serta memungkinkan pengujian dan evaluasi lanjutan pada data yang baru. *File pickle* yang dihasilkan berfungsi sebagai representasi dari *model* KNN yang sudah optimal, siap untuk digunakan dalam berbagai aplikasi prediktif.

3. KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma *K-Nearest Neighbors* (KNN) dengan tingkat akurasi sebesar 92,86% memberikan hasil yang lebih baik dalam klasifikasi kelayakan pembiayaan anggota koperasi dibandingkan dengan algoritma *Naive Bayes* yang memiliki akurasi sebesar 85,71%. Implementasi sistem pendukung prediksi kelayakan pembiayaan ini dapat berkontribusi secara signifikan dalam mengurangi risiko gagal bayar bagi koperasi. Selain itu, penggunaan *model* klasifikasi yang telah dikembangkan dapat membantu pengambilan keputusan yang lebih tepat dan efisien terkait pembiayaan anggota koperasi.

4. DAFTAR PUSTAKA

- [1] M. H. Effendy, D. Anggraeni, Y. S. Dewi, and A. F. Hadi, "Classification of Bank Deposit Using Naïve Bayes Classifier (NBC) and K –Nearest Neighbor (K -NN) ," *Proc. Int. Conf. Math. Geom. Stat. Comput. (IC-MaGeStiC 2021)*, vol. 96, pp. 163–166, 2022, doi: 10.2991/acsr.k.220202.031.
- [2] J. W. Iskandar and Y. Nataliani, "Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1120–1126, 2021, doi: 10.29207/resti.v5i6.3588.
- [3] A. Azevedo and M. F. Santos, "KDD, SEMMA and CRISP-DM: A parallel overview," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 11, no. 1, p. e1398, 2021, doi: <https://doi.org/10.1002/widm.1398>.
- [4] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011, doi: 10.5555/1953048.2078195.
- [5] A. Kumar and P. K. Mallick, "Heart disease prediction using hybrid ensemble classifier," *Int. J. Eng. \& Technol.*, vol. 7, no. 3.12, pp. 106–109, 2018, doi: 10.14419/ijet.v7i3.12.15993.

- [6] P. Narayan, G. Jothi, and B. Mani, “A comparative analysis of machine learning algorithms for credit scoring,” *J. Comput. Theor. Nanosci.*, vol. 17, no. 9, pp. 4307–4312, 2020, doi: 10.1166/jctn.2020.9110.
- [7] N. Sharma, R. Mishra, and S. Kaur, “Deployment of a machine learning model for credit risk assessment,” *Procedia Comput. Sci.*, vol. 171, pp. 1048–1055, 2020, doi: <https://doi.org/10.1016/j.procs.2020.04.112>.
- [8] M. A. Borg, J. A. Briffa, and J. Buhagiar, “Practical deployment of machine learning models in real-world applications,” *AI Ethics*, vol. 2, pp. 123–133, 2021, doi: <https://doi.org/10.1007/s43681-021-00064-2>.
- [9] H. Henderi, T. Wahyuningsih, and E. Rahwanto, “Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer,” *Int. J. Informatics Inf. Syst.*, vol. 4, no. 1, pp. 13–20, 2021, doi: <https://doi.org/10.47738/ijis.v4i1.73>.
- [10] A. A. Shujaaddeen, F. Mutaher Ba -Alwi, A. T. Zahary, and A. Sultan Alhegami, “A Model for Measuring the Effect of Splitting Data Method on the Efficiency of Machine Learning Models: A Comparative Study,” in *2024 4th International Conference on Emerging Smart Technologies and Applications (eSmarTA)*, IEEE, Aug. 2024, pp. 1–13. doi: 10.1109/eSmarTA62850.2024.10639022.