Bina Informatika dan Komputer (BINER) DOI: ...
Vol. x, No. x, November 2025 ISSN: 3026-796X

PREDIKSI KEBERHASILAN UMKM DI KAWASAN SEKITAR JEMBATAN SUNGAI SAMBAS BESAR MENGGUNAKAN ALGORITMA GAUSSIAN NAÏVE BAYES

Cyndi Kartika Kumala Dewi[1], Muhammad Usman[2]

[1]Program Studi D3 Manajemen Informatika, Politeknik Negeri Sambas
[2]Program Studi D3 Manajemen Informatika, Politeknik Negeri Sambas
Email: [1]koemala.cys@gmail.com, [2]usman.mtaib2@gmail.com

Abstrak

Pembangunan Jembatan Sungai Sambas Besar (JSSB) secara tiba-tiba mengubah lanskap ekonomi di Kabupaten Sambas, memicu lonjakan pembukaan UMKM baru di kawasan tersebut. Kondisi ini menimbulkan urgensi tinggi karena banyak pelaku usaha mengambil risiko investasi modal tanpa memahami karakteristik usaha yang secara matematis relevan dengan keberhasilan di lingkungan dengan aksesibilitas baru. Jika tidak diprediksi sejak dini, hal ini berpotensi menyebabkan tingkat kegagalan usaha yang tinggi pasca euforia pembangunan jembatan. Penelitian ini berkontribusi dalam memitigasi risiko tersebut dengan membangun sistem prediksi keberhasilan UMKM menggunakan algoritma *Gaussian Naïve Bayes*. Sistem ini secara spesifik mengintegrasikan delapan variabel meliputi: modal awal, omzet usaha, lama usaha, jenis usaha, jarak lokasi usaha dari JSSB, jumlah tenaga kerja, jumlah pelanggan per bulan, dan metode pemasaran. Dari 41 data UMKM valid yang diuji menggunakan *Leave-One-Out Cross-Validation* (LOOCV), model menghasilkan akurasi 82,92%, *Precision* 72,73%, *Recall* 94,12%, dan *F1-Score* 82,05%. Algoritma ini berhasil diimplementasikan ke dalam sistem website prediksi, dan diharapkan dapat menjadi alat bantu evaluasi objektif bagi pemerintah daerah dan panduan bagi pelaku usaha dalam merancang strategi bertahan di kawasan JSSB.

Kata kunci: UMKM, *Gaussian Naïve Bayes*, Prediksi Keberhasilan Usaha, Jembatan Sungai Sambas Besar, Sistem Prediksi.

PREDICTION OF MSMEs SUCCESS IN THE SAMBAS BESAR RIVER BRIDGE AREA USING THE GAUSSIAN NAÏVE BAYES ALGORITHM

Abstract

The construction of the Jembatan Sungai Sambas Besar *(JSSB) has suddenly altered the economic landscape in* Sambas *Regency, triggering a surge in the establishment of new MSMEs in the region. This condition creates high urgency, as many business owners are taking capital investment risks without understanding the business characteristics that are mathematically relevant to success in an environment with newly improved accessibility. If not predicted early, this situation has the potential to cause a high rate of business failure following the post-construction euphoria of the bridge. This research contributes to mitigating these risks by developing an MSME success prediction system using the Gaussian Naïve Bayes algorithm. The system specifically integrates eight variables: initial capital, business turnover, business duration, business type, business location distance from JSSB, number of workers, number of monthly customers, and marketing method. Evaluated on 41 valid MSME data using Leave-One-Out Cross-Validation (LOOCV), the model achieved an accuracy of 82.92%, Precision 72,73%, Recall 94,12%, dan F1-Score 82,05%. The algorithm was successfully implemented into a website-based prediction system, which is expected to serve as an objective evaluation tool for local governments and a guide for business actors in designing survival strategies within the JSSB area.*

Keywords: *MSMEs, Gaussian Naïve Bayes, Business Success Prediction, Jembatan Sungai Sambas Besar, Prediction System.*

1. PENDAHULUAN

Pembangunan infrastruktur transportasi berperan vital dalam meningkatkan konektivitas antarwilayah, memperlancar distribusi barang dan jasa, serta membuka peluang aktivitas ekonomi baru. Peluang tersebut sangat identik dengan sektor Usaha Mikro, Kecil, dan Menengah (UMKM) yang menyumbang lebih dari 60% terhadap Produk Domestik Bruto nasional dan menyerap sekitar 97% tenaga kerja di Indonesia (PILIANG, 2024). Salah

satu infrastruktur yang mengubah dinamika ekonomi di Kabupaten Sambas, Kalimantan Barat, adalah Jembatan Sungai Sambas Besar (JSSB) yang mulai beroperasi pada tahun 2025.

Sebelum JSSB beroperasi, mobilitas masyarakat bergantung pada jalur penyeberangan yang terbatas. Kehadiran jembatan ini memicu lonjakan pembukaan UMKM baru di kawasannya. Namun, peluang pasar yang luas belum tentu memberikan tingkat keberhasilan yang sama bagi setiap UMKM. Perbedaan karakteristik usaha seperti modal awal, omzet, jenis usaha, lokasi, hingga strategi pemasaran menyebabkan kemampuan pelaku usaha dalam memanfaatkan peluang menjadi beragam. Masyarakat yang ingin membangun usaha di kawasan tersebut sering kali tidak memiliki dasar pertimbangan matematis mengenai faktor-faktor yang berkaitan dengan keberhasilan usaha.

Kajian sistem yang sedang berjalan saat ini menunjukkan beberapa celah. Penelitian terdahulu seperti yang dilakukan oleh MUHAMMAD REZA ROMADHONI, (2025) menggunakan *Machine Learning Random Forest* untuk prediksi keberhasilan UMKM, namun mengandalkan dataset sekunder umum (*Kaggle*) yang belum merepresentasikan dinamika wilayah pasca-pembangunan infrastruktur fisik. Di sisi lain, penelitian oleh RIANA RAHMADANTI, (2024) dan PUTRA BAGUS UTAMA, (2026) yang menggunakan algoritma *Naïve Bayes* untuk mengklasifikasi UMKM dilakukan pada konteks perkotaan umum tanpa memasukkan variabel spesifik seperti "jarak lokasi usaha dari jembatan".

Berdasarkan celah tersebut, penelitian ini bertujuan untuk menerapkan Algoritma *Gaussian Naïve Bayes* untuk memprediksi keberhasilan UMKM pasca pembangunan JSSB, mengevaluasi akurasi model menggunakan metode *Leave-One-Out Cross-Validation*, serta mengimplementasikannya ke dalam sebuah sistem *website* prediksi yang dapat dijadikan alat bantu pengambilan keputusan objektif.

2. LANDASAN TEORI

2.1. UMKM dan Keberhasilan Usaha

Usaha Mikro, Kecil, dan Menengah (UMKM) diatur dalam Undang-Undang Nomor 20 Tahun 2008 sebagai bentuk usaha ekonomi produktif yang bertujuan membangun perekonomian nasional berdasarkan demokrasi ekonomi (EKO PURWANTO, 2023). Keberhasilan UMKM didefinisikan sebagai kondisi di mana usaha mengalami peningkatan, mencapai keadaan lebih baik, serta mampu bertahan dan berkembang secara berkelanjutan (Fauzan Mawadah, 2023). Faktor-faktor seperti modal, manajerial, lokasi, dan pemasaran menjadi determinasi utama keberhasilan ini. Pada konteks pembangunan infrastruktur besar seperti JSSB, variabel lokasi tidak lagi diukur secara umum, melainkan dijabarkan spesifik menjadi "jarak lokasi usaha dari jembatan".

2.2. Data Mining dan Machine Learning

Data mining adalah proses sistematis mengekstraksi informasi berguna dari basis data besar untuk membangun model yang mampu mengenali pola-pola tersembunyi. Fungsi *predictive* dari *data mining* digunakan untuk memprediksi nilai variabel yang belum diketahui (RAHMADINI, 2023). Pendekatan ini diwujudkan melalui *Machine Learning*, khususnya paradigma *Supervised Learning*, di mana algoritma dilatih menggunakan dataset yang sudah memiliki label target (TRIULI NOVIANTI, 2025).

2.3. Algoritma Gaussian Naïve Bayes

Gaussian Naïve Bayes merupakan varian dari algoritma klasifikasi *Naïve Bayes* yang berlandaskan teorema *Bayes*. Keunikannya terletak pada kemampuan menangani atribut data numerik kontinu dengan mengasumsikan distribusi normal (*Gaussian*). Algoritma ini menghitung nilai rata-rata (μ) dan standar deviasi (σ) dari setiap fitur kontinu untuk disubstitusikan ke dalam fungsi kepadatan probabilitas (YOGA NAUFAL RAY PUTRO, 2024).

Asumsi "*Naïve*" diterapkan dengan menganggap setiap atribut bersifat independen satu sama lain terhadap kelas (*conditional independence*). Meskipun dalam realitas bisnis fitur-fitur seperti modal dan omzet mungkin berkorelasi, asumsi ini menyederhanakan komputasi probabilitas yang kompleks. Algoritma ini dipilih karena terbukti efisien, memiliki waktu komputasi cepat, dan sangat stabil ketika dihadapkan pada dataset berukuran kecil.

2.4. Metode CRISP-DM dan Evaluasi Model

Penelitian ini menggunakan kerangka kerja CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang terdiri dari enam tahapan: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* (MSY AULIA HASANAH, 2021). Pada tahap evaluasi, kinerja model diukur menggunakan *Confusion Matrix* yang menghasilkan empat metrik utama: *Accuracy*, *Precision*, *Recall*, dan *F1-Score* (JOSE JULIAN HIDAYAT, 2026). Mengingat keterbatasan data, pengujian model dilakukan menggunakan metode *Leave-One-Out Cross-Validation* (LOOCV), di mana satu data dijadikan data uji dan sisanya menjadi data latih, dilakukan bergantian sebanyak jumlah data (N) untuk meminimalkan bias evaluasi (TRIULI NOVIANTI, 2025).

3. METODOLOGI

Metode penelitian ini mengadaptasi tahapan CRISP-DM sebagai berikut:

1. *Business Understanding*, mengidentifikasi urgensi prediksi keberhasilan UMKM di kawasan JSSB akibat lonjakan pembukaan usaha pasca pembangunan jembatan.

2. *Data Understanding*, pengumpulan data primer melalui survei lapangan terhadap pelaku UMKM di kawasan JSSB. Data mencakup 8 variabel input (modal awal, omzet, lama usaha, jenis usaha, jarak dari JSSB, jumlah tenaga kerja, jumlah pelanggan per bulan, metode pemasaran) dan 1 variabel *output* (Keberhasilan: Berhasil/Tidak Berhasil).
3. *Data Preparation*, proses yang dilakukan mencakup *data cleaning* (menghapus data kosong dan duplikat), *data selection* (menghapus data tidak valid), serta *encoding* (transformasi data kategorikal menjadi numerik menggunakan pemetaan label agar sesuai dengan kebutuhan *input* algoritma). Dari proses ini diperoleh 41 data valid.
4. *Modeling*, penerapan algoritma *Gaussian Naïve Bayes* dengan skema validasi LOOCV. Pada setiap iterasi ke-i, model dilatih menggunakan 40 data dan diuji pada 1 data sisanya. Proses ini diulang sebanyak 41 kali.
5. *Evaluation*, menghitung akumulasi *Confusion Matrix* dari seluruh iterasi LOOCV untuk mendapatkan nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.
6. *Deployment*, model final dilatih ulang menggunakan keseluruhan 41 data, disimpan dalam format file, dan diintegrasikan ke dalam antarmuka sistem *website* prediksi berbasis *web*.

4. HASIL DAN PEMBAHASAN

4.1. Penerapan Algoritma dan Evaluasi Model

Algoritma *Gaussian Naïve Bayes* diuji menggunakan 41 data UMKM valid melalui mekanisme LOOCV. Pada setiap iterasi, sistem melakukan pelatihan probabilitas terhadap data latih dan mengklasifikasikan data uji ke dalam kelas "Berhasil" atau "Tidak Berhasil".

Secara keseluruhan, mekanisme *Leave-One-Out Cross-Validation* berhasil mengklasifikasikan keberhasilan dan kegagalan UMKM dengan tingkat akurasi mencapai 82,92%. Dari matriks kebingungan (*confusion matrix*), tercatat 18 data UMKM berhasil diklasifikasikan dengan benar ke kategori "Berhasil" dan 16 data ke kategori "Tidak Berhasil". Meskipun akurasinya tergolong baik, terdapat sebuah pola kesalahan yang menarik untuk dikaji lebih dalam. Dari total 7 kesalahan prediksi, sebanyak 6 di antaranya merupakan *False Positive*, yaitu UMKM yang faktualnya "Berhasil" justru diprediksi "Tidak Berhasil" sementara hanya terdapat 1 False Negative.

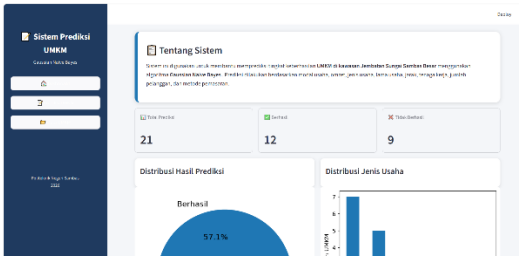
Dominasi kesalahan False Positive ini mengindikasikan sebuah fenomena unik yang dipicu oleh kombinasi dari metode LOOCV dan karakteristik algoritma Naïve Bayes. Pada mekanisme LOOCV, ketika data UMKM dengan kategori "Berhasil" dikeluarkan sementara dari dataset untuk dijadikan data uji, model cenderung kehilangan sedikit kemampuan dalam mengenali pola keberhasilan spesifik dari data tersebut. Ditambah lagi, *Naïve Bayes* sangat bergantung pada probabilitas prior. Diduga kuat, keenam UMKM tersebut memiliki karakteristik yang berada di batas ambang keberhasilan, misalnya omzet yang pas di angka median atau jarak lokasi yang relatif jauh. Kondisi ini membuat nilai probabilitas kondisional dari fitur-fitur mereka justru lebih kuat mendorong prediksi ke arah kelas "Tidak Berhasil", membuat model terkesan sedikit "kaku" dalam menilai keberhasilan usaha.

Kekakuan model ini secara langsung tercermin pada angka *Precision* yang tercatat sebesar 72,73%. Angka ini berarti bahwa ketika model menyatakan sebuah UMKM "Gagal", hanya sekitar 72,73% yang benar-benar gagal di dunia nyata, sementara sisanya (sekitar 27%) adalah kesalahan menuduh usaha yang sebenarnya berhasil sebagai gagal. Namun di balik kelemahan tersebut, model ini memiliki keunggulan luar biasa dalam hal sensitivitas, yang dibuktikan dengan nilai *Recall* sangat tinggi yakni 94,12%. Angka ini menunjukkan bahwa "mata" sistem sangat tajam dalam mendeteksi kegagalan; dari total 17 UMKM yang faktualnya memang gagal, model berhasil menangkap hampir semuanya (16 data) dan hanya membiarkan 1 UMKM gagal lolos dari radar (False Negative). Dalam konteks bisnis, *Recall* yang tinggi ini merupakan aset berharga karena sistem menjadi sangat peka terhadap sinyal bahaya kegagalan usaha. Akhirnya, keseimbangan antara kelemahan dalam presisi dan keunggulan dalam recall ini berhasil dijaga dengan baik, yang ditunjukkan oleh nilai F1-Score sebesar 82,05%. Secara keseluruhan, meskipun sedikit keras dalam menilai keberhasilan, kinerja model dalam mendeteksi kegagalan UMKM tetap bisa dikatakan sangat baik dan layak untuk dijadikan alat bantu deteksi dini.

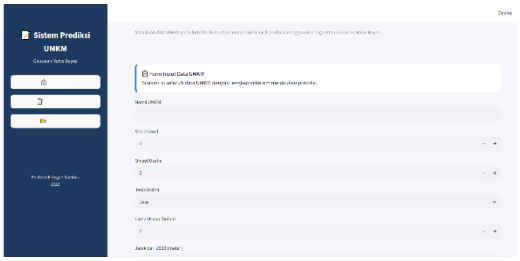
4.2. Implementasi Sistem Prediksi

Model yang telah tervalidasi selanjutnya di-*deploy* ke dalam bentuk sistem *website* prediksi. Sistem ini dirancang tidak hanya sebagai mesin penghitung, tetapi sebagai alat bantu evaluasi yang komprehensif. Fitur utama sistem meliputi:

1. *Dashboard* dan *Input Data*: Pengguna (pelaku usaha atau pemerintah) dapat memasukkan karakteristik usaha melalui formulir interaktif.

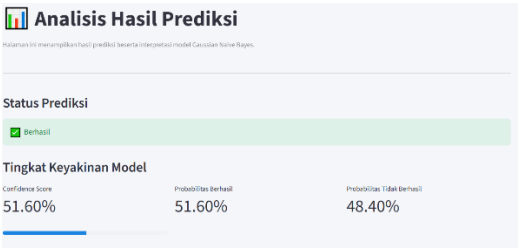


Gambar 1. Halaman *dashboard*



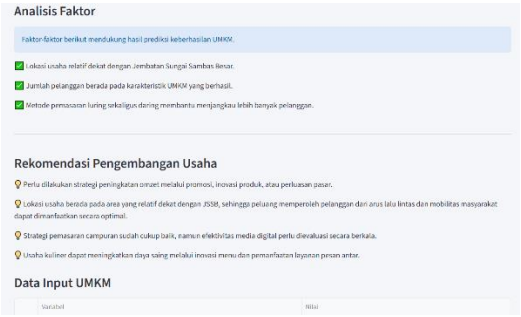
Gambar 2. Halaman *input* data dan prediksi

2. Proses Prediksi dan Probabilitas: Sistem tidak hanya menampilkan label "Berhasil/Tidak Berhasil", tetapi juga menampilkan nilai probabilitas dan tingkat keyakinan model terhadap keputusan tersebut.



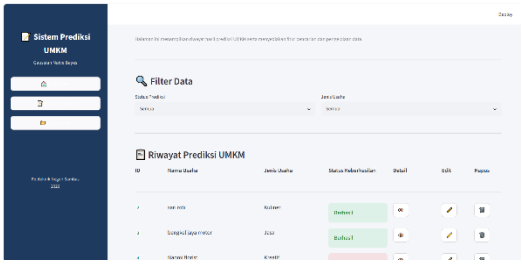
Gambar 3. Status prediksi dan probabilitas

3. Analisis Faktor dan Rekomendasi: Sistem memberikan analisis terhadap faktor yang paling mempengaruhi hasil prediksi, memberikan panduan strategis bagi pelaku usaha.



Gambar 4. Analisis faktor dan rekomendasi

4. Manajemen Data (CRUD): Terdapat fitur riwayat prediksi, edit data, dan analisis ulang otomatis jika data UMKM mengalami perubahan karakteristik di kemudian hari.



Gambar 5. Halaman riwayat data prediksi UMKM

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pembahasan, dapat disimpulkan bahwa: (1) Algoritma *Gaussian Naïve Bayes* berhasil diterapkan untuk memprediksi keberhasilan UMKM di kawasan JSSB dengan mengintegrasikan delapan variabel karakteristik usaha, di mana variabel "jarak lokasi usaha dari JSSB" berhasil menjadi penentu spesifik yang membedakan penelitian ini dengan kajian umum lainnya; (2) Tingkat akurasi yang dihasilkan oleh algoritma *Gaussian Naïve Bayes* dengan skema validasi LOOCV mencapai 82,92%, dengan rincian *Precision* 72,73%, *Recall* 94,12%, dan *F1-Score* 82,05%. Nilai Recall yang sangat tinggi ini membuktikan bahwa algoritma mampu meminimalkan risiko tertutupnya peluang UMKM potensial, sehingga sangat andal dan stabil untuk dataset berukuran kecil; (3) Algoritma tersebut berhasil diimplementasikan ke dalam sistem *website* prediksi yang mampu menyajikan hasil prediksi beserta tingkat probabilitas, analisis faktor, dan rekomendasi secara interaktif.

Saran yang dapat diberikan untuk pengembangan penelitian selanjutnya meliputi: (1) Pengembangan dataset dengan menambah jumlah sampel UMKM seiring berjalannya waktu pasca-operasional jembatan, sehingga pola ekonomi yang terbentuk dapat direpresentasikan dengan lebih komprehensif; (2) Penambahan variabel lain yang relevan seperti pendapatan bersih dan tingkat persaingan usaha; (3) Melakukan perbandingan performa algoritma *Gaussian Naïve Bayes* dengan algoritma klasifikasi lainnya yang mampu menangani ketergantungan antar variabel, untuk mengetahui algoritma mana yang paling optimal dalam konteks spesifik ini.

DAFTAR PUSTAKA

eko Purwanto, R. I. (2023). Digital Marketing Sebagai Strategi Pengembangan Pasar Pada Umkm Di Kelurahan Rungkut Mananggal. *Jics: Journal Of International Community Service*, *2*(2), 51.

Jose Julian Hidayat, M. R. (2026). Evaluasi Kinerja Algoritma Ensemble Learning Pada Klasifikasi Penyakit Diabetes Berbasis Boosting Method. *Jurnal Surya Informatika*, *16*(1), 75.

Msy Aulia Hasanah, S. S. (2021). Implementasi Crisp-Dm Model Menggunakan Metode Decision Tree Dengan Algoritma Cart Untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal Of Applied Informatics And Computing (Jaic)*, *5*(2), 103-108.

Muhammad Reza Romadhoni, R. D. (2025). Penerapan Machine Learning Algoritma Random Forest Untuk Prediksi Keberhasilan Umkm. *Journal Of Software Engineeringand Technology*.

Piliang, L. H. (2024). Umkm Penggerak Roda Perekonomian Nasional. *Public Administration Journal*, *8*(1).

Putra Bagus Utama, D. A. (2026). Tingkat Keberhasilan Umkm Berdasarkan Modal Dan Strategi Pemasaran Dengan Naive Bayes. *Journal Of Science And Social Research*, *9*(1), 156-162.

Rahmadini, E. E. (2023). Penerapan Data Mining Untuk Memprediksi Harga Bahan Pangan Di Indonesia Menggunakan Algoritma K-Nearest Neighbor. *Jurnal Mahasiswa Akuntansi Samudra (Jmas)*, *4*(4), 228-229.

Riana Rahmadanti, S. S. (2024). Penerapan Algoritma Naive Bayes Untuk Prediksi Tren Pertumbuhan Umkm Berdasarkan Jenis Usaha di Kota Depok . *Jurnal INSAN (Journal of Information Systems Management Innovation)*, *4*(2).

TRIULI NOVIANTI, I. S. (2025). Implementasi machine Learning pada Industri PLN Untuk Klasifikasi Gangguan Listrik Berdasarkan Data DFR. *Journal of manufacturing in Industrial Engineering and Technology (MINE-TECH)*, *4*(2).

YOGA NAUFAL RAY PUTRO, A. A. (2024). Penggunaan Algoritma Gaussian Naive Bayes & Decision Tree Untuk Klasifikasi Tingkat Kemenangan Pada Game Mobile Legends. *Jurnal Komputer dan Informatika*, *6*(1), 12-24.