

# Implementasi Algoritma Naive Bayes pada Sistem Klasifikasi Kelayakan UMKM Penerima Bantuan Modal Usaha

Agung Anggara<sup>1</sup>, M. Khadafi<sup>2\*</sup>, Muhammad Reza Zulman<sup>3</sup>

<sup>1,2,3</sup> Jurusan Teknologi Informasi dan Komputer, Politeknik Negeri Lhokseumawe, Indonesia

\*Penulis Korespondensi: mkhadafi@pnl.ac.id

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



## Abstrak

Keberhasilan program pemberdayaan UMKM sangat bergantung pada ketepatan dalam menentukan kelayakan penerima bantuan modal usaha. Penelitian ini membahas perancangan dan pembangunan sistem klasifikasi berbasis web untuk menentukan kelayakan Usaha Mikro, Kecil, dan Menengah (UMKM) dalam menerima bantuan modal usaha dari Baitul Mal Aceh Utara. Permasalahan utama yang dihadapi adalah proses seleksi penerima bantuan yang masih dilakukan secara manual, memakan waktu, dan berpotensi menimbulkan subjektivitas. Sebagai solusi, penelitian ini mengimplementasikan algoritma *Naive Bayes* karena kemampuannya memproses data dengan cepat, sederhana, dan konsisten. Data penelitian mencakup empat belas variabel bebas dan satu variabel terikat, yaitu status kelayakan (Layak/Tidak Layak) serta kategori penerima modal (Rp3.000.000, Rp5.000.000, dan Rp10.000.000). Proses penelitian meliputi pengumpulan data, preprocessing, pelabelan, pelatihan model, evaluasi menggunakan classification report, serta implementasi sistem berbasis web terintegrasi Flask API dan PHP. Hasil pengujian menunjukkan bahwa model *Naive Bayes* mencapai akurasi sebesar 95%, dengan nilai precision, recall, dan F1-score masing-masing 0,95, yang menandakan bahwa model mampu melakukan klasifikasi kelayakan dan kategori penerima modal usaha dengan baik.

**Kata kunci:** *Naive Bayes*, UMKM, klasifikasi, *machine learning*, Baitul Mal Aceh Utara

## 1. Pendahuluan

Usaha Mikro, Kecil, dan Menengah (UMKM) berperan penting dalam memperkuat perekonomian daerah. Namun, proses seleksi penerima bantuan modal usaha di lembaga seperti Baitul Mal Aceh Utara masih dilakukan secara manual sehingga rawan kesalahan dan subjektivitas. Kondisi ini mendorong kebutuhan sistem pendukung keputusan berbasis teknologi yang mampu menilai kelayakan secara objektif [1].

Berbagai penelitian telah menerapkan algoritma *machine learning* untuk klasifikasi keputusan, seperti Decision Tree, Random Forest, dan *Naive Bayes*. Di antara algoritma tersebut, *Naive Bayes* banyak digunakan karena sederhana, efisien, dan mampu memberikan hasil akurat pada dataset terbatas [2]. Oleh karena itu, penelitian ini merancang sistem klasifikasi berbasis web menggunakan algoritma *Naive Bayes* untuk menentukan kelayakan UMKM penerima bantuan modal di Baitul Mal Aceh Utara.

Pemanfaatan teknologi *machine learning* memungkinkan sistem untuk mempelajari pola data dari penerima sebelumnya dan menghasilkan prediksi kelayakan yang lebih objektif berdasarkan parameter tertentu seperti omzet, lama usaha, aset, dan pengalaman bisnis

Tujuan utama penelitian ini adalah mengevaluasi performa algoritma *Naive Bayes* berdasarkan metrik accuracy, precision, recall, dan F1-score sehingga dapat diketahui tingkat keakuratan model dalam menentukan keputusan kelayakan serta kategori penerimaan modal usaha [3]. Hasil penelitian ini diharapkan dapat mengklasifikasikan kelayakan UMKM, sekaligus memberikan gambaran seberapa akurat model tersebut dalam mendukung proses penentuan penerima modal usaha di Baitul Mal Aceh Utara.

## 2. Metode

### 2.1 Dataset

Penelitian ini menggunakan data sekunder yang diperoleh dari lembaga Baitul Mal Aceh Utara. Data tersebut merupakan arsip digital dan dokumen resmi yang mencatat informasi terkait profil UMKM dan status penerimaan bantuan modal usaha. Desain data ditunjukkan pada Gambar 1.

Tabel 1 Variabel Input

No	Nama Variabel	Jenis Data	Kategori/Nilai	Keterangan
1	omzet_harian	Numerik/Kategorikal	Nilai rupiah (contoh: 90.000, 200.000, dst)	Pendapatan rata-rata per hari dari hasil penjualan
2	lama_usaha	Numerik/Kategorikal	1, 2, 3, ... (tahun)	Lama usaha telah berjalan berdasarkan tahun berdiri
3	keuntungan_bulan	Numerik/Kategorikal	Rendah, Sedang, Tinggi	Rata-rata laba bersih yang diperoleh dalam satu bulan
4	stok_bergerak	Kategorikal	Barang, Produksi, Ekspansi	Jenis perputaran atau aktivitas utama dalam usaha
5	promosi	Kategorikal	Sering, Kadang, Tidak Pernah	Frekuensi pelaku usaha melakukan promosi
6	tujuan_modal	Kategorikal	Produksi, Barang, Ekspansi, Peralatan	Tujuan utama penggunaan dana bantuan modal
7	catatan_keuangan	Kategorikal	Ada, Tidak Ada	Status keberadaan pembukuan atau catatan keuangan usaha
8	utang_bermasalah	Kategorikal	Ya, Tidak	Kondisi apakah pelaku usaha memiliki utang bermasalah
9	lokasi_tempat	Kategorikal	Strategis, Tidak Strategis	Lokasi tempat usaha dilihat dari akses dan potensi pasar
10	kebersihan	Kategorikal	Baik, Sedang, Buruk	Kondisi kebersihan dan kerapian tempat usaha
11	tenaga_kerja	Numerik/Kategorikal	0, 1, 2, 3, dst	Jumlah pekerja yang terlibat dalam kegiatan usaha
12	daya_saing	Kategorikal	Rendah, Sedang, Tinggi	Tingkat kemampuan usaha dalam bersaing dengan usaha lain
13	pernah_dibantu	Kategorikal	1 (Pernah), 0 (Belum Pernah)	Status apakah pelaku usaha pernah menerima bantuan sebelumnya
14	kelompok_individu	Kategorikal	Kelompok, Individu	Status kepemilikan usaha berdasarkan pelaku

Gambar 1 memperlihatkan Parameter input yang di gunakan untuk klasifikasi data umkm sehingga menghasilkan kelas layak dan tidak layak serta kategori penerimaan modal usaha. Variabel yang digunakan dalam penelitian ini diambil dari data hasil observasi dan wawancara pelaku UMKM penerima bantuan modal usaha di Baitul Mal Aceh Utara.

## 2.2 Algoritma Naïve Bayes

Algoritma Naïve Bayes adalah algoritma klasifikasi probabilistik yang berdasarkan pada Teorema Bayes dengan asumsi independensi antar atribut. Naïve Bayes sering digunakan karena keunggulannya dalam

kecepatan dan kemudahan implementasi, terutama untuk dataset besar. Teorema Bayes seperti berikut:

$$P(H | X) = \frac{P(X|H)xP(H)}{P(X)}$$

X : Data yang kelasnya belum diketahui

H : Hipotesis bahwa data X termasuk ke dalam kelas tertentu

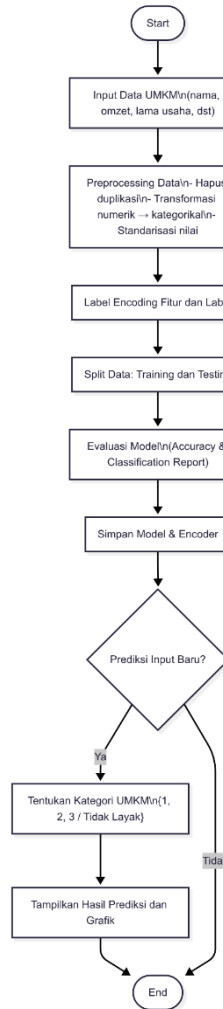
P(H | X) : Probabilitas hipotesis H dengan kondisi data X (posterior probability).

P(H) : Probabilitas awal hipotesis H (prior probability).

P(X | H) : Probabilitas data X dengan asumsi hipotesis H.

P(X) : Probabilitas data X secara keseluruhan

Model *Naive Bayes* digunakan untuk memproses klasifikasi data UMKM dengan *input* berupa data *historis* dari baitul mal aceh utara. Arsitektur model ditunjukkan pada Gambar 2.



Gambar 1. Desain Arsitektur Naive Bayes

Gambar 2 menunjukkan alur proses sistem klasifikasi kelayakan UMKM menggunakan algoritma Naive Bayes. Proses diawali dengan input data UMKM seperti nama, omzet, dan lama usaha. Data kemudian dibagi menjadi dua bagian, yaitu Training dan Testing dengan membagi dataset menjadi data latih sebesar 80% dan data uji sebesar 20%. untuk melatih dan menguji performa model. Setelah pelatihan, dilakukan Evaluasi Model menggunakan accuracy, precision, recall, F1-score, dan confusion matrix. Hasil evaluasi tidak hanya menunjukkan tingkat ketepatan model, tetapi juga memberikan gambaran kelas mana yang sering salah prediksi, sehingga dapat menjadi dasar perbaikan sistem agar rekomendasi penyaluran modal lebih akurat dan adil.

Data tersebut kemudian melewati tahap preprocessing, yaitu penghapusan duplikasi, transformasi nilai numerik menjadi kategorikal, serta standarisasi data. Selanjutnya dilakukan label encoding pada fitur dan label agar dapat diproses oleh model. Data kemudian dibagi menjadi data training dan testing untuk melatih serta menguji performa model. Hasil evaluasi diperoleh melalui nilai akurasi dan classification report, kemudian model dan encoder disimpan. Jika terdapat input baru, sistem akan melakukan prediksi kategori UMKM (misalnya kategori 1, 2, 3, atau tidak layak), lalu menampilkan hasil prediksi dan grafik visualisasi. Jika tidak ada input baru, proses langsung berakhir.

### 2.3 Confusion Matrix

Confusion Matrix merupakan metode evaluasi yang digunakan untuk menilai kinerja model klasifikasi dengan membandingkan antara hasil prediksi model dan data aktual. Matriks ini berbentuk tabel dua dimensi yang menampilkan jumlah prediksi benar dan salah untuk setiap kelas [4]. Dalam penelitian ini, Confusion Matrix digunakan untuk mengetahui seberapa baik algoritma Naïve Bayes dalam mengklasifikasikan kelayakan UMKM ke dalam kategori yang sesuai [5].

Tabel Confusion Matrix terdiri atas empat komponen utama, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Nilai TP menunjukkan jumlah data yang benar diprediksi sebagai kelas positif, sedangkan TN menunjukkan data negatif yang berhasil diprediksi dengan benar. Sementara itu, FP menggambarkan jumlah data negatif yang salah diprediksi sebagai positif, dan FN menunjukkan data positif yang salah diprediksi sebagai negatif.

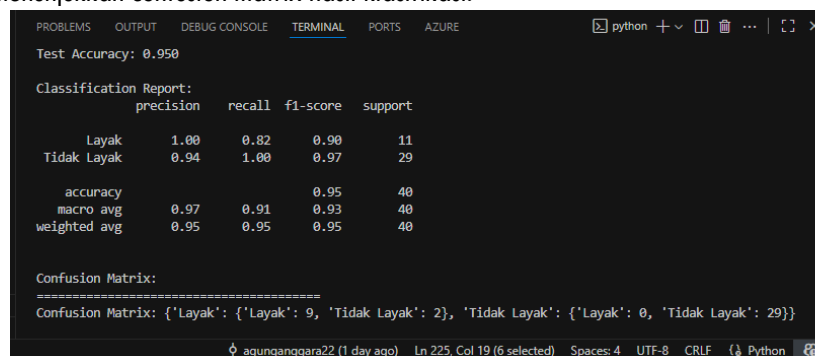
Dari Confusion Matrix tersebut, dapat dihitung berbagai metrik evaluasi seperti akurasi, presisi, recall, dan f1-score, yang memberikan gambaran menyeluruh tentang performa model dalam mengenali dan mengklasifikasikan data UMKM. Dengan demikian, Confusion Matrix menjadi dasar penting dalam mengevaluasi efektivitas model klasifikasi Naïve Bayes yang digunakan pada penelitian ini.

## 3. Hasil dan Pembahasan

Bagian ini membahas hasil implementasi dari perancangan yang sudah dijabarkan pada bab sebelumnya. Bab ini juga menjelaskan beberapa fitur yang ada pada web klasifikasi kelayakan umkm.

### 3.1 Pengujian Model Layak dan Tidak Layak

Model Naive Bayes diuji menggunakan dataset Baitul Mal Aceh Utara. Dari hasil evaluasi, diperoleh akurasi 95%, yang berarti model dapat memprediksi kelayakan dengan sangat baik. Precision, recall, dan F1-score masing-masing 0.95 menandakan model memiliki performa yang seimbang untuk kelas Layak dan Tidak Layak. Gambar 2 berikut menunjukkan confusion matrix hasil klasifikasi.



```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS AZURE
Test Accuracy: 0.950

Classification Report:
precision    recall  f1-score   support

   Layak      1.00    0.82    0.90        11
  Tidak Layak  0.94    1.00    0.97        29

 accuracy          0.95        40
 macro avg         0.97        40
 weighted avg      0.95        40

Confusion Matrix:
=====
Confusion Matrix: {'Layak': {'Layak': 9, 'Tidak Layak': 2}, 'Tidak Layak': {'Layak': 0, 'Tidak Layak': 29}}

```

Gambar 2. Hasil Confusion Matrix Label

Gambar 2 di atas menunjukkan hasil evaluasi model Naïve Bayes berdasarkan classification report dan confusion matrix. Nilai akurasi 0.95 menunjukkan bahwa 95% dari total data uji berhasil diklasifikasikan dengan benar. Kelas Layak memiliki nilai precision 1.00 yang berarti semua prediksi “Layak” benar, namun recall-nya 0.82 menandakan masih ada beberapa data Layak yang tidak terdeteksi dengan benar. Sebaliknya, kelas Tidak Layak memiliki recall 1.00 yang berarti semua data Tidak Layak berhasil dikenali, dengan precision 0.94 yang menunjukkan ada sebagian kecil prediksi Tidak Layak yang keliru. Nilai F1-score 0.90 pada kelas Layak dan 0.97 pada kelas Tidak Layak menunjukkan keseimbangan antara presisi dan sensitivitas model. Berdasarkan confusion matrix, terdapat 9 data Layak dan 29 data Tidak Layak yang diklasifikasikan benar, sedangkan 2 data Layak salah diklasifikasikan sebagai Tidak Layak.

```

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS  AZURE

Kategori Test Accuracy: 0.975
Kategori Classification Report:
      precision  recall  f1-score  support
0         0.97    1.00    0.98     31
0         0.97    1.00    0.98     31
1         1.00    1.00    1.00     5
2         1.00    0.50    0.67     2
3         1.00    1.00    1.00     2

accuracy                0.97     40
macro avg             0.99    0.88    0.91     40
weighted avg         0.98    0.97    0.97     40

```

Gambar 3. Hasil Confusion Matrix Kategori

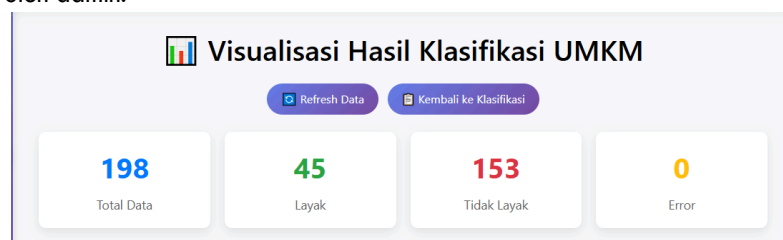
Gambar 3 di atas menampilkan hasil evaluasi model Naive Bayes untuk klasifikasi kategori penerima modal usaha, dengan tingkat akurasi sebesar 0.975 atau 97,5%. Nilai ini menunjukkan bahwa model mampu mengklasifikasikan sebagian besar data kategori dengan benar. Kelas kategori 0 dan 1 memiliki nilai precision dan recall sebesar 0.97–1.00, menandakan bahwa hampir seluruh data pada kelas tersebut terprediksi secara tepat. Kelas 2 memiliki recall 0.50 dan F1-score 0.67, menunjukkan masih ada beberapa data kategori 2 yang salah klasifikasi. Sedangkan kelas 3 memiliki nilai sempurna (1.00) untuk semua metrik, menunjukkan prediksi yang sangat akurat. Secara keseluruhan, nilai rata-rata macro F1-score sebesar 0.91 dan weighted F1-score sebesar 0.97 menunjukkan bahwa model memiliki performa yang sangat baik dan konsisten dalam mengklasifikasikan kategori penerima modal usaha dengan tingkat kesalahan yang sangat rendah.

### 3.2 Fitur Aplikasi

Fitur utama sistem untuk melakukan klasifikasi kelayakan UMKM menggunakan algoritma Naive Bayes. Admin memilih file CSV yang berisi data Kriteria UMKM yang sudah di nilai oleh petugas dan mengunggahnya ke sistem. Dataset ini kemudian akan diproses oleh model Naive Bayes untuk menentukan kelayakan. Jika Sudah klik tombol mulai klasifikasi dan tunggu hasil di proses dan akan muncul hasil klasifikasi seperti berapa jumlah data yang di uji, jumlah error jika ada data yang tidak valid, jumlah layak dan tidak layak. Serta statistik per kategori di mana pelaku umkm yang layak di beri modal usaha akan di bagi menjadi tiga kategori yaitu kategori I 3 juta, kategori II 5 juta dan kategori III 10 juta.

### 3.3 Hasil Klasifikasi Kelayakan UMKM

visualisasi hasil klasifikasi dalam bentuk grafik. Grafik diagram lingkaran dan diagram batang digunakan untuk memperlihatkan distribusi status kelayakan dan kategori bantuan yang diterima oleh UMKM, agar memudahkan analisis oleh admin.



Gambar 4 Visualisasi Hasil Klasifikasi

Gambar 4 menampilkan Hasil Data yang di klasifikasi dan tombol refresh data jika ada dataset baru yang di klasifikasi dan tombol kembali ke klasifikasi jika belum ada data klasifikasi.

## 4. Kesimpulan

Hasil pengujian ini berhasil mengembangkan Model klasifikasi kelayakan UMKM penerima bantuan modal usaha menggunakan algoritma Naive Bayes dengan akurasi 95%. Nilai precision, recall, dan F1-score 0,95 menunjukkan keseimbangan performa model dalam mengklasifikasikan kelas Layak dan Tidak Layak. Sistem berbasis web yang terintegrasi antara Flask dan PHP mampu membantu proses seleksi menjadi lebih Cepat dan efisien. Penelitian selanjutnya dapat membandingkan hasil ini dengan algoritma lain seperti Decision Tree atau Random Forest untuk mengukur peningkatan akurasi.

## REFERENSI

- [1] G. Nguyen et al., 'Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey,' *Artificial Intelligence Review*, vol. 52, no. 1, 2019.
- [2] C. Shorten and T. M. Khoshgoftaar, 'A survey on Image Data Augmentation for Deep Learning,' *Journal of Big Data*, vol. 6, no. 1, 2019.
- [3] R. Vinayakumar et al., 'Deep Learning Approach for Intelligent Intrusion Detection System,' *IEEE Access*, vol. 7, 2019.
- [4] I. P. Sari dan D. Fitriani, 'Penerapan Algoritma Naive Bayes untuk Klasifikasi KelayakanUMKM Penerima Bantuan Modal,' *Jurnal Teknologi dan Sistem Informasi*, vol. 9, no. 2, 2023.
- [5] S. Fadli, 'Sistem Pendukung Keputusan Pemberian Bantuan Modal UMKM Menggunakan Metode Naive Bayes,' *Jurnal Informatika Politeknik Negeri Lhokseumawe*, vol. 4, no. 1, 2022.
- [6] A. Yuliana et al., 'Perbandingan Algoritma Naive Bayes dan Decision Tree dalam Prediksi Kelayakan Kredit UMKM,' *Jurnal Teknologi Informasi dan Komputer*, vol. 8, no. 3, 2021.
- [7] B. Rahman, 'Implementasi Machine Learning dalam Sistem Pendukung Keputusan,' *Jurnal Teknologi Rekayasa*, vol. 7, no. 2, 2022.
- [8] H. Pratama, 'Analisis Kinerja Model Naive Bayes untuk Klasifikasi Data Keuangan,' *Jurnal Sistem Cerdas*, vol. 5, no. 4, 2021.