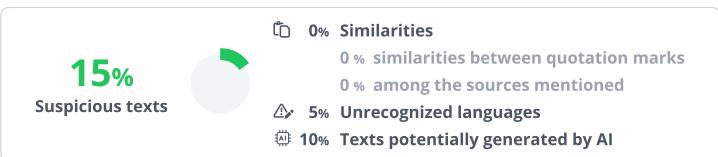


211080200090_archive umsida



Document name: 211080200090_archive umsida.pdf
Document ID: 7bcc9a93c2753e0a87989f46b71af89349ce0e29
Original document size: 607.24 KB

Submitter: UMSIDA Perpustakaan
Submission date: 10/8/2025
Upload type: interface
analysis end date: 10/8/2025

Number of words: 4,319
Number of characters: 32,835

Location of similarities in the document:

☰ Sources of similarities

Referenced source (without similarities detected) These sources were cited in the paper without finding any similarities.

1  <https://idm.kemendesa.go.id/>

Classification of the Village Development Index in East Java Province

2024 Using Backpropagation Neural Network and Naïve Bayes

Klasifikasi Indeks Desa Membangun Provinsi Jawa Timur 2024

Menggunakan Jaringan Saraf Tiruan Backpropagation dan Naïve

Bayes



Naila Farah Diba, Novia Ariyanti, Ade Eviyanti, Mochamad Alfan Rosid, Yulian Findawati

Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi: noviaariyanti@umsida.ac.id

Abstract.

The Village Development Index (Indeks Desa Membangun/IDM) is an important instrument for measuring village development. This study compares two classification methods, Backpropagation and Naïve Bayes, in determining village status using the 2024 IDM dataset of East Java. The evaluation was carried out using accuracy, precision, recall, F1-score, and confusion matrix metrics. The results indicate that Naïve Bayes is more stable, achieving an accuracy of 0.80 on the training set and slightly increasing to 0.81 on the test set, which demonstrates good generalization ability. In contrast, Backpropagation achieved 0.7802 accuracy on the training set but dropped to 0.7540 on the test set, showing less consistent performance, particularly across several categories. In conclusion, simpler methods such as Naïve Bayes can outperform more complex models in regional datasets that are relatively small with linear features, while Backpropagation is more suitable for national-scale data with larger size and more complex non-linear patterns.

Keywords – Village Development Index (IDM),



Classification, Naïve Bayes, Backpropagation,

Machine Learning

Abstrak. Indeks Desa Membangun (IDM) merupakan instrumen penting untuk mengukur perkembangan desa. Penelitian ini membandingkan dua metode klasifikasi, yaitu Backpropagation dan Naïve Bayes, dalam menentukan status desa pada data IDM Jawa Timur 2024. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, F1-score, serta

confusion matrix.

Hasil menunjukkan bahwa Naïve Bayes lebih stabil dengan akurasi 0,80 pada training set dan meningkat menjadi 0,81 pada test set, menandakan kemampuan generalisasi yang baik. Sebaliknya, Backpropagation mencatat akurasi 0,7802 pada training set namun menurun menjadi 0,7540 pada test set, dengan performa yang kurang konsisten terutama pada beberapa kategori. Kesimpulannya, metode sederhana seperti Naïve Bayes dapat

lebih unggul pada dataset regional yang berukuran relatif kecil dengan fitur linear, sementara Backpropagation lebih sesuai untuk skala nasional dengan data besar dan pola non-linear yang kompleks.

Kata Kunci – Indeks Desa Membangun (IDM),



Klasifikasi, Naïve Bayes, Backpropagation,

Pembelajaran Mesin

I. PENDAHULUAN

Menurut Undang-Undang Nomor 6 Tahun 2014 tentang Desa, desa adalah batas wilayah terkecil yang berwenang untuk mengatur dan mengurus urusan pemerintahan demi kepentingan masyarakat setempat. Peraturan tersebut memberikan sudut pandang baru tentang pembangunan desa, sehingga desa memiliki hak penuh untuk mengelola pembangunan di daerahnya masing-masing untuk mengurangi tingkat kesenjangan yang terjadi pada wilayah pedesaan [1]. Pembangunan di wilayah pedesaan merupakan upaya strategis pemerintah untuk mencapai pembangunan nasional secara merata.

Permendagri No.114 Tahun 2014 menjadi asas dasar upaya pembangunan desa, dengan tujuan pembangunan dan mensejahterahkan masyarakat desa. Untuk memetakan setiap kondisi desa guna merancang kebijakan pembangunan yang tepat untuk mendorong kemajuan desa, Kementerian Desa menyusun Indeks Desa Membangun (IDM) [2]. IDM sendiri dikembangkan untuk menentukan lokasi prioritas dalam upaya pengentasan desa tertinggal dan percepatan pembangunan desa mandiri [3].

Berbeda dari format indikator sebelumnya, yang terdiri dari tiga indikator utama, Indeks Ketahanan Sosial (IKS), Indeks Ketahanan Ekonomi (IKE), dan Indeks Ketahanan Lingkungan (IKL) [2]. Dilansir dari situs resmi milik Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Republik Indonesia (KemenPAN-RB), pemerintah meluncurkan indikator tunggal dalam data IDM yang terdiri dari enam dimensi utama, yakni layanan dasar, sosial, ekonomi, lingkungan, aksesibilitas, dan tata kelola pemerintahan desa.

Berdasarkan Keputusan Menteri Desa PDTT No. 400 Tahun 2024, Provinsi Jawa Timur (Jatim) kembali berhasil menorehkan provinsi tertinggi dengan desa terindeks mandiri di Indonesia. Dilansir dari laman resmi Kominfo Jatim tercatat jumlah desa mandiri di provinsi tersebut sebanyak 4.019 desa. Jumlah desa yang naik menjadi desa mandiri kian naik melesat setiap tahunnya. Pada tahun 2024 jumlah desa mandiri di Jatim meningkat sebanyak 43.54%. Lonjakan yang konsisten ini membuat Provinsi Jatim konsisten berada di peringkat pertama nasional provinsi dengan indeks desa mandiri tertinggi selama lima tahun berturut-turut sejak tahun 2019-2024.

2 | Page

Meski sebagian besar desa di provinsi ini telah masuk dalam kategori maju dan mandiri, namun demikian, masih terdapat sejumlah desa yang memerlukan perhatian lebih dalam indikator IDM. Hal ini menunjukkan bahwa masih terdapat kesenjangan pada suatu wilayah yang berkembang pesat. Kesenjangan internal antar desa tetap perlu diperhatikan melalui pemetaan yang lebih detail [4].

Kemajuan teknologi telah mendorong perkembangan machine learning yang kini banyak diterapkan di berbagai sektor, salah satunya sektor publik. Dengan bantuan algoritma pada machine learning, klasifikasi pada data IDM dengan machine learning membantu mengidentifikasi pola dan mengevaluasi capaian pembangunan desa pada pengelompokan kategori status desa menurut IDM, yakni mandiri, maju, membangun, berkembang, terbelakang, dan sangat terbelakang [2].



Pada penelitian ini, penulis menggunakan dua metode klasifikasi, yaitu Jaringan Saraf Tiruan (JST)

Backpropagation dan Naïve Bayes. Algoritma backpropagation diterapkan sebagai metode pelatihan pada model JST.

Backpropagation sebagai salah satu algoritma dasar dalam JST yang mampu mempelajari pola data non-linear secara efektif serta cukup fleksibel untuk diaplikasikan dalam berbagai data mining tasks, baik klasifikasi maupun prediksi

[5]. Penelitian ini menggunakan Multilayer Perceptron (MLP) dengan algoritma Backpropagation sebagai classifier.

Karena tujuan penelitian adalah klasifikasi, maka diperlukan lebih dari satu neuron pada lapisan output, sesuai dengan

jumlah kategori [6]. Sementara itu, Naïve Bayes Classifier merupakan metode machine learning yang telah lama

digunakan dan terbukti baik dalam menyelesaikan berbagai permasalahan klasifikasi. Algoritma ini bekerja dengan

asumsi independensi antar variabel prediktor, sehingga disebut "naïve" [7]. Penelitian ini menggunakan metode

Gaussian Naïve Bayes karena model ini sesuai untuk klasifikasi pada dataset IDM Jatim yang memiliki atribut bersifat

kontinu [8].

II. METODE

Tahapan Penelitian

Tahapan penelitian merupakan serangkaian langkah sistematis yang dilakukan mulai dari perencanaan awal hingga

evaluasi akhir guna mencapai tujuan penelitian. Pada penelitian ini, proses dimulai dari studi literatur, dilanjutkan

dengan pengumpulan data, tahap pre-processing, pemrosesan data menggunakan algoritma Backpropagation dan

Naïve Bayes, hingga tahap validasi model. Penelitian ini dilakukan dengan menggunakan data IDM tahun 2024 di

Kabupaten Sidoarjo. Rangkaian tahapan penelitian ditunjukkan pada gambar berikut:

Gambar 1. Tahapan Penelitian

1. Pengumpulan Data

Pengumpulan data dilakukan dengan mengakses IDM tahun 2024 melalui laman resmi dari Kemendesa

"<https://idm.kemendesa.go.id/>". Provinsi Jawa Timur secara keseluruhan, termasuk setiap kecamatan dan

desanya, menjadi fokus area dalam pengumpulan data. Dataset yang diperoleh mencakup 66 atribut, dengan

informasi yang berkaitan dengan dimensi utama IDM dan status perkembangan desa, disertai 7.721 baris

Page | 3

data. Variabel target dalam penelitian ini adalah status desa, yang akan diklasifikasikan menggunakan

algoritma Naïve Bayes. Data ini dianalisis menggunakan bahasa pemrograman Python melalui platform

Jupyter Notebook untuk membangun model klasifikasi prediktif. Di bawah ini tabel atribut data IDM

Tabel 1. Atribut Data IDM

No. Atribut Dimensi Atribut

X1-X17 Layanan Dasar Layanan Dasar,



SUB-DIMENSI

PENDIDIKAN, Akses terhadap PAUD/ TK/

Sederajat,

Akses terhadap SD/ MI/ Sederajat,

Akses terhadap SMP/ MTs/ Sederajat, akses

terhadap SMA/ SMK/ MA/ MAK/ Sederajat,



SUB-DIMENSI KESEHATAN, Layanan

Sarana Kesehatan,

Fasilitas Kesehatan

Poskesdes/ Polindes, Aktivitas Posyandu,

Layanan Dokter, Layanan Bidan, Layanan

Tenaga Kesehatan Lainnya, Jaminan Kesehatan

Nasional, SUB-DIMENSI UTILITAS DASAR,

Air Minum, dan Persentase Rumah Tidak

Layak Huni

X18-X28 Sosial SOSIAL,



SUB-DIMENSI AKTIVITAS,

Kearifan Sosial/ Budaya,

Frekuensi Gotong

Royong, Kegiatan Olahraga, Mitigasi dan

Penanganan Konflik Sosial, Satkamling, SUB-

DIMENSI FASILITAS MASYARAKAT,

Taman Bacaan Masyarakat/ Perpustakaan Desa,

Fasilitas Olahraga, dan Keberadaan Ruang

Publik Terbuka

X29-X41 Ekonomi EKONOMI, Keragaman Aktivitas Ekonomi,

Produk Unggulan Desa, Ekonomi Kreatif, SUB-

DIMENSI FASILTAS PENDUKUNG

EKONOMI, Akses Terhadap Pendidikan Non-

formal/ Pusat Keterampilan/ Kursus,



'Pasar

Rakyat, 'Toko/ Pertokoan, Kedai/ Rumah

Makan, Penginapan,

Layanan Pos dan/

Logistik, Lembaga Ekonomi, dan Layanan

Keuangan

X42-X49 Lingkungan LINGKUNGAN, SUB-DIMENSI

PENGELOLAAN LINGKUNGAN, Kearifan

Lingkungan, Sistem Pengelolaan Sampah,

Tingkat Pencemaran Lingkungan, Sistem

Pembuangan Air Limbah Domestik (Rumah

Tangga), SUB-DIMENSI

PENANGGULANGAN BENCANA, dan

Penanggulangan Bencana

X50-X57 Aksesibilitas AKSESIBILITAS, SUB-DIMENSI KONDISI

AKSES JALAN, Kondisi Jalan di desa, Kondisi

Penerangan Jalan Utama Desa, SUB-DIMENSI

KEMUDAHAN AKSES, Keberadaan

Angkutan Perdesaan/ Angkutan Lokal/ Sejenis,

Akses Listrik, dan Layanan Telekomunikasi

X58-X65 Tata Kelola Pemerintahan Desa TATA KELOLA PEMERINTAHAN DESA,

SUB-DIMENSI KELEMBAGAAN DAN

PELAYANAN DESA, Pelaksanaan Pelayanan

dan Administrasi Desa, Pemanfaatan Teknologi

dalam Pelayanan Desa (SPBE), Musyawarah

Desa, SUB-DIMENSI TATA KELOLA

KEUANGAN DESA, Pendapatan Asli Desa

4 | Page

(PADes) dan Dana Desa, dan Jumlah

Kepemilikan dan Produktivitas Aset Desa.

X66 Skor Total Skor

Y Status Desa STATUS ID 2024

2. Pra-Pemrosesan

Pra-pemrosesan data merupakan tahap awal dalam implementasi algoritma Naïve Bayes pada IDM, yang mencakup proses pembersihan data, transformasi data, seleksi fitur, split data. Tahapan ini penting agar data dapat diproses dan dikomputasi oleh algoritma Backpropagation maupun Gaussian Naïve Bayes dengan lancar dan menguji kedua algoritma tersebut secara obyektif.

a. Data Cleaning

Data cleaning dilakukan untuk membersihkan dataset dari nilai kosong, teks tidak konsisten, dan data yang kurang relevan [9]. Pada dataset ini ditemukan beberapa masalah, seperti nilai “-” (kosong), data tekstual seperti “Cek Kembali” dan “Tidak Teridentifikasi” yang perlu diubah ke bentuk numerik, serta kolom dengan nilai yang seragam di seluruh baris. Penanganan dilakukan dengan simple imputer, yaitu mengganti data kosong menggunakan nilai mean [10], mengubah data teks menjadi angka, dan menghapus atribut yang tidak memberikan informasi penting agar data siap digunakan dalam proses klasifikasi [11].

b. Data Transformation

Data transformation bertujuan untuk mengonversi data kategorikal berbentuk teks menjadi nilai numerik. Pada penelitian ini, digunakan manual mapping untuk algoritma Naïve Bayes dan one-hot encoding untuk Backpropagation Multilayer Perceptron (MLP). Pada manual mapping, kategori

pada dataset IDM dikodekan menjadi 0 hingga 4, yaitu: 0 = sangat terbelakang, 1 = terbelakang, 2 = berkembang, 3 = maju, dan 4 = mandiri [12]. Sementara itu, one-hot encoding merepresentasikan kelas dalam bentuk biner menggunakan lima perceptron secara ascending, sesuai dengan pendekatan yang penulis terapkan pada manual mapping [6].

c. Seleksi Fitur

Dataset yang digunakan terdiri dari 66 atribut prediktor yang mewakili indikator-indikator dalam data IDM. Untuk menyederhanakan proses klasifikasi dan meningkatkan akurasi model, dilakukan seleksi fitur dengan menghapus beberapa kelompok atribut yang dianggap kurang berpengaruh terhadap tingkat partisipasi pemilu [13]. Penelitian ini menggunakan korelasi antar-atribut dengan metode pearson [14]. Penelitian ini memilih korelasi antar-atribut dengan nilai minimal 0,1, dan diaplikasikan pada train set. Berikut tabel di bawah ini menampilkan atribut setelah dilakukan seleksi fitur:

Tabel 2. Atribut yang tersisa setelah seleksi fitur

No. Atribut Dimensi Atribut

X7, X8,



X11, X12, X13,

X14, dan X16

Layanan Dasar SUB-DIMENSI KESEHATAN,

Layanan

Sarana Kesehatan, Fasilitas Kesehatan

Poskesdes/ Polindes, Layanan Tenaga

Kesehatan Lainnya, Jaminan Kesehatan

Nasional, SUB-DIMENSI UTILITAS

DASAR, dan Persentase Rumah Tidak Layak

Huni

X18,



X19, X20, X21,

X23, X24, X25, dan X26

Sosial SOSIAL, SUB-DIMENSI AKTIVITAS,

Kearifan Sosial/ Budaya, Kegiatan Olahraga,

Mitigasi dan Penanganan Konflik Sosial,

Satkamling, dan SUB-DIMENSI FASILITAS

MASYARAKAT

X29, X30, X31, X32,

X33, 34, X35, X37, X38,

X40, dan X41

Ekonomi EKONOMI, Keragaman Aktivitas Ekonomi,

Produk Unggulan Desa, Ekonomi Kreatif,



SUB-DIMENSI FASILITAS PENDUKUNG

EKONOMI,

Akses Terhadap Pendidikan

Non-formal/ Pusat Keterampilan/ Kursus,



'Pasar Rakyat, Kedai/ Rumah Makan,

Penginapan, Layanan Pos dan/ Logistik,

Page | 5

Lembaga Ekonomi, dan Layanan Keuangan

X42, X43, X44, X45,

X47, X48, dan X49

Lingkungan LINGKUNGAN, SUB-DIMENSI

PENGELOLAAN LINGKUNGAN, Kearifan

Lingkungan, Sistem Pengelolaan Sampah,

Sistem Pembuangan Air Limbah Domestik

(Rumah Tangga), SUB-DIMENSI

PENANGGULANGAN BENCANA, dan

Penanggulangan Bencana

X51, X52, X53, X55, dan

X57

Aksesibilitas SUB-DIMENSI KONDISI AKSES JALAN,

Kondisi Jalan di desa, Kondisi Penerangan

Jalan Utama Desa, Keberadaan Angkutan

Perdesaan/ Angkutan Lokal/ Sejenis, dan

Layanan Telekomunikasi

X59, X60, X61, dan X64 Tata Kelola

Pemerintahan Desa

SUB-DIMENSI KELEMBAGAAN DAN

PELAYANAN DESA, Pelaksanaan

Pelayanan dan Administrasi Desa,

Pemanfaatan Teknologi dalam Pelayanan

Desa (SPBE), dan Pendapatan Asli Desa

(PADes) dan Dana Desa

d. Random Dataset

Pengacakan data dilakukan untuk memastikan setiap baris memiliki peluang yang sama dalam mewakili seluruh atribut. Dengan mengacak urutan data, model Naïve Bayes dapat belajar dari data secara merata dan tidak terpengaruh oleh susunan awal dalam dataset. Proses ini dilakukan sebelum data dibagi menjadi data latih dan data uji [15].

e. Split Data

Dataset yang digunakan terdiri dari 66 atribut prediktor yang mewakili dimensi dalam data IDM. Untuk menyederhanakan proses klasifikasi dan meningkatkan akurasi model, dilakukan seleksi fitur dengan menghapus beberapa kelompok atribut yang dianggap kurang berpengaruh terhadap tingkat partisipasi pemilu [16]. Setelah diseleksi fitur, tersisa 40 atribut pada dataset tersebut. Lalu aplikasikan penyaringan fitur tersebut di training set saja untuk mencegah data leakage [17].

3. Implementasi Algoritma

Tahap ini merupakan implementasi algoritma klasifikasi pada data IDM Jawa Timur 2024. Algoritma yang digunakan adalah Multiple Layer Perceptron (MLP) Backpropagation dan Gaussian Naïve Bayes.



Pemilihan

Gaussian Naive Bayes didasarkan pada kemampuannya menangani data kontinu [7], sedangkan MLP Backpropagation dipilih karena mampu memodelkan hubungan non-linear yang kompleks [6]. Berikut ditunjukkan implementasi algoritma MLP Backpropagation Classifier:

A. Langkah-Langkah Algoritma MLP Backpropagation

1. Normalisas

i Data

Normalisasi nilai pada seluruh isi dataset dari rentang 0 hingga 1 sebagai syarat untuk melanjutkan pada proses fungsi aktivasi sigmoid. Dalam penelitian ini, penulis menggunakan adjusted normalization untuk mencegah nilai input berada dalam rentang yang konsisten [18].

2. Arsitektur Jaringan

Dalam menentukan arsitektur jaringan dan banyaknya neuron pada hidden layer, penulis di sini menggunakan aturan praktis, dengan rumus [19].

Dalam aplikasi rumus di atas menghasilkan 35 neuron pada lapisan tersembunyi, sehingga arsitektur jaringannya adalah 65-35-5.

3. Menentukan Bobot dan Bias

$$x' = \\frac{x_i - x_{min}}{x_{max} - x_{min}}$$

$$Nj = Ni + Nk$$

secara acak dari rentang -0,5–0,5 [20]. Namun pemilihan random memiliki peraturan dimana inisialisasi bobot dan bias tidak boleh memiliki nilai yang sama persis dengan input maupun output [21].

4. Menentukan Learning Rate

Learning rate yang digunakan untuk penelitian ini adalah, yakni 0,001. Untuk mencapai hasil prediksi yang akurat dan presisi pada dataset berskala besar, learning rate 0,001 cenderung memberikan performa yang baik karena memungkinkan proses pembelajaran yang lebih stabil dan terkontrol [22]. Selain itu, perubahan bobot dan bias yang dihasilkan tidak terlalu besar, sehingga model menjadi lebih stabil dalam melakukan prediksi pada dataset berskala kecil.

5. Menentukan Fungsi Aktivasi dan Fungsi Loss

Fungsi aktivasi yang digunakan pada penelitian ini yaitu fungsi aktivasi softmax. Karena softmax mengubah output JST menjadi distribusi probabilitas antarkelas. Kombinasi softmax dan Cross Entropy (CE) sebagai pendekatan klasifikasi secara teoritis paling efektif dalam klasifikasi [23]. Hal ini dikarenakan CE bekerja dengan memodelkan probabilitas kelas, yang memiliki prinsip yang sama dengan klasifikasi [24].

6. Propagasi Maju

Propagasi maju pada MLP backpropagation classifier adalah proses menghitung keluaran jaringan dengan cara mengalikan input dengan bobot, menjumlahkan bias, lalu melewatkannya ke fungsi aktivasi di setiap neuron [25].

Output dari Z_{netj} dihitung menggunakan sigmoid [26]:

Menghitung aktivasi di output jaringan menggunakan softmax [23]:

Menghitung fungsi loss dengan Cross Entropy (CE) [27]:

7. Fase Propagasi Mundur (Backward)

Fase Backward adalah proses menghitung error dengan cara mengurangkan output prediksi dengan target sebenarnya, kemudian menyebarluaskan error tersebut ke layer sebelumnya menggunakan aturan rantai turunan untuk memperoleh gradien [21].

8. Update Bobot dan Bias

Pada MLP backpropagation classifier, bobot dan bias diperbarui dengan cara menguranginya menggunakan gradien error yang telah dihitung dikalikan dengan learning rate. Proses ini memastikan bobot dan bias bergerak sedikit demi sedikit ke arah yang meminimalkan nilai loss function [28].

$$Z_{netj} = \sum_{i=1}^n x_i \cdot v_{ij} + b_j$$

$$Z_j = \frac{1}{1 + e^{-Z_{netj}}}$$

$$y_k = \exp(y_{netj})$$

$$y_k = \frac{\exp(y_{netj})}{\sum_{i=1}^c \exp(y_{neti})}$$

$$j=1$$

CE = -
1

n
 $ti \log(yi)$
 c
 $j=1$
 n
 $i=1$)

Page | 7

B. Langkah-Langkah Algoritma Naïve Bayes

1. Presentasi Data

Informasi dari IDM diubah menjadi serangkaian fitur numerik yang merefleksikan kondisi sosial, ekonomi, lingkungan, aksesibilitas, dan tata kelola pemerintahan desa secara terstruktur dalam bentuk vektor data [12]. Setiap entri data mewakili satu desa dengan atribut-atribut seperti layanan dasar, pendidikan, penanggulangan bencana, dan sebagainya. Label kelas yang digunakan adalah status, yang diklasifikasikan ke dalam kelas:

2. Perhitungan Probabilitas Posterior (Teorema Bayes)

Algoritma Naïve Bayes menghitung probabilitas kemunculan suatu kelas Y terhadap sekumpulan fitur X dengan menggunakan Teorema Bayes, yang dirumuskan sebagai berikut:

3. Prediksi Kelas

Dalam proses klasifikasi, algoritma Naïve Bayes akan menghitung peluang atau probabilitas setiap kelas berdasarkan nilai-nilai fitur yang dimiliki oleh suatu data. Kemudian, model akan memilih kelas yang memiliki probabilitas paling besar sebagai hasil prediksi. Artinya, kelas dengan kemungkinan tertinggi dianggap sebagai representasi paling tepat dari data tersebut.

Dengan asumsi independensi antar fitur. Asumsi ini menyederhanakan proses perhitungan, sehingga algoritma dapat bekerja secara efisien meskipun dengan jumlah fitur yang besar. Naïve Bayes menghitung:

4. Validasi Model

Untuk menguji kinerja algoritma Naïve Bayes, dilakukan validasi menggunakan data uji. Beberapa matrik evaluasi yang digunakan [29]:

1. Accuracy

Mengukur persentase prediksi yang benar terhadap keseluruhan data:

2. Precision

Menunjukkan seberapa akurat prediksi terhadap kelas positif [30]:

3. Recall

Mengukur kemampuan model dalam menemukan seluruh data yang benar-benar positif:

4. F-1 Score

Merupakan rata-rata harmonik dari Precision dan Recall, yang memberikan keseimbangan antara keduanya:



$Y = \arg \max P(Y|X)$

$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$

$Precision + Recall$

$y \in \{mandiri, maju, membangun, terbelakang, sangat terbelakang\}$

$P(Y|X) = \frac{P(X|Y) \cdot P(Y)}{P(X)}$

$P(X)$

$P(X|Y) = P(x_1|Y) \cdot P(x_2|Y) \cdots P(x_n|Y)$

$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$

$Precision = \frac{TP}{TP + FP}$

$Recall = \frac{TP}{TP + FN}$

$x = x_1, x_2, \dots, x_n$

8 | Page

III.

HASIL DAN PEMBAHASAN

A. Distribusi Data Status Desa IDM Jawa Timur 2024

Statistik deskriptif merupakan cabang dari statistika yang berfungsi untuk menyajikan data agar lebih mudah dipahami [31]. Pada bagian ini, penulis menampilkan sebaran data berdasarkan kategori status desa menurut IDM melalui bentuk visual (diagram). Sebaran tersebut menggambarkan distribusi status desa di Provinsi Jawa Timur serta menampilkan peringkat 10 kabupaten dengan persentase desa mandiri dan maju tertinggi. Gambar berikut menyajikan diagram lingkaran mengenai sebaran lima kategori status desa berdasarkan IDM Jawa Timur tahun 2024:

Gambar 2. Sebaran status desa IDM 2024 di Provinsi Jawa Timur

Berdasarkan Gambar 1, status desa dengan jumlah tertinggi di Jawa Timur adalah desa berstatus maju, yaitu sebanyak 2.887 desa (37,4%). Selanjutnya, desa berkembang menempati urutan kedua dengan total 1.966 desa (25,5%), diikuti desa mandiri sebanyak 1.736 desa (22,5%). Sementara itu, terdapat 846 desa (11,0%) berstatus tertinggal, dan yang paling sedikit adalah desa sangat tertinggal dengan jumlah 286 desa (3,7%). Selanjutnya, dua gambar di bawah ini akan menyajikan diagram batang yang menampilkan 10 kabupaten/kota dengan persentase tertinggi desa berstatus mandiri.

Gambar 3. Persentase 10 Kabupaten/Kota dengan Desa Berstatus Mandiri Tertinggi

Page | 9

Dari gambar di atas menunjukkan Kota Batu meraih peringkat tertinggi desa terindeks mandiri, dimana terdapat 18 desa berstatus mandiri dengan total keseluruhan 19 desa (94,7%). Di urutan ke-dua adalah Kabupaten Banyuwangi dengan 80 desa berstatus mandiri dari 189 desa. Dan di peringkat ke-tiga adalah Kabupaten Mojokerto dengan 100 desa berstatus mandiri dari 299 desa (33,4%). Urutan selanjutnya Kabupaten Lamongan (33.1%), Kabupaten Malang (31%), Kabupaten Sidoarjo (28.9%), Kabupaten Blitar (28,6%), Kabupaten Lumajang (28.3%), Kabupaten Nganjuk (27.7%), dan terakhir Kabupaten Trenggalek (27.6%). %). Selanjutnya, dua gambar di bawah ini akan menyajikan diagram batang yang menampilkan 10 kabupaten/kota dengan persentase tertinggi desa berstatus maju..

Gambar 3. Persentase 10 Kabupaten/Kota dengan Desa Berstatus Maju

Dari gambar di atas menunjukkan Kabupaten Magetan meraih peringkat tertinggi desa terindeks maju, dimana terdapat 111 desa berstatus maju dari total keseluruhan 207 desa (53,6%). Di urutan ke-dua adalah Kabupaten Situbondo dengan 66 desa berstatus mju dari 132 desa (50%). Dan di peringkat ke-tiga adalah Kabupaten Gresik dengan 159 desa berstatus maju dari 330 desa (48.2%). Urutan selanjutnya Kabupaten Malang (45%), Kabupaten Bojonegoro (44.6%), Kabupaten Madiun (43.9%), Kabupaten Ngawi (43.2%), Kabupaten Blitar (42.3%), Kabupaten Pasuruan (41.94%), dan terakhir Kabupaten Pacitan (41.92%).

B. Hasil Uji Validasi Model

Akurasi dan Presisi MLP Backpropagation

Pada Tabel 3, hasil uji akurasi pada test set menunjukkan bahwa performa model bervariasi antar kategori. Kategori mandiri memiliki performa terbaik dengan precision sebesar 0.84, recall 0.80, dan F1-score 0.82 pada 521 data. Kategori maju juga relatif baik dengan precision 0.72, recall 0.89, dan F1-score 0.80. Selanjutnya, kategori berkembang mencapai precision 0.77, recall 0.83, dan F1-score 0.80 pada 866

data, serta kategori tertinggal memperoleh precision 0.82, recall 0.72, dan F1-score 0.77 pada 586 data.



Namun, performa model menurun drastis pada kategori sangat tertinggal yang hanya mencapai precision

0.20, recall 0.03, dan F1-score 0.05 pada 254 data.

Secara keseluruhan, akurasi model pada test set MLP Backpropagation adalah 0.7540, sedikit menurun

dibandingkan akurasi pada training set sebesar 0.7802. Penurunan ini mengindikasikan adanya kesulitan

model dalam mempelajari pola pada data uji, yang kemungkinan disebabkan oleh keterbatasan jumlah data,

hanya 30% dari total dataset (2317 data) serta kompleksitas fitur input yang relatif tinggi. Kondisi ini

menunjukkan bahwa peningkatan jumlah data pelatihan dapat berkontribusi terhadap peningkatan performa

model.



10 | Page

Tabel 3. MLP Backpropagation Test Set Classification Report.

Precision Recall F1-Score Support

Sangat Tertinggal 0.20 0.03 0.05 254

Tertinggal 0.82 0.72 0.77 86

Berkembang 0.77 0.83 0.80 590

Maju 0.72 0.89 0.80 866

Mandiri 0.84 0.80 0.82 521

Accuracy 0.75 2317

Macro Avg 0.75 0.65 0.65 2317

Weighted Avg 0.77 0.75 0.72 2317

Akurasi dan Presisi Naïve Bayes

Pada Tabel 4, hasil uji akurasi pada test set menggunakan Naïve Bayes menunjukkan bahwa performa

model relatif lebih stabil dibandingkan Backpropagation Classifier. Kategori mandiri memperoleh precision

0,90, recall 0,9, dan F1-score 0,78 pada 520 data. Kategori maju menunjukkan hasil cukup baik dengan

precision 0,



76, recall 0,90, dan F1-score 0,83. Selanjutnya, kategor

i berkembang mencapai precision 0,85,

recall 0,87, dan F1-score 0,86, sedangkan kategori tertinggal memperoleh precision 0,74, recall 0,94, dan F1-score 0,83 pada 597 data. Performa menurun pada kategori sangat tertinggal, yaitu precision 0,82, recall 0,53, dan F1-score 0,64 pada 254 data. Meskipun demikian, penurunan ini tidak sedastis yang terjadi pada Backpropagation Classifier.

Secara keseluruhan, akurasi model pada test set Naïve Bayes adalah 0,81, sedikit lebih tinggi dibandingkan akurasi pada training set sebesar 0,80. Hal ini menunjukkan bahwa Naïve Bayes mampu bekerja cukup baik pada jumlah data uji lebih sedikit, karena sifat algoritma ini yang lebih sesuai untuk dataset berukuran kecil dengan asumsi distribusi fitur yang sederhana dan relatif linear. Peningkatan akurasi sebesar 0,01 antara training dan testing set mengindikasikan bahwa model tidak mengalami overfitting yang berarti, serta menunjukkan generalisasi yang lebih baik dibandingkan Backpropagation Classifier.

Tabel 4. Naive Bayes Test Set Classification report

Precision Recall F1-Score Support

Sangat Tertinggal 0.82 0.53 0.64 254

Tertinggal 0.74 0.94 0.83 86

Berkembang 0.85 0.87 0.86 590

Maju 0.76 0.91 0.83 866

Mandiri 0.90 0.69 0.78 521

Accuracy 0.81 2317

Macro Avg 0.81 0.79 0.79 2317

Weighted Avg 0.82 0.81 0.81 2317

Tabel 5. Perbandingan Akurasi Model

Model Akurasi Train Set Akurasi Test Set

Backpropagation 0.7802 0.7540

Naïve Bayes 0.80 0.81

Confusion Matrix

Pada Gambar 5 terlihat bahwa kategori mandiri dan maju memiliki tingkat presisi yang relatif baik dalam mengklasifikasikan data. Namun, pada kategori lainnya masih terdapat banyak kesalahan, terutama pada kategori sangat tertinggal. Dari total 254 data sangat tertinggal, hanya 52 data yang berhasil diklasifikasikan dengan benar, sedangkan sisanya salah diprediksi sebagai tertinggal (8 data), berkembang (52 data), maju (102 data), dan mandiri (40 data). Hal ini menunjukkan bahwa model masih kesulitan membedakan kategori sangat tertinggal dengan kategori maju dan berkembang.



Untuk kategori tertinggal, sebanyak 46 data berhasil diklasifikasikan dengan benar dari total 86 data,

sementara 18 data salah diprediksi sebagai sangat tertinggal, 21 data sebagai berkembang, dan 1 data sebagai

mandiri. Kondisi ini mengindikasikan bahwa model masih mengalami kebingungan dalam membedakan

kategori tertinggal dengan kategori berkembang maupun sangat tertinggal.

Page | 11

Sementara itu, pada kategori berkembang, terdapat 384 data yang terklasifikasi dengan benar dari total 590 data. Kesalahan klasifikasi terjadi pada 102 data yang diprediksi sebagai sangat tertinggal, 16 data sebagai tertinggal, 84 data sebagai maju, dan 4 data sebagai mandiri. Hal ini menunjukkan bahwa model masih sering menyamakan kategori berkembang dengan kategori sangat tertinggal dan maju.



Gambar 5. Confusion Matrix MLP Backpropagation

Pada Gambar 6 terlihat bahwa kategori maju dan tertinggal memiliki tingkat klasifikasi yang relatif stabil. Namun, pada kategori lainnya masih terdapat cukup banyak kesalahan, terutama pada kategori sangat tertinggal. Dari total 254 data sangat tertinggal, hanya 134 data yang berhasil diklasifikasikan dengan benar, sedangkan sisanya salah diprediksi sebagai tertinggal (3 data), berkembang (21 data), maju (61 data), dan mandiri (35 data). Kondisi ini menunjukkan bahwa model masih mengalami kesulitan dalam membedakan kategori sangat tertinggal dengan maju dan mandiri.

Untuk kategori berkembang, sebanyak 511 data berhasil diklasifikasikan dengan benar dari total 590 data. Kesalahan klasifikasi terjadi pada 4 data yang diprediksi sebagai sangat tertinggal, 26 data sebagai tertinggal, dan 49 data sebagai maju. Kondisi ini mengindikasikan bahwa model masih mengalami kebingungan dalam membedakan kategori berkembang dengan maju maupun tertinggal. Secara keseluruhan, kesalahan klasifikasi pada model Naïve Bayes relatif lebih stabil apabila dibandingkan dengan Backpropagation Classifier.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

IV. SIMPULAN

Hasil perbandingan performa antara Backpropagation dan Naïve Bayes menunjukkan bahwa Naïve Bayes memiliki kinerja lebih unggul. Pada training set, akurasi Naïve Bayes sebesar 0,80, bahkan sedikit meningkat pada test set menjadi 0,81. Hal ini mengindikasikan bahwa model tidak mengalami overfitting dan mampu melakukan generalisasi dengan baik. Sebaliknya, Backpropagation menunjukkan akurasi 0,7802 pada training set, namun turun menjadi 0,7540 pada test set. Pada uji validasi, Backpropagation mengalami penurunan akurasi secara gradual, sedangkan pada Naïve Bayes penurunan lebih terlihat pada kategori sangat tertinggal. Hal ini terjadi karena model masih kesulitan membedakan kategori tersebut dengan kategori lain, meskipun penurunan performanya tidak sedastis Backpropagation. Secara keseluruhan, Naïve Bayes tetap menunjukkan performa yang lebih stabil. Temuan ini menegaskan bahwa model baru tidak selalu lebih baik daripada model yang lebih sederhana. Faktor pra-pemrosesan dan pelatihan data juga berpengaruh besar terhadap akurasi dan presisi klasifikasi. Naïve Bayes dapat bekerja efektif pada data dengan jumlah fitur linear yang cukup banyak, sebagaimana pada data IDM Jatim 2024 yang relatif tidak terlalu besar jika dibandingkan dengan jumlah fiturnya. Sebaliknya, Backpropagation lebih sesuai untuk data berskala besar dengan pola non-linear yang kompleks.



Dengan demikian, pemilihan model perlu disesuaikan dengan karakteristik data. Untuk perhitungan IDM dalam skala nasional, Backpropagation dapat dipertimbangkan mengingat kemampuannya dalam mengolah data besar dan kompleks. Namun, untuk skala regional dengan jumlah data yang relatif lebih kecil dan fitur yang cenderung linear, metode tradisional seperti Naïve Bayes, justru lebih tepat digunakan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada keluarga atas segala dukungan, baik secara emosional maupun material, dalam menjalani proses akademik dari awal hingga saat ini. Ucapan terima kasih juga disampaikan kepada teman-teman, rekan sejawat, serta akademisi Informatika Umsida yang telah memberikan bimbingan dan dukungan sehingga penelitian ini dapat terselesaikan dengan baik.

REFERENSI