

EKSPLORASI METODE *NAIVE BAYES CLASSIFIER* PADA *MACHINE LEARNING* UNTUK SISTEM *ALTERNATIVE CREDIT SCORING*

SKRIPSI

Diajukan Sebagai Salah Satu Persyaratan Untuk Memperoleh Gelar
Strata Satu (S1) Jurusan Teknik Elektro



**PRODI TEKNIK ELEKTRO
FAKULTAS TEKNIK
UNIVERSITAS ISLAM MALANG
2025**

ABSTRAK

AININA MIFTAHUL JANNAH, 22001053015, Jurusan Teknik Elektro, Fakultas Teknik, Universitas Islam Malang, “EKSPLORASI METODE *NAIVE BAYES* CLASSIFIER PADA *MACHINE LEARNING* UNTUK SISTEM ALTERNATIVE *CREDIT SCORING*”, Dosen Pembimbing 1 : Anang Habibi, S.ST., M.T., Dosen Pembimbing II : Bambang Minto B., S.T., M.T.

Permasalahan dalam sistem kredit konvensional sering kali terletak pada keterbatasan akses masyarakat terhadap penilaian kelayakan kredit, khususnya bagi individu yang tidak memiliki riwayat kredit formal. Penelitian ini bertujuan untuk mengembangkan sistem alternatif *Credit Scoring* menggunakan metode *Naïve Bayes Classifier (NBC)* yang mampu mengevaluasi kelayakan kredit berdasarkan variabel sosial ekonomi dari data peserta Jaminan Kesehatan Nasional (JKN) Kota Malang. Variabel yang digunakan meliputi usia, total pendapatan, lama bekerja (*days employed*), dan total tagihan bulanan. Tahapan penelitian dimulai dengan proses preprocessing data, normalisasi, pembobotan skor berdasarkan skala 1–5, dan modeling yang dilakukan menggunakan *GaussianNB* untuk mempelajari hubungan antara fitur-fitur numerik dan label target kelayakan kredit kemudian menghitung nilai rata-rata dan varians dari setiap fitur pada masing-masing kelas serta menghitung probabilitas awal tiap kelas yang secara internal membentuk probabilistik berbasis distribusi. Evaluasi model dilakukan menggunakan metrik akurasi, presisi, *recall*, dan *f1-score*. Hasil pengujian menunjukkan bahwa model *Naïve Bayes* dapat mengklasifikasikan kelayakan kredit dengan akurasi sebesar 84%, presisi 85%, *recall* 84%, dan *f1-score* 84%. Penelitian ini menunjukkan bahwa pendekatan *Machine Learning* dengan metode *Naïve Bayes* memiliki potensi untuk dijadikan sebagai sistem alternatif dalam penilaian kelayakan kredit, terutama untuk masyarakat yang tidak terjangkau oleh sistem kredit konvensional.

Kata Kunci : *Machine Learning*, *Naïve Bayes*, *Credit Scoring*, Lembaga Keuangan, JKN.

ABSTRACT

AININA MIFTAHUL JANNAH, 22001053015, Departement of Electrical Engineering, Faculty of Engineering, Islamic University of Malang, “*EXPLORATION OF NAIVE BAYES CLASSIFIER METHOD IN MACHINE LEARNING FOR ALTERNATIVE CREDIT SCORING SYSTEM*”, Supervisor 1 : Anang Habibi, S.ST., M.T., Supervisor II : Bambang Minto B., S.T., M.T.

Problems in conventional credit systems often lie in the limited access people have to creditworthiness assessments, particularly for individuals without a formal credit history. This study aims to develop an alternative Credit Scoring system using the Naïve Bayes Classifier (NBC) method, which can evaluate creditworthiness based on the socioeconomic variables of participants in the National Health Insurance (JKN) program in Malang City. The variables used include age, total income, length of employment (days employed), and total monthly bills. The research stages began with data preprocessing, normalization, score weighting based on a scale of 1–5, and modeling using GaussianNB to learn the relationships between numerical features and the creditworthiness target label. It then calculated the mean and variance of each feature within each class and computed the prior probabilities of each class, internally forming a distribution-based probabilistic model. Model evaluation was carried out using accuracy, precision, recall, and F1-score metrics. The test results showed that the Naïve Bayes model could classify creditworthiness with an accuracy of 84%, precision of 85%, recall of 84%, and an F1-score of 84%. This study demonstrates that a Machine Learning approach using the Naïve Bayes method has the potential to serve as an alternative system for creditworthiness assessment, especially for communities underserved by conventional credit systems.

Keywords— *Machine Learning, Naïve Bayes, Credit Scoring, Financial Instituion, JKN.*

BAB I

PENDAHULUAN

1.1 Latar Belakang Penelitian

Pemberian kredit merupakan praktik umum dan memiliki peran penting dalam operasional lembaga keuangan. Salah satu contoh utama lembaga keuangan yang sering menawarkan kredit adalah bank. Kredit menjadi salah satu sumber pendapatan bagi bank, perusahaan swasta, maupun lembaga lain yang menyediakan layanan pembiayaan kepada konsumen. Oleh karena itu, lembaga keuangan harus mampu mengidentifikasi calon nasabah yang memiliki kemampuan untuk melunasi kredit yang diberikan. Salah satu metode yang digunakan untuk menilai kelayakan nasabah dalam menerima pinjaman adalah *Credit Scoring* [1].

Saat ini, sistem *Credit Scoring* tradisional masih menjadi standar utama dalam menilai kelayakan kredit, baik bagi individu maupun perusahaan. Evaluasi kredit secara konvensional umumnya mempertimbangkan berbagai faktor seperti riwayat kredit, tingkat penghasilan, kepemilikan aset, dan rekam jejak pekerjaan. Namun, metode ini sering kali membatasi akses bagi calon peminjam yang belum memiliki riwayat kredit atau yang baru pertama kali mengajukan pinjaman.

Untuk mengatasi kendala tersebut, penelitian ini mengusulkan pengembangan Sistem *Alternative Credit Scoring* sebagai solusi inovatif. Sistem ini mengintegrasikan sumber data non-tradisional, seperti riwayat tagihan serta informasi pribadi dari data Jaminan Kesehatan Nasional, guna menyusun profil kredit seseorang secara lebih komprehensif. Keunggulan utama dari pendekatan ini adalah kemampuannya dalam memperluas akses kredit bagi individu yang sebelumnya tidak memenuhi kriteria sistem tradisional. Selain itu, metode alternatif ini juga mampu memberikan penilaian kredit yang lebih objektif dan adil dengan mengurangi potensi bias manusia.

Dalam penelitian ini, analisis data dilakukan menggunakan metode *Machine Learning Naïve Bayes Classifier*, yang efektif dalam mengidentifikasi pola dan hubungan kompleks dalam data. *Naïve Bayes Classifier* dianggap mampu menyelesaikan permasalahan probabilitas dengan mempertimbangkan kriteria yang relevan bagi penerima kredit. Dengan demikian, penelitian ini bertujuan untuk

mengembangkan sistem yang dapat memberikan penilaian akurat terkait kelayakan konsumen dalam mengajukan kredit berdasarkan data Jaminan Kesehatan Nasional.

1.2 Rumusan Masalah

Dari permasalahan latar belakang di atas, perumusan masalah untuk penelitian yang akan dilakukan adalah:

1. Bagaimana merancang dan membangun sistem *Alternative Credit Scoring* berbasis *Machine Learning* untuk mengatasi keterbatasan sistem konvensional dalam menilai kelayakan kredit?
2. Bagaimana efektivitas eksplorasi metode *Naïve Bayes Classifier* dalam sistem *Alternative Credit Scoring* dengan memanfaatkan data kepesertaan Jaminan Kesehatan Nasional (JKN) masyarakat Kota Malang sebagai sumber data alternatif?

1.3 Batasan Masalah

Untuk meminimalisir perluasan pembahasan penelitian, adapun batasan-batasan masalah yang diberikan sebagai berikut:

1. Sistem hanya menggunakan Data Sample Jaminan Kesehatan Nasional masyarakat kota Malang
2. Hasil klasifikasi yang dihasilkan terbatas pada angka atau huruf yang menunjukkan nilai kelayakan
3. Model klasifikasi kelayakan kredit hanya dibangun menggunakan empat variabel numerik
4. Sistem yang dibuat menggunakan *library* yang ada di *Machine Learning* python
5. Sistem tidak mencakup proses integrasi dengan sistem lembaga keuangan dan hanya dapat digunakan pada masyarakat yang mempunyai Jaminan Kesehatan Nasional
6. Penelitian hanya berfokus pada algoritma *Naïve Bayes Classifier*, tanpa melakukan perbandingan performa terhadap algoritma klasifikasi lainnya seperti *Decision Tree*, *K-Nearest Neighbors*, *Random Forest*, atau *Support Vector Machine*

7. Penelitian tidak membangun aplikasi atau antarmuka pengguna (*user interface*)

1.4 Tujuan Penelitian

1. Dapat menghasilkan sebuah *Sistem Alternative Credit Scoring* untuk membantu lembaga keuangan mengetahui kelayakan konsumen dalam mengajukan kredit
2. Mengetahui seberapa efektif penggunaan data Jaminan Kesehatan Nasional masyarakat kota Malang nilai kredit hasil pengujian *Sistem Alternative Credit Scoring*
3. Menganalisis performa dan keakuratan model *Naïve Bayes Classifier* dibandingkan dengan metode tradisional dalam proses *Credit Scoring*

1.5 Manfaat Penelitian

Manfaat Penelitian yang di harapkan dalam penelitian ini adalah :

1. Manfaat dalam bidang keilmuan.
2. Membantu lembaga keuangan untuk mendapatkan alternatif pengecekan *scoring* terhadap nasabah yang akan melakukan peminjaman dengan mendapatkan analisis data berdasarkan data Jaminan Kesehatan Nasional masyarakat kota Malang
3. Menambah wawasan penyelesaian masalah dengan menggunakan *Machine Learning* yang diimplementasikan kedalam sistem
4. Manfaat praktis

BAB V

KESIMPULAN

5.1 Kesimpulan

Kesimpulan dari penelitian skripsi yang berjudul “Implementasi Metode *Naïve Bayes Classifier* Pada *Machine Learning* Untuk Sistem Alternative *Credit Scoring*” adalah sebagai berikut :

1. Sistem Alternative *Credit Scoring* dibangun dengan pendekatan *Machine Learning* menggunakan algoritma *Naïve Bayes Classifier (GaussianNB)* yang diimplementasikan melalui platform Google Colab dengan bahasa pemrograman *Python*. Data yang digunakan berasal dari Jaminan Kesehatan Nasional (JKN) masyarakat Kota Malang periode 2020–2023. Proses pengembangan sistem mencakup beberapa tahapan utama. **Pengumpulan data** melalui metode web scraping dari Portal Data JKN. **Preprocessing data**, termasuk data cleaning, normalisasi, dan penanganan missing value. **Penentuan variabel**, dengan empat variabel independen utama yaitu: usia (USIA), penghasilan (INCOME_TOTAL), jumlah tagihan bulanan (TOTALBILL_MONTH), dan durasi kerja (DAYS_EMPLOYED). **Penentuan kriteria kelayakan kredit**, berdasarkan kebijakan OJK dan literatur terkait. **Pelatihan model Gaussian Naïve Bayes** untuk mengklasifikasikan kelayakan kredit menjadi 5 skor, dari skor 1 (sangat layak) hingga skor 5 (tidak layak).
2. Hasil implementasi sistem menunjukkan bahwa model *Naïve Bayes Classifier (GaussianNB)* mampu melakukan klasifikasi kelayakan kredit dengan akurasi tinggi. Berdasarkan pengujian terhadap dataset yang berjumlah 30 sampel valid (yang digunakan untuk uji manual dan Sklearn). Model Sklearn *GaussianNB* menghasilkan akurasi 84%, yaitu seluruh prediksi skor kredit. Hasil ini membuktikan bahwa algoritma *Naïve Bayes* efektif untuk dikembangkan dalam sistem alternative *Credit Scoring* berbasis data JKN.

5.2 Saran

Sistem alternative kredit scoring yang dibuat masih jauh dari kata sempurna dan masih banyak kekurangan. Untuk itu diperlukan penyempurnaan sistem agar dapat menjadi lebih baik. Adapun saran agar sistem ini bisa berjalan dengan lebih optimal sebagai berikut :

1. Penggunaan Dataset yang Lebih Luas dan Beragam

Penelitian ini menggunakan sampel data dari Jaminan Kesehatan Nasional (JKN) masyarakat Kota Malang. Untuk pengembangan lebih lanjut, disarankan agar sistem diujikan pada dataset yang lebih besar dan mencakup wilayah geografis yang lebih luas agar hasil klasifikasi lebih general dan representatif terhadap kondisi masyarakat Indonesia secara keseluruhan.

2. Pengembangan Antarmuka Pengguna (User Interface)

Sistem yang dibangun pada penelitian ini masih berfokus pada pengolahan dan analisis data berbasis kode di Google Colab. Untuk penerapan nyata di lembaga keuangan, disarankan agar sistem dikembangkan lebih lanjut dalam bentuk aplikasi dengan antarmuka pengguna (*UI*) yang ramah, agar dapat digunakan oleh staf operasional non-teknis maupun pengguna umum.



DAFTAR PUSTAKA

- [1] S. A. Nurrizqi Rahmadania and N. Nurismalatri, “Analisis Credit Scoring Dan Faktor-Faktor Yang Mempengaruhi Npf Pembiayaan Murabahah Pada Pt Bank Muamalat Indonesia,” *JABI (Jurnal Akunt. Berkelanjutan Indones.*, vol. 3, no. 2, pp. 204–221, 2020, doi: 10.32493/jabi.v3i2.y2020.p204-221.
- [2] D. A. Hasri, “Penerapan Fuzzy Set Theory Dalam Credit Scoring Usaha Mikro, Kecil Dan Menengah (Ukm),” *J. Ekon. dan Bisnis Indones.*, vol. 12, no. 1, pp. 1976–1993, 1997.
- [3] A. Ramaquita and A. Alamsyah, “Model Credit Scoring Berdasarkan Data Demografi Dan Jejaring Sosial Di Media Sosial,” *e-Proceeding Manag.*, vol. 7, no. 2, pp. 2214–2219, 2020.
- [4] D. W. Triscowati and L. D. Jayanti, “Penilaian Kredit Pada Data Tak Seimbang Menggunakan Random Forest Credit Scoring In Imbalance Data Using Random Forest,” *JIKOSTIK – J. Ilm. Komputasi dan Stat.*, vol. 1, no. 1, pp. 25–31, 2021.
- [5] S. Informasi, U. Darwan Ali, J. Batu Berlian, and K. Tengah, “Penerapan Metode *Naive bayes* Dalam Mengklasifikasi Penerima BLT Pada Desa Pelangian,” *J. JUPITER*, vol. 14, no. 2, pp. 64–70, 2022.
- [6] R. M. Samhadi1 and S. , Bambang Minto Basuki2, “Implementasi Multiple Linear Regression Pada Machine Learning Untuk Prediksi,” pp. 1–7, 2022.
- [7] A. Khahar and Y. Yamasari, “Eksplorasi Teknik Pre-Processing Berbasis eXtreme Gradient Boosting (XGBoost) pada Klasifikasi Kredit Default,” vol. 06, pp. 453–464, 2024.
- [8] A. Afandi, H. B. Agtriadi, M. Nur, and I. Susanti, “Jurnal E-Komtek Advanced Credit Scoring with *Naive bayes* Algorithm : Improving Accuracy and Reliability in Financial Risk Assessment,” vol. 8, no. 2, pp. 399–409, 2024.
- [9] M. Prameswari, P. E. Kania, I. G. De Ayu, S. Namira, and P. Harnoko,

- “Penerapan Metode Stacking Ensemble Untuk Klasifikasi Status Pinjaman Nasabah Bank,” vol. 2024, no. Senada, pp. 802–811, 2024.
- [10] A. Hasna, “PENILAIAN KEMAMPUAN PEMBAYARAN KREDIT DENGAN MENGGUNAKAN MACHINE LEARNING LOGISTIC REGRESSION DAN RANDOM FOREST CLASSIFIER PADA HOME CREDIT Oleh,” *Univ. Lampung*, vol. 11, no. 1, pp. 1–14, 2019, [Online]. Available:
http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEMBETUNGAN_TERPUSAT_STRATEGI_MELESTARI
- [11] M. Anam, “PENGEMBANGAN MODEL CREDIT RISK SCORING UNTUK PEMBIAYAAN DI LEMBAGA KEUANGAN SYARIAH,” *Univ. PTIQ Jakarta*, vol. 15, no. 1, pp. 37–48, 2024.
- [12] A. Nuriksan, T. Hendro Pudjiantoro, and P. Nurul Sabrina, “Klasifikasi Kelayakan Kredit Motor Menggunakan Metode Naïve Bayes Dan Model Credit Scoring,” *Semin. Nas. Inform. dan Apl.*, pp. G1–G6, 2021.
- [13] DJSN, “Statistik JKN 2016-2021,” *Dewan Jaminan Sos. Nas.*, pp. 1–252, 2022.
- [14] M. R. A. A. S, B. M. Basuki, A. Habibi, I. Malang, and U. I. Malang, “IMPLEMENTASI ALGORITMA MACHINE LEARNING UNTUK PREDIKSI PENJUALAN CROISSANT PADA CAFE RETAWUDELI BERBASIS HISTORI PENJUALAN TIME SERIES,” pp. 1–7.
- [15] A. S. Lombu, S. Hidayat, and A. F. Hidayatullah, “Pemodelan Klasifikasi Gaji Menggunakan Support Vector Machine,” *J. Comput. Syst. Informatics*, vol. 3, no. 4, pp. 363–370, 2022, doi: 10.47065/josyc.v3i4.2137.
- [16] H. F. Putro, R. T. Vlandari, and W. L. Y. Saptomo, “Penerapan Metode *Naive bayes* Untuk Klasifikasi Pelanggan,” *J. Teknol. Inf. dan Komun.*, vol. 8, no. 2, 2020, doi: 10.30646/tikomsin.v8i2.500.
- [17] Fatimah, M. Abdul Mukid, and A. Rusgiyono, “Analisis Credit Scoring Menggunakan Metode Bagging K-Nearest Neighbor,” *J. Gaussian*, vol. 6,

- no. 1, pp. 161–170, 2017, [Online]. Available: <http://ejournal-s1.undip.ac.id/index.php/gaussian>
- [18] E. T. Kurniawati, “Pengembangan sistem credit scoring pembiayaan UMKM,” *Inovasi*, vol. 17, no. 3, pp. 609–616, 2021.
- [19] Ahmed, “The Motivation for Train-Test Split,” medium. Accessed: Jun. 13, 2025. [Online]. Available: <https://medium.com/%40nahmed3536/the-motivation-for-train-test-split-2b1837f596c3>
- [20] N. Nurhayati aris, “Penerapan Algoritma *Naive bayes* Dalam Menentukan Kelayakan Nasabah Pada Koperasi Simpan Pinjam,” *J. Comput. Inf. Syst. (J-CIS)*, vol. 1, no. 1, pp. 1–9, 2021, doi: 10.31605/jcis.v2i2.811.
- [21] A. Perguruan, T. Aperti, N. C. Siregar, R. R. A. Siregar, and M. Y. D. Sudirman, “Jurnal Teknologi Implementasi Metode *Naive bayes* Classifier (NBC) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (PJJ) Jurnal Teknologi,” vol. 3, no. 1, pp. 102–110, 2020.
- [22] B. Tandika, “Bahasa Pemrograman *Python*: Pelajari Arti, Fungsi, dan Keunggulannya,” 2024, [Online]. Available: <https://glints.com/id/lowongan/apa-itu-bahasa-pemrograman-python/>
- [23] I. Binanto, N. F. Sianipar, N. M. D. Aprilianti, J. Gein, and P. D. Paska, “Analisis Perbandingan Algoritma Knn, Gaussian *Naive bayes*, Random Forest Untuk Data Tidak Seimbang Dan Data Yang Diseimbangkan Dengan Metode Tomek Link Undersampling Pada Dataset Lcms Tanaman Keladi Tikus,” *Pros. Sains Nas. dan Teknol.*, vol. 13, no. 1, p. 156, 2023, doi: 10.36499/psnst.v13i1.9002.
- [24] A. P. Candra, “Analisis Data Menggunakan *Python* : Memperkenalkan Pandas dan NumPy,” vol. 3, no. 1, pp. 11–16, 2025.
- [25] A. Talia, A. A. Tumanggor, and F. Hasibuan, “Implementasi Metode Numerik dan Simbolik dengan *Python* untuk Penentuan Nilai Limit Fungsi,” no. 4, 2024.
- [26] M. N. Fahmi, “Implementasi Machine Learning menggunakan *Python* Library : Scikit-Learn (Supervised dan Unsupervised Learning),” *Sains*

Data J. Stud. Mat. dan Teknol., vol. 1, no. 2, pp. 87–96, 2023, doi:
10.52620/sainsdata.v1i2.31.

- [27] R. Kaestria *et al.*, “Penerapan Matplotlib dalam Visualisasi Data untuk Analisis Hubungan Penggunaan Gadget dan Hasil Belajar,” *J. Digit. Bus. Inf. Technol.*, pp. 29–39, 2024, doi: 10.23971/jobit.v1i1.204.
- [28] I. D. S. Tarigan, Roni Habibi, and Rd. Nuraini Siti Fatonah, “Evaluasi Algoritma Klasifikasi Machine Learning Kategori Nilai Akhir Tunjangan Kinerja Pegawai,” *J. Sist. Cerdas*, vol. 6, no. 3, pp. 251–261, 2023, doi: 10.37396/jsc.v6i3.246.
- [29] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.
- [30] AGUSTINA, I., ALGORITMA *NAIVE BAYES* CLASSIFIER UNTUK KLASIFIKASI PEMBERIAN PINJAMAN PADA PT PNM MEKAAR KABUPATEN CIANJUR.

