Perbandingan Algoritma *Naive Bayes*, *Logistic Regression* Dan *Artificial Neural Network* Untuk Klasifikasi Tingkat Kesejahteraan Keluarga

Natalia R. Simon 1[1], Alter Lasarudin 2[2], Wahyudin Hasyim 3[3]*
[1,2,3,]Program Studi Sistem Informasi, Universitas Muhammadiyah Gorontalo
nataliasimon769@gmail.com*
Universitas Muhammadiyah Gorontalo, Jl. Prof. Mansoer Pateda, Kab. Gorontalo, Indonesia

ABSTRACT

Batu Hijau Village is one of the coastal villages in Bonepantai District, Bone Bolango Regency. Currently, the classification of family welfare levels in the area is not entirely accurate, which leads to the inappropriate targeting of subsidy distribution. Therefore, it is crucial to understand the village's family welfare level. The author used the Naive Bayes, Logistic Regression, and Artificial Neural Network algorithms in this study. Based on the results of experiments using the RapidMiner and Google Colab tools, the Naive Bayes, Logistic Regression, and Artificial Neural Network algorithms had the same accuracy value of 100%, precision of 100%, and recall of 100%. Based on the results of testing using the RapidMiner and Google Colab tools, the Naive Bayes, Logistic Regression, and Artificial Neural Network algorithms could classify family welfare level data perfectly, indicating that the model could classify data without errors.

Keywords: Family Welfare Level Classification; Naive Bayes; Logistic Regression; Artificial Neural Network.

ABSTRAK

Desa Batu Hijau adalah salah satu desa di pesisir pantai yang ada di Kecamatan Bonepantai Kabupaten Bone Bolango. Saat ini klasifikasi tingkat kesejahteraan keluarga di wilayah tersebut belum sepenuhnya tepat, yang mengakibatkan distribusi subsidi yang tidak tepat sasaran. Oleh karena itu, sangat penting untuk memahami tingkat kesejahteraan keluarga di Desa tersebut. Dalam penelitian ini, penulis menggunakan algoritma *Naive Bayes, Logistic Regression,* dan *Artificial Neural Network.* Berdasarkan hasil percobaan menggunakan tool rapidminer dan google colab algoritma *naive bayes, logistic regression,* dan *artificial neural networ k* memiliki nilai akurasi yang sama yaitu 100%, precision 100%, serta recall 100%. Berdasarkan Hasil pengujian menggunakan tool rapidminer dan google colab algoritma *Naive Bayes, Logistic Regression,* dan *Artificial Neural Network* mampu mengklasifikasikan data tingkat kesejahteraan keluarga dengan sempurna yang menunjukkan bahwa model mampu mengklasifikasikan data tanpa kesalahan.

Kata Kunci: Klasifikasi Tingkat Kesejahteraan Keluarga, Naive Bayes, Logistic Regression, Artificial Neural Network**.**

1. PENDAHULUAN

Kesejahteraan adalah tujuan dari setiap keluarga. Kesejahteraan adalah kemampuan setiap keluarga atau kelompok untuk memenuhi semua kebutuhan agar dapat hidup layak, produktif, dan sehat. Menurut data dari Badan Pusat Statistik pada September 2020, presentase penduduk miskin sebesar 10,19%. Ini meningkat 0,41% dari maret 2020 dan naik lagi 0,97% dari September 2019.

Keluarga merupakan unit paling kecil yang terapat dalam struktur sosial bermasyarakat yang di dalamnya terdiri dari kepala keluarga dan beberapa
orang anggota keluarga lainnya yang berkumpul dan tinggal disuatu tempat dalam satu atap yang saling terikat dan saling berketergantungan antar satu sama lain. Hubungan dalam lingkungan sosial keluarga lebih dekat dibandingkan dengan lingkungan masyarakat sekitar.

Kesejahteraan keluarga merupakan suatu kondisi dimana seluruh kebutuhan dasar setiap anggota keluarga terpenuhi baik itu secara material, sosial, dan spritual, sehingga mereka dapat hidup layak dan sehat. Keluarga sejahtera juga mencakup hubungan yang harmonis antar anggota keluarga, masyarakat serta lingkungan.Kualitas keluarga sangat dipengaruhi oleh tingkat kesejahteraannya, dimana peran anggota keluarga sangat penting untuk mencapai tujuan kesejahteraan keluarga itu sendiri. Namun, jumlah dan kualitas masalah kesejahteraan terus meningkat. Keterbatasan ekonomi sosial menyebabkan banyak orang tidak dapat memenuhi kebutuhan dasar mereka.

Di daerah pesisir pantai, khususnya di Desa Batu Hijau, kehidupan ekonomi mereka bergantung secara langsung pada pemanfaatan sumber daya laut dan pesisir. Sebagian besar masyarakat yang ada di daerah pesisir pantai tergolong miskin karena mereka tidak memiliki sumber modal, tidak memiliki akses ke pasar, dan tidak terlibat dalam pengelolaan sumber daya alam.Oleh karena itu, sangat penting untuk memahami tingkat kesejahteraan keluarga di Desa tersebut. Saat ini klasifikasi tingkat kesejahteraan keluarga di wilayah tersebut belum sepenuhnya tepat yang mengakibatkan distribusi subsidi yang tidak tepat sasaran.

Desa Batu Hijau berada di Kecamatan Bonepantai, Kabupaten Bone Bolango. Lokasi Desa Batu Hijau adalah pada 0°21'31.94" Lintang Utara dan 123°13'40.43" Bujur Timur. Terletak di atas garis khatulistiwa atau garis equator, sejauh 3,8 km2 dari Pusat Kecamatan Bonepantai, 47,9 km2 dari Pusat Pemerintah Daerah Kabupaten Bone Bolango, dan 38,4 km2 dari Pusat Pemerintah Propinsi Gorontalo. Wilayah administratif Desa Batu Hijau seluas ± 6 km2, dengan luas lautan 1,47 km2 dan panjang pantai 1,8 km2. Desa Batu Hijau berbatasan dengan Desa Kemiri di sebelah Utara, Desa Tongo di sebelah Barat, dan Desa Uabanga di sebelah Timur.

Dalam penelitian ini, penulis menggunakan beberapa algoritma perbandingan yaitu Naive Bayes, Logistic Regression, dan Artificial Neural Network. Algoritma ini dipilih karena merupakan algoritma klasifikasi yang memiliki performa dan akurasi yang baik.

Dengan fokus pada pengestimasian probabilitas, algoritma Naive Bayes adalah metode pengklasifikasian yang didasarkan pada probabilitas sederhana dan mengansumsikan antar variabel penjelas saling bebas (independen). Keuntungan algoritma Naive Bayes termasuk kecepatan dan akurasi yang lebih tinggi ketika digunakan pada dataset yang lebih besar, dan tingkat nilai error yang lebih rendah ketika dataset yang lebih besar digunakan.

Algoritma Logistic Regression adalah teknik statistik yang digunakan untuk memprediksi kemungkinan suatu instance termasuk ke dalam salah satu kelas target biner berdasarkan nilai nilai fiturnya.

Model penalaran yang didasarkan pada otak manusia disebut Algoritma Jaringan Syaraf Artificial. ANN terdiri dari banyak prosesor saling berhubungan yang sangat sederhana yang disebut neuron. Sinyal dikirim dari neuron satu ke neuron lain melalui pembobotan atau berat.

Dengan demikian tingkat kesejahteraan keluarga penting untuk diklasifikasi dengan tujuan penentuan target masyarakat miskin agar

subsidi pemerintah dapat disalurkan secara tepat sasaran.

2. TINJAUAN PUSTAKA

a. Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. Data mining merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Data mining disebut juga dengan pattern recognition yang merupakan metode dalam pengolahan data untuk menemukan pola tersembunyi dari data yang diolah kemudian menghasilkan suatu pengetahuan baru yang bersumber dari data lama dan dapat digunakan dalam menentukan keputusan dimasa yang akan datang.(Romli, 221)

b. Klasifikasi

Klasifikasi adalah metode yang mengklasifikasikan objek berdasarkan model yang dimiliki objek klasifikasi. Metode ini membuat model pada data pelatihan sebelumnya dan kemudian menggunakan model tersebut untuk mengklasifikasikan data baru. Klasifikasi mempelajari suatu fungsi objektif yang mengaitkan setiap himpunan atribut (karakteristik) dengan sejumlah abel kelas yang ada. Beberapa algoritma yang termasuk dalam klasifikasi adalah C4.5, Kmeans, Priori, Support Vector Machines, dan Naive Bayes. (Homepage et al., 2022)

Klasifikasi adalah proses menemukan model atau fungsi yang menggambarkan dan membedakan kelas data, dan dimaksudkan untuk memungkinkan anda menggunakan model untuk memprediksi kelas objek dari label kelas yang tidak diketahui.l
(Mandar, 2022)

c. Algoritma

Dalam ilmu matematika dan komputer, pengertian algoritma merupakan prosedur dari beberapa langkah demi langkah untuk penghitungan. Algoritma dipakai untuk penghitungan, penalaran otomatis, dan pemrosesan data. Pengertian algoritma ialah suatu metode yang efektif diekspresikan sebagai rangkaian yang terbatas dari beberapa instruksi yang telah dijelaskan dengan baik guna menghitung sebuah fungsi. Susunan algoritma dimulai dari kondisi awal dan input awal, instruksi tersebut mendeskripsikan komputasi yang apabila itu dieksekusi serta diproses dengan melewati urutan-urutan kondisi terbatas yang terdefinisi dengan baik, sehingga dapat menghasilkan output atau keluaran dan berhenti di kondisi akhir yang telah ditentukan.(Rizq & Jodi, 2020)

d. Naive Bayes

Dalam pengklasifikasian, metode Naive Bayes menggunakan konsep probabilitas sederhana untuk menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Keuntungan dari metode ini adalah bahwa itu hanya membutuhkan jumlah data pelatihan atau pelatihan yang kecil untuk menemukan perkiraan parameter yang diperlukan untuk proses pengklasifikasian.. (Tri et al., 2021)
Karena metode Naive Bayes tidak memiliki aturan, dia menggunakan cabang matematik, konsep probabilitas, untuk menemukan kemungkinan tertinggi dari proses klasifikasi yang dilakukan. Dia melakukan pengamatan terhadap bagaimana setiap klasifikasi muncul dalam data pelatihan.(Julianto & Andayani, 2024)

e. Logistic Regression

Logistic regression memodelkan hubungan antara berbagai faktor independen (seperti memori perangkat, ukuran perangkat, dan durasi baterai) dan variabel hasil (dari kelompok harga rendah, sedang, tinggi, atau sangat tinggi) dalam data nominal atau ordinal. Fungsi logistik menunjukkan kemungkinan bahwa sebuah ponsel akan termasuk dalam kategori harga tertentu. Tiga kategori logistic regression

berbeda: binar, multinomial, dan ordinal. (Najwa Ardelia et al., 2020)

Dalam ilmu biologi dan sosial, regresi logistik banyak digunakan. Ini adalah jenis regresi yang menghubungkan satu atau beberapa variabel independen, juga disebut variabel bebas, dengan variabel dependen yang berupa kategori, biasanya antara 0 dan 1. Metode ini digunakan ketika variabel yang ada berkategori 0 atau 1. (Achmad, 2022)

f. Artificial Neural Network

ANN biasa digunakan untuk memproses kumpulan data yang besar, untuk dapat menyediakan informasi analisis yang dapat berguna untuk memprediksi kanat maupun identifikasi klasifikasi suatu data. Fungsi aktivasi merupakan fungsi yang digunakan pada jaringan saraf untuk mengaktifkan atau tidak mengaktifkan neuron. Karakteristik yang harus dimiliki oleh fungsi aktivasi jaringan perambatan balik antara lain harus kontinu, terdiferensial, dan tidak menurun secara monoton (monotonically non-decreasing). Fungsi aktivasi dibutuhkan oleh hidden layer pada ANN untuk membuat ANN menjadi non linear. Fungsi aktivasi dapat berupa fungsi linear, threshold, atau sigmoid. Fungsi sigmoid biasa digunakan karena fungsi ini mengombinasikan nearly linear behavior, curvi linear behavior, dan nearly constant behavior. (Sutoyo et al., 2020)

Kesejahteraan Keluarga

Keluarga Sejahtera menurut UU No 10 tahun 1992 adalah keluarga yang dibentuk atas perkawinan yang sah, mampu memenuhi kebutuhan hidup spiritual dan materiil yang layak, bertaqwa pada Tuhan YME, memiliki hubungan yang serasi, selaras, dan seimbang antar anggota dan antara keluarga dengan masyarakat dan lingkungan.(Bagus Bambang Sumantri et al., 2020)

g. Rapid Miner

RapidMiner merupakan suatu perangkat lunak untuk sains data yang dibentuk oleh rapidminer, yang didalamnya menyajikan data yang terintegrasi text mining, machine learning, dan analisis prediktif. Serta membantu langkah dalam suatu proses pembelajaran mesin seperti hasi visualisasi, optimasi, persiapan data dan validasi model. Dalam pengertian lain rapidminer merupakan sebuah aplikasi yang digunakan dalam mengolah sebuah data dengan berbagai teknik dan metode dalam sebuah data mining, sehingga data dapat menjadi informasi yang berguna. Rapidminer adalah salah satu perangkat lunak yang bersifat terbuka yang berguna untuk pengolahan suatu data mining.(Waldi Setiawan dkk., 2024).

h. Google Collab

Google Collab (Anjani, 2024) merupakan sebuah platform berbasis cloud yang memungkinkan pengguna untuk menulis dan mengeksekusi kode Python secara daring melalui browser web tanpa perlu menginstal perangkat lunak tambahan(Tock, 2020). Platform ini memiliki keunggulan dalam kemudahan aksesibilitas, kolaborasi real-time, dan integrasi dengan layanan Google lainnya, seperti Google Drive dan Google Docs. Sebagai hasilnya, Google Colaboratory menjadi salah satu pilihan utama bagi pendidik dan siswa dalam menyajikan materi dan melakukan praktikum dalam bidang ilmu komputer.(Soen et al., 2022).

i. Kriteria

Kriteria yang digunakan dalam penelitian ini yaitu: Pendidikan kk, Pekerjaan Kk, Kepemilikan rumah, Bantuan, Bahan bakar dan Sumber Penerangan

j. Penelitian terkait

Penelitian yang dilakukan oleh(Nasien, Dkk 2024) Perbandingan Implementasi machine learning menggunakan mètode knn, naive bayes, dan regresi logistic untuk mengklasifikasi penyakit diabetes. Penyakit diabetes menjadi sorotan karena sifatnya yang kronis, dengan gejala utama berupa peningkatan kadar gula darah di atas batas normal. Diabetes terjadi ketika tubuh tidak dapat efisien mengambil glukosa ke dalam sel untuk diubah menjadi energi, menyebabkan menumpuknya gula ekstra dalam aliran darah. Penelitian ini menggunakan ekstraksi fitur dengan Analisis Komponen

Utama (Principal Component Analysis-PCA) dengan ambang batas 80%, menghasilkan 5 fitur utama. Fitur-fitur ini kemudian digunakan sebagai input untuk tiga classifier, yaitu K-Nearest Neighbors (KNN), Naive Bayes, dan Regresi Logistik. Data yang digunakan berasal dari Kaggle, dengan pembagian data 70:30 dan 80:20. Hasil penelitian menunjukkan bahwa metode Naive Bayes memberikan akurasi terbaik, mencapai 79% pada pembagian data 80:20. Oleh karena itu, dapat disimpulkan bahwa algoritma Naive Bayes adalah pilihan terbaik untuk klasifikasi data diabetes dalam penelitian ini.

3. METODE PENELITIAN

a. Objek Penelitian

Objek penelitian ini berfokus pada klasifikasi tingkat kesejahteraan keluarga dengan menggunakan Perbandingan Algoritma Naive Bayes, Logistic Regression dan Artificial Neural Network.

b. Pengumpulan data

Pengumpulan data penduduk untuk tingkat kesejahteraan keluarga dilakukan di Desa Batu Hijau. Data yang digunakan yaitu sebanyak 156 data kepala keluarga (KK). Dari hasil pengumpulan data tersebut akan ditemukan berapa jumlah keluarga yang termasuk kategori pra sejahtera untuk dijadikan sebagai acuan dalam pemberian bantuan sehingga tepat sasaran. Data tersebut akan diuji menggunakan Perbandingan Algoritma Naive Bayes, Logistic Regression dan Artificial Neural Network Untuk Klasifikasi Tingkat Kesejahteraan Keluarga.

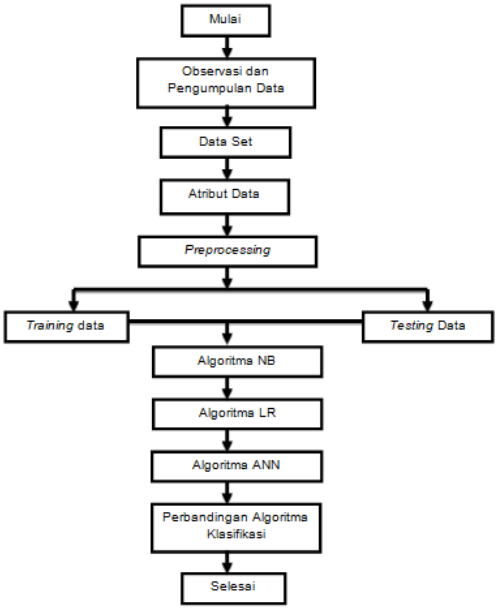
c. Observasi

Untuk mengumpulkan informasi dan pengambilan data, maka peneliti turun langsung untuk mengamati kondisi yang ada di desa tersebut.

d. Wawancara

Dalam penelitian ini, wawancara dilakukan dengan aparat Desa Batu Hijau untuk mendapatkan informasi yang diperlukan dalam penelitian.

e. Tahapan Penelitian



f. Tahapan Perhitungan Algoritma Naive Bayes

Tahapan naive bayes adalah sebagai berikut :
Menghitung jumlah kelas atau label
Menghitung jumlah kasus per kelas
Kalikan semua variabel
Bandingkan hasil perkelas
Rumus algoritma naive bayes adalah sebagai berikut:
P (c|x) = ...**(3.1)**
Keterangan :
x : data dengan class yang belum diketahui
c : hipotesis data merupakan suatu class spesifik
p (c|x) : Probabilitas hipotesis berdasar kondisi
p(c) : Probabilitas hipotesis (prior probability)
p(x|c) : Probabilitas berdasarkan kondisi pada hipotesis
p(x) : Probabilitas c

g. Tahapan Perhitungan Algoritma Logistic Regression

Tahapan Logistic Regression adalah sebagai berikut:
Tentukan koefisien regresi b0,b1,b2

Hitung nilai logit dengan mengambil logaritma peluang probabilitas (p) terjadinya suatu peristiwa menggunakan rumus

Logit (p) =.log($\frac{p}{1-p}$).....**(3.2)**

Buat nilai e logit yang dipangkatkan dengan nilai di kolom logit

e logit = logit ^2 .....**(3.3)**

hitung nilai probabilitas

P =$\frac{e^{a+bx}}{1+e^{a+bx}}$ .....**(3.4)**

Keterangan :

p adalah probabilitas nilai 1 (proposi angka 1, rata rata Y)

e adalah konstanta

a dan b adalah parameter algoritma

Menghitung nilai kemungkinan log likelihood

Log kemungkinan = Ln (Probabilitas).....**(3.5)**

Jumlahkan semua nilai likelihood

Likelihood = Sum (nilai likelihood)

menghitung estimasi koefisien regresi secara otomatis.

Dalam excel perhitungan koefisien regresi adalah sebagai berikut:

Instal add-in Solver Excel dengan mengeklik terlebih dahulu menu Beranda, lalu menu Add-In.

Cari dan instal Solver dengan mengikuti petunjuknya.

Pilih menu Data dari navigasi tingkat atas dan klik Solver di sisi kanan untuk menjalankan add-in.

Di panel Parameter Pemecah Masalah, masukkan nilai berikut Jumlah likelihood

Tetapkan Tujuan: pilih sel dengan jumlah kemungkinan logaritma

Untuk: Max

Dengan Mengubah Sel Variabel: Pilih sel yang berisi koefisien regresi Anda

Jadikan Variabel Tak Terbatas Menjadi Non-Negatif: Hapus centang

Pilih Metode Penyelesaian: GRG Nonlinie.

Nilai koefisien regresinya otomatis berubah

Masukan data baru untuk melihat peluang suatu kejadian

h. Tahapan Perrhitungan Algoritma Artificial Neural Network

Pada tahap ini dilakukan perancangan model ANN dengan menentukan:

Normalisasikan Data dengan rumus

x_i 0,8 $\left(\frac{x_{i\,min}}{max-min}\right)$ + 0,1 ……**(3.6)**

Penentuan struktur ANN seperti input layer, hidden layer dan output layer.

Menentukan Bobot dan Bias ANN

Melakukan tahapan feedforward

Setiap neuron input dengan i = 1, 2, 3,…n mendapatkan sinyal dan diteruskan ke semua neuron pada hidden layer.

Setiap neuron pada hidden layer dengan j = 1, 2, 3,…n menjumlahkan sinyal-sinyal input berbobot, menggunakan persamaan sebagai berikut :

$z_in_j = v_{oj} + \sum_{i=1}^{n} x_i\ v_{ij}$ ……**(3.7)**

Hitung sinyal output pada hidden layer kemudian kirimkan sinyal yang diperoleh ke semua neuron pada lapisan output.

Pengaktifan bobot pada hidden layer dengan menggunakan persamaan

y =f (x) = ($\frac{1}{1+\ e^{-zj}}$) ……**(3.8)**

Setiap neuron output dengan k = 1, 2, 3,…,n menjumlahkan sinyal-sinyal input berbobot :

$y_in_k = w_{ok} + \sum_{i=1}^{p} z_j\ w_{jk}$ …….... **(3.9)**

hitung sinyal output pada output layer dengan menggunakan fungsi aktivasi berikut :

$y_k \frac{1}{1+e^{-y_ink}}$…… **(3.10)**

Melakukan tahapan backpropagation

Setiap neuron input dengan k = 1, 2, 3,…n menerima target pola output yang berkaitan dengan pola input dengan pelatihan. Hitung error pada output layer :

$\delta_k = (T_k - y_k) * (y_k) * 1_{yk}$ ……**(3.11)**

δ_k adalah target output yang diharapkan.

Selanjutnya hitung perubahan bobot (digunakan untuk memperbaiki nilai bobot w_{jk}

$\Delta w_{jk} = a\delta_k\ z_j$ ……..**(3.12)**

Kirimkan ini ke neuron pada lapisan tersembunyi.

Setiap neuron pada hidden layer dengan j = 1, 2, 3,…n menjumlahkan faktor delta pada hidden layer, menggunakan persamaan sebagai berikut :

$\delta_in_k = \sum_{k=1}^{m} s_k\ w_{jk}$……**(3.13)**

Nilai ini dikalikan dengan turunan dari fungsi aktivasinya dan digunakan untuk menghitung informasi error pada hidden layer dengan menggunakan persamaan berikut :

$\delta_k = \delta_{in} * (\frac{1}{1+e^{-zj}}) * (1- (\frac{1}{1+e^{-zj}}))$ …….(3.14)

Selanjutnya hitung perubahan bobot pada setiap unit keluaran (output).

$\Delta v_{jk} = a\delta_k x_i$ ……**(3.15)**

Selanjutnya hitung perubahan bobot input layer ke hidden layer.

$\Delta v_{jo} = a\ \delta_k$ ……..….**(3.16)**

Perbaharui bobot awal input layer ke hidden layer

v_{ij} (Baru) = v_{jk} (Lama) + Δv_{jk} ……..**(3.17)**

Setiap neuron output dengan k = 1,2,3,…,n memperbaharui bobot hidden layer ke output layer (j = 0,1,2,…,n):

w_{jk} (Baru) = w_{jk} (lama) + Δw_{jk}……**(3.18)**

Setiap neuron pada hidden layer , dengan memperbaharui bobot persamaan (14) dan biasnya persamaan (15) sebagai berikut :

v_{jo} (Baru) = v_{jo} (Lama) + Δv_{jo}……**(3.19)**

w_{jo} (Baru) = w_{jo} (Lama) + Δw_{jo}……**(3.20)**

Pengujian

menjumlahkan semua produk antara nilai input dan bobotnya, ditambah dengan bias. Proses ini adalah langkah penting dalam forward propagation, di mana informasi bergerak melalui jaringan dari lapisan input ke lapisan output.

$z_in_j = v_{oj} + \sum_{i=1}^{n} x_i\ v_{ij}$ ……..……(3.21)

kemudian dapat diteruskan ke fungsi aktivasi untuk mendapatkan output akhir dari neuron tersebut.

$z_j = (\frac{1}{1+e^{-z_{inj}}})\ v_{ij}$ ……(3.22)

menjumlahkan semua produk antara nilai output dari neuron sebelumnya dan bobotnya, ditambah dengan bias.

$y_in_k = w_{ok} + \sum_{i=1}^{p} z_i\ w_{jk}$ ……(3.23)

kemudian dapat diteruskan ke fungsi aktivasi untuk menghasilkan output akhir dari neuron tersebut.

$y_k = (\frac{1}{1+e^{-y_{ink}}})\ v_{ij}$ ……(3.24)

4. HASIL DAN PEMBAHASAN

a. Hasil Penelitian

Tujuan dari penelitian ini adalah membandingkan Algoritma Klasifikasi Naive Bayes, Logistic Regression, dan Artificial Neural Network. Dataset yang digunakan dalam penelitian ini adalah Data Tingkat kesejahteraan keluarga yang terdiri dari dua kategori yaitu sejahtera dan prasejahtera. Penelitian ini diawali dengan tahap pengolahan data awal. Pada tahap pengolahan data awal atau prepocessing peneliti merubah data dari nominal ke numerik agar data dapat diolah di tool Rapidminer.

b. Hasil Pengolahan Data Awal

Tahapan prepocessing adalah merubah data nominal ke numerik sehingga data dapat diolah. Pada penelitian ini penulis menggunakan 6 atribut yang digunakan dalam mengklasifikasikan tingkat kesejahteraan keluarga. Adapun atribut-atribut tersebut adalah Pendidikan (X1), Pekerjaan (X2), Kepemilikan Rumah (X3), Bantuan (X4), Bahan Bakar (X5), dan Sumber Penerangan (X6). Dan yang menjadi targetnya adalah Status Kesejahteraan (Y). Data tingkat kesejahteraan keluarga seperti tampak pada tabel berikut.

c. Data Awal

Data awal yang digunakan yaitu data sebanyak

155 data tingkat kesejahteraan keluarga

Tabel 1. Data Awal

NAMA	PENDIDIKAN	PEKERJAAN	K.RUMAH	BANTUAN	B. BAKAR	S. PENERANGAN	STATUS KESEJAHTERAAN
MIN IDRUS	TIDAK SEKOLAH	IBU RUMAH TANGGA	MILIK ORANG LAIN	TIDAK ADA	LPG	PLN	SEJAHTERA
RAHMAN HANTULI	TAMAT SD/SEDERAJAT	IBU RUMAH TANGGA	MILIK SENDIRI	ADA	KAYU	PLN	PRA SEJAHTERA
LULUN PAKILI	TIDAK TAMAT SD	NELAYAN	MILIK SENDIRI	TIDAK ADA	LPG	PLN	SEJAHTERA
ABD.RASYID YUSUF	SLTA/SEDERAJAT	PERANGKAT DESA	MILIK SENDIRI	TIDAK ADA	LPG	PLN	SEJAHTERA
RUSTAM THALIB	SLTA/SEDERAJAT	PENSIUNAN	MILIK SENDIRI	ADA	LPG	PLN	PRA SEJAHTERA
ASWIN KABATIA	TAMAT SD/SEDERAJAT	NELAYAN	MILIK ORANG LAIN	TIDAK ADA	LPG	PLN	SEJAHTERA
FATMAH A. LAKORO	TAMAT SD/SEDERAJAT	PEDAGANG	MILIK SENDIRI	TIDAK ADA	LPG	PLN	SEJAHTERA
ZABIR DARISE	SLTP/SEDERAJAT	NELAYAN	MILIK SENDIRI	ADA	LPG	PLN	PRA SEJAHTERA

d. Atribut data

dalam penelitian ini atribut yang digunakan yaitu pendidikan kepala keluarga, pekerjaan kk, kepemilikan rumah, bantuan, bahan bakar dan sumber peneranga

Tabel 2. Atribut Data

NO	KRITERIA	KODE	SUB KRITERIA	SIMBOL
1	PENDIDIKAN KEPALA KELUARGA	X1	TIDAK PUNYA IJAZAH	0
			SD/SEDERAJAT	1
			SMP/SEDERAJAT	2
			SMA/SEDERAJAT	3
			PERGURUAN TINGGI	4
2	PEKERJAAN KEPALA KELUARGA	X2	BERUSAHA SENDIRI	1
			BURUH/TIDAK TETAP/TIDAK DIBAYAR	2
			BURUH TETAP/DIBAYAR	3
			BURUH/KARYAWAN/ PEGAWAI SWASTA	4
			PEKERJA BEBAS	5
			PEKERJA KELUARGA/TIDAK DIBAYAR	6
3	KEPEMILIKAN RUMAH	X3	MILIK SENDIRI	1
			KONTRAK/SEWA	2
			LAINNYA	3
4	BANTUAN	X4	ADA	1
			TIDAK ADA	2
5	BAHAN BAKAR	X5	LPG	1
			MINYAK TANAH	2
			KAYU	3
6	SUMBER PENERANGAN	X6	LISTRIK PLN	1

e. Data Pre processing

Pre processing bertujuan merubah data awal menjadi data yang lebih sederhana atau mengubah data yang berupa kategori menjadi angka agar lebih mudah diolah

Tabel 3.Data Preprocessing

ID	X1	X2	X3	X4	X5	X6	CLASS(Y)
1	0	6	3	2	1	1	SEJAHTERA
2	1	6	1	1	3	1	PRA SEJAHTERA
3	0	5	1	2	1	1	SEJAHTERA
4	3	4	1	2	1	1	SEJAHTERA
5	3	4	1	1	1	1	PRA SEJAHTERA
6	1	5	3	2	1	1	SEJAHTERA
7	1	1	1	2	1	1	SEJAHTERA
8	2	5	1	1	1	1	PRA SEJAHTERA
9	2	4	3	2	1	1	SEJAHTERA
10	3	4	1	2	1	1	SEJAHTERA
11	3	6	1	2	1	1	SEJAHTERA
12	4	5	1	2	1	1	SEJAHTERA
13	4	4	3	2	1	1	SEJAHTERA
14	1	2	1	2	1	1	SEJAHTERA
15	1	5	1	1	1	1	PRA SEJAHTERA
16	1	5	1	2	1	1	SEJAHTERA
17	1	5	1	1	3	2	PRA SEJAHTERA
156	1	6	1	2	1	1	SEJAHTERA

f. Training Testing

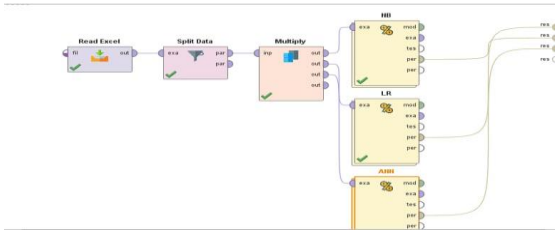
Tabel 4. Training Testing

X1 = PENDIDIKAN	X2 = PEKERJAAN	X3 = KRUMAH	X4 = BANTUAN	X5 = BAHAN BAKAR	X6 = SUMBER PENERANGAN	CLASS(Y)	KET
0	6	3	2	1	1	SEJAHTERA	Training
1	6	1	1	3	1	PRA SEJAHTERA	Training
0	5	1	2	1	1	SEJAHTERA	Training
3	4	1	2	1	1	SEJAHTERA	Training
3	4	1	1	1	1	PRA SEJAHTERA	Training
1	5	3	2	1	1	SEJAHTERA	Training
1	1	1	2	1	1	SEJAHTERA	Training
2	5	1	1	1	1	PRA SEJAHTERA	Training
2	4	3	2	1	1	SEJAHTERA	Training
3	4	1	2	1	1	SEJAHTERA	Training
3	6	1	2	1	1	SEJAHTERA	Training
4	5	1	2	1	1	SEJAHTERA	Training
4	4	3	2	1	1	SEJAHTERA	Training
1	2	1	2	1	1	SEJAHTERA	Training
1	5	1	1	1	1	PRA SEJAHTERA	Training
1	5	1	2	1	1	SEJAHTERA	Training
1	5	1	1	3	2	PRA SEJAHTERA	Training
1	6	1	2	1	1	???	Testing

g. Pengujian Menggunakan RapidMiner

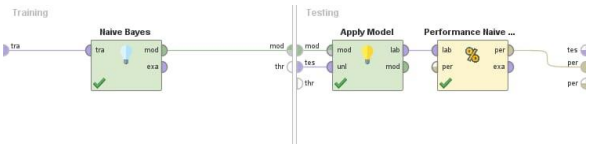
Berikut tampilan pola rapidminer menggunakan

aplikasi rapidminer menggunakan algoritma naive bayes, logistik regression dan artifisial Neural network dapat dilihat pada gambar 2.

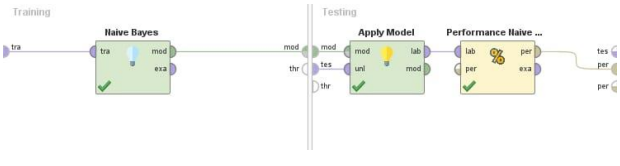


Gambar 2. Tampilan pola pada rapid miner Pada gambar diatas diawali dengan import data, Split data, dan Multiplay. Split data digunakan untuk membagi data training dan data testing menggunakan perbandingan 80:20%. Jumlah data training 124 data dan data training 32 data. Ketiga operator ini dihubungkan ke Cross validation.

Berdasarkan pengujian menggunakan aplikasi rapidminer Miner dengan algoritma naive bayes dapat dilihat pada gambar 3.berikut dan nilai performance dapat dilihat pada gambar 4.



Gambar 3. Tampilan X- Validation Naive Bayes



accuracy: 100.00% +/- 0.00% (micro average: 100.00%)

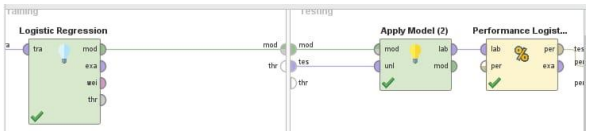
	true SEJAHTERA	true PRA SEJAHTERA	class precision
pred. SEJAHTERA	94	0	100.00%
pred. PRA SEJAHTERA	0	30	100.00%
class recall	100.00%	100.00%	

Gambar 4. Accuracy Naive Bayes

Berdasarkan data pada gambar 4. disimpulkan bahwa hasil akurasi algoritma Naive Bayes yaitu

sebesar 100%, dengan jumlah data yang di prediksi sejahtera dan true sejahtera sebanyak 94, prediksi sejahtera dan true prsejahtera sebanyak 0 dan prediksi prsejahtera true sejahtera sebanyak 0 dan prediksi prsejahtera dan true prsejahtera sebanyak 30. Untuk nilai Precision 100%.dan Recall 100%.

Berdasarkan pengujian menggunakan aplikasi rapidminer Miner dengan algoritma logistic Regression dapat dilihat pada gambar 5.dan nilai performance dapat dilihat pada gambar 6.



Gambar 5. Tampilan X- Validation Logistik Regression

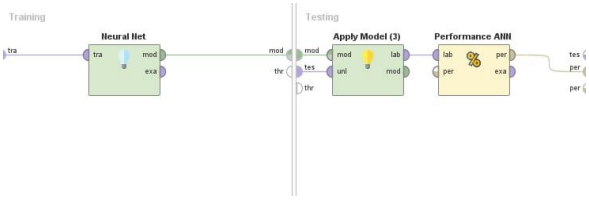
accuracy: 100.00% +/- 0.00% (micro average: 100.00%)

	true SEJAHTERA	true PRA SEJAHTERA	class precision
pred. SEJAHTERA	94	0	100.00%
pred. PRA SEJAHTERA	0	30	100.00%
class recall	100.00%	100.00%	

Gambar 6. Accuracy Logistik Regression

Berdasarkan data pada gambar 6. disimpulkan bahwa hasil akurasi algoritma Logistik Regression yaitu sebesar 100%, dengan jumlah data yang di prediksi sejahtera dan true sejahtera sebanyak 94, prediksi sejahtera dan true prsejahtera sebanyak 0 dan prediksi prsejahtera true sejahtera sebanyak 0 dan prediksi prsejahtera dan true prsejahtera sebanyak 30. Untuk nilai Precision 100%.dan Recall 100%.sebanyak 30. Untuk nilai Precision 100%.dan Recall 100%.

Berdasarkan pengujian menggunakan aplikasi rapidminer Miner dengan algoritma Artificial Neural Network dapat dilihat pada gambar 7.dan nilai performance dapat dilihat pada gambar 8.



Gambar 7. Tampilan X -Validation Artificial Neural Network

accuracy: 100.00% +/- 0.00% (micro average: 100.00%)

	true SEJAHTERA	true PRA SEJAHTERA	class precision
pred. SEJAHTERA	94	0	100.00%
pred. PRA SEJAHTERA	0	30	100.00%
class recall	100.00%	100.00%	

Gambar 8. Accuracy Artificial Neural Network

Berdasarkan data pada gambar 8. disimpulkan bahwa hasil akurasi algoritma Artificial Neural Network yaitu sebesar 100%, dengan jumlah data yang di prediksi sejahtera dan true sejahtera sebanyak 94, prediksi sejahtera dan true prsejahtera sebanyak 0 dan prediksi prsejahtera true sejahtera sebanyak 0 dan prediksi prsejahtera dan true prsejahtera sebanyak 30. Untuk nilai Precision 100%.dan Recall 100%.

h. Pengujian MenggunakanCollab



Gambar 9. script import library

Dalam tahap ini, peneliti menggunakan platform google collab untuk menjalankan program python dengan mengimpor library yang digunakan seperti perintah pandas dan sklearn.



Gambar 10. Script yang menghubungkan Excel ke drive

Dalam tahap ini, peneliti menggunakan format drive.mount untuk mengimpor model drive yang memungkinkan mengakses file dari drive, dan format df read excel untuk untuk membaca file excel dengan perbandingan 80:20 untuk data training dan testing.



Gambar 11.Potongan Kode python untuk data frame, label encorder dan encoder fitur kategorikal

Pada gambar ini peneliti menggunakan script data untuk fitur dari kategori yang digunakan yaitu pendidikan, pekerjaan, kepemilikan rumah, batuan, bahan bakar dan sumber penerangan. Script df = pd. DataFrame(data) Menggunakan fungsi dari library pandas untuk mengubah dictionary menjadi DataFrame, yaitu struktur data 2 dimensi seperti tabel (baris dan kolom). df adalah variabel yang menyimpan hasil konversi tersebut dalam bentuk DataFrame, yang akan digunakan untuk analisis selanjutnya seperti: Pra-pemrosesan (encoding, scaling), Pelatihan model klasifikasi, dan Evaluasi performa model. Menggunakan LabelEncoder untuk mengubah kategori menjadi angka dengan format label_encoder = Label Encoder() untuk mengubah data kategorikal menjadi angka



Gambar 12.Mengubah fitur kategorikal, ecode target label, membagi data training dan testing, dan menstandarisasi fitur

Pada gambar ini peneliti menggunakan script encode target label untuk mengubah label kategori menjadi angka agar bisa diproses. Membagi data training dan testing dengan format 80:20. Standard Scaler():Membuat objek untuk menstandarkan fitur (fitur akan memiliki mean = 0 dan standar deviasi = 1). fit_transform(X_train): Melakukan fit (menghitung mean dan std dari X_train) dan langsung transform data pelatihan agar distandarkan. transform(X_test): Menstandarkan data uji menggunakan mean dan std dari data pelatihan, bukan dari data uji sendiri.



Gambar 13. Membuat dan melatih model naïve bayes, logistic regression dan artificial neural network

Pada gambar 13. ini peneliti membuat dan melatih model algoritma denga model GaussianNB():Ini merupakan sebuah model Naive Bayes untuk klasifikasi yang mengasumsikan bahwa fitur-fitur yang ada dii data terdistribusi normal atau (Gaussian). Logistic Regression(max_iter=200): Membuat model regresi logistik yang digunakan untuk klasifikasi. Parameter max_iter=200 menyatakan jumlah maksimal iterasi dalam proses pelatihan, untuk memastikan model cukup waktu untuk konvergen (sesuai kebutuhan). fit(X_train_scaled, y_train): digunakan untuk melatih sebuah model regresi logistik pada data pelatihan (X_train_scaled) dan label target (y_train). predict(X_test_scaled): menghasilkan prediksi berdasarkan data uji yang sudah distandarisasi (X_test_scaled). y_pred_lr: Menyimpan hasil prediksi dari model regresi logistik. MLPClassifier (hidden_layer_sizes= (10,), dengan max_iter=2000): Membuat model Neural Network untuk klasifikasi. Ini hidden_layer_sizes=(10,) menyatakan bahwa terdapat satu layer tersembunyi dengan 10 neuron. max_iter=2000 menunjukkan bahwa model diizinkan untuk melakukan 2000 iterasi selama pelatihan (untuk memastikan konvergensi model). fit(X_train_scaled, y_train): Melatih model ANN pada data pelatihan yang sudah distandarisasi (X_train_scaled) dan label target (y_train). predict(X_test_scaled): Menghasilkan prediksi dari model ANN berdasarkan data uji yang sudah distandarisasi (X_test_scaled). y_pred_ann:

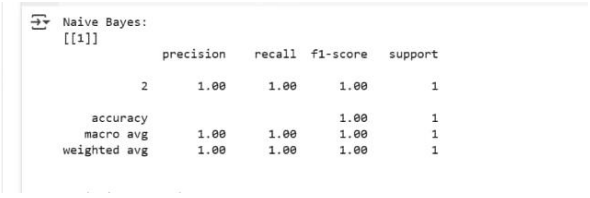
Menyimpan hasil prediksi dari model Artificial Neural Network (ANN).



Gambar 14. Menghitung confusion matrix dan clasification report untuk masing-masing model

Pada gambar 14. confusion_matrix(y_test, y_pred_nb): Menghitung confusion matrix antara nilai sebenarnya (y_test) dan hasil prediksi (y_pred_nb). Confusion Matrix memberikan sebuah gambaran visual mengenai jumlah prediksi yang benar dan salah dari setiap kelas, termasuk true positives, true negative, false positives, dan false negatives. classification_report(y_test, y_pred_nb):Menghasilkan classification report yang mencakup metrik seperti: Precision: Kemampuan model untuk mengklasifikasikan dengan benar kelas positif. Recall: Kemampuan model untuk menangkap seluruh kelas positif yang benar. F1-score: Rata-rata harmonis antara precision dan recall. Akurasi: Rasio prediksi yang benar dengan jumlah total prediksi.

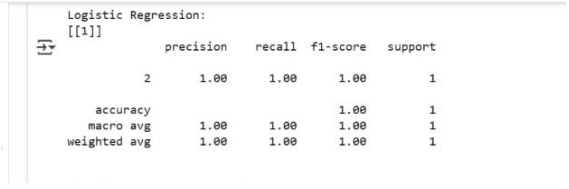
Untuk hasil accuracy menggunakan google collab pada algoritma naive bayes dapat dilihat pada gambar 16. Logistic Regression gambar 17. dan Artificial neural network pada gambar 18



Naive Bayes:
[[1]]

	precision	recall	f1-score	support
2	1.00	1.00	1.00	1
accuracy			1.00	1
macro avg	1.00	1.00	1.00	1
weighted avg	1.00	1.00	1.00	1

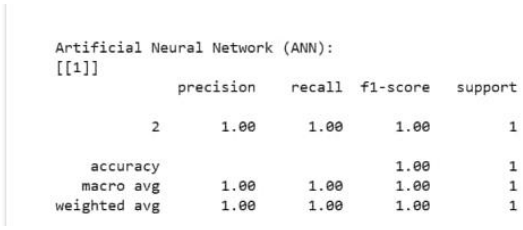
Gambar 15.Accuracy naive bayes



Logistic Regression:
[[1]]

	precision	recall	f1-score	support
2	1.00	1.00	1.00	1
accuracy			1.00	1
macro avg	1.00	1.00	1.00	1
weighted avg	1.00	1.00	1.00	1

Gambar 16. Accuracy Logistik Regression

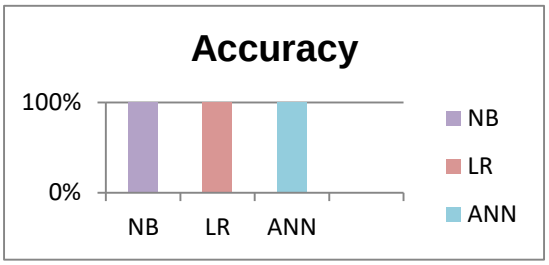


Artificial Neural Network (ANN):
[[1]]

	precision	recall	f1-score	support
2	1.00	1.00	1.00	1
accuracy			1.00	1
macro avg	1.00	1.00	1.00	1
weighted avg	1.00	1.00	1.00	1

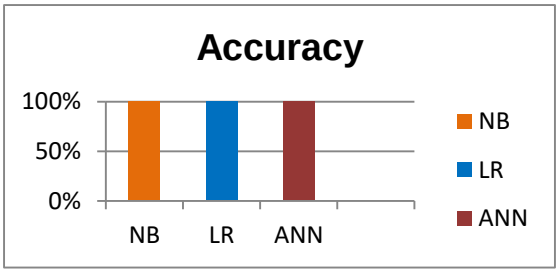
Gambar 17. Accuracy Artificial Neural Network

i. Perbandingan algoritma naive bayes, logistik Regression dan artificial neural network menggunakan rapidminer dan colab



Gambar 17. Hasil Perbandingan Algoritma Naive Bayes, Logistic Regression dan Artificial Neural Network Menggunakan Rapidminer

Pada gambar diatas dapat dilihat bahwa perbandingan rata-rata tingkat akurasi hasil klasifikasi menggunakan Naive Bayes, Logistic Regression, dan Artificial neural network memiliki akurasi, precission, dan recall yang sama. Hasil percobaan menggunakan tool rapidminer untuk algoritma naive bayes, nilai akurasinya 100%, dan Algoritma Logistic Regression akurasi 100%, dan Algoritma Artificial neural network nilai akurasinya 100%. Berdasarkan akurasi yang didapatkan Algoritma Naive Bayes, Logistic Regression dan Artificial Neural Network memiliki tingkat akurasi yang baik untuk Klasifikasi Tingkat Kesejahteraan keluarga.



Gambar 18. Hasil Perbandingan Algoritma Naive Bayes, Logistic Regression dan Artificial Neural Network Menggunakan Colab

Pada gambar 18. Diatas dapat dilihat bahwa perbandingan yang didapatkan nilai rata-rata tingkat akurasi menggunakan algoritma naive bayes, logistic regression dan artificial neural network memiliki accuracy, Precision dan recall yang sama.

5. KESIMPULAN DAN SARAN

Berdasarkan Hasil pengujian menggunakan tool rapidminer dan google colab algoritma Naive Bayes, Logistic Regression, dan Artificial Neural Network mampu mengklasifikasikan data tingkat kesejahteraan keluarga dengan sempurna, yang menunjukkan bahwa model mampu mengklasifikasikan data secara sempurna tanpa kesalahan. Dari penjelasan diatas baik rapidminer maupun google collab memiliki kemampuan yang sangat baik dalam membangun sebuah model klasifikasi yang akurat dengan akurasi yang sama pada kedua platform. Oleh karena itu tidak ada perbedaan yang signifikan dalam hal akurasi antara penggunaan rapid miner dan google colab, dan keberhasilan klasifikasi lebih bergantung pada kualitas data, pemilihan fitur, dan evaluasi model yang tepat. Dengan demikian seluruh algoritma dan tool yang digunakan terbukti memiliki akurasi yang baik untuk klasifikasi tingkat kesejahteraan keluarga.

Untuk penelitian selanjutnya diharapkan peneliti dapat membandingkan beberapa algoritma klasifikasi lainnya selain algoritma Naive Bayes, Logistic Regression dan Artificial Neural Network dengan data yang lebih kompleks dengan menggunakan tool rapid atau google Colab. Hasil yang diperoleh dari penelitian ini diharapkan dapat mendukung hasil-hasil penelitian yang sama dimasa yang akan datang. Selain itu, penelitian ini dapat dijadikan acuan untuk melakukan penelitian lebih lanjut dengan algoritma yang berbeda.

DAFTAR PUSTAKA

1. Achmad, A. D. (2022). KLASIFIKASI BREAST CANCER MENGGUNAKAN METODE LOGISTIC REGRESSION. JTRISTE, 9(1), 143–148.
2. Ainurrohmah. (2021). Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka. 4, 493–499.
3. Nasien, Dewi, dkk. "Perbandingan Implementasi Machine Learning Menggunakan Metode KNN, Naive Bayes, Dan Regresi Logistik Untuk Mengklasifikasi Penyakit Diabetes." JEKIN-Jurnal Teknik Informatika 4.1 (2024): 10-17.
4. Bagus Bambang Sumantri, R., Utami, E., Komputama Majenang, S., & AMIKOM Yogyakarta, U. (2020). Penentuan Status Tahapan Keluarga Sejahtera Kecamatan Sidareja Menggunakan Teknik Data Mining. Jurnal Teknologi Informasi, 3.
5. Damanik, S. F., Wanto, A., & Gunawan, I. (2022). Penerapan Algoritma Decision Tree C4 . 5 untuk Klasifikasi Tingkat Kesejahteraan Keluarga pada Desa Tiga Dolok. 1, 21–32.
6. Galih Pradana, D., Alghifari, M. L., Farhan Juna, M., & Dwisiwi Palaguna, S. (2022). Klasifikasi Penyakit Jantung Menggunakan Metode Artificial Neural Network. Indonesian Journal of Data and Science (IJODAS), 3(2), 55–60.
7. Homepage, J., Delvika, B., Nurhidayarnis, S., Rinada, P. D., Abror, N., & Hidayat, A. (2022). MALCOM: Indonesian Journal of Machine Learning and Computer Science Comparison of Classification Between Naive Bayes and K-Nearest Neighbor on Diabetes Risk in Pregnant Women Perbandingan Klasifikasi Antara Naive Bayes dan K-Nearest Neighbor Terhadap Resiko Diabetes Pada Ibu Hamil. 2, 68–75.
8. Julianto, A., & Andayani, S. (2024). Penerapan Data Mining Untuk Klasifikasi Produk Terlaris Menggunakan Algoritma Naive Bayes Pada Bengkel Motor.
9. Mandar, A. A. (2022). DATA MINING KLASIFIKASI KELULUSAN

MAHASISWA MENGGUNAKAN METODE NAÏVE BAYES. 4.
10. Najwa Ardelia, D., Desfianty Arifin, H., Daniswara, S., Puspita Sari, A., Studi Informatika -UPN, P., & Timur Jl Raya Rungkut Madya Gunung Anyar -Surabaya, J. (2020). Klasifikasi Harga Ponsel Menggunakan Algoritma Logistic Regression. In JOURNAL INFORMATICS AND ELECTRONICS ENGINEERING (Vol. 04, Issue 01).
11. Palingga Ninditama, I., Cholil, W., Akbar, M., & Antoni, D. (2021). Klasifikasi Keluarga Sejahtera Study Kasus : Kecamatan Kota Palembang. 15(2).
12. Rahmawati, P., & Larasati, A. (2022). PENGEMBANGAN MODEL PERSETUJUAN KREDIT NASABAH BANK DENGAN ALGORITMA KLASIFIKASI NAÏVE BAYES , DECISION TREE , DAN ARTIFICIAL NEURAL NETWORK. 17(1), 1–12.
13. Rizq, M., & Jodi, D. (2020). Algoritma dan Struktur data Fakultas komputer Algoritma.
14. Romli, I. (2021). PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK KLASIFIKASI PENYAKIT ISPA. Indonesian Journal of Business Intelligence (IJUBI), 4(1), 10. https://doi.org/ 10.21927/ijubi. v4i1. 1727
15. Sandra, M., & Erny, D. (2024). Klasifikasi Kualitas Air Menggunakan K-Nearest Neighbors , Naïve Bayes , Dan Logistic Regression. 6(4), 851–858.
16. Saraswati, E. (2021). Penerapan Algoritma Artificial Neural Network untuk. 5(2), 109–118.
17. Nasional Matematika Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5. 3, 64–71. https://journal.unnes.ac.id/sju/index.php/prisma/
18. Sutoyo, E., Asri Fadlurrahman, M., Telekomunikasi Jl Terusan Buah Batu, J., Dayeuhkolot, K., Bandung, K., & Barat, J. (2020). JEPIN (Jurnal Edukasi dan Penelitian Informatika) Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network.
19. Tri, B., Doni, R., Susanti, S., & Mubarok, A. (2021). 12~19 Naskah diterima 23 Januari 2021; direvisi 17 Februari. JURNAL RESPONSIF, 3(1). http://ejurnal.ars.ac.id/index.php/jti
20. Waldi Setiawan, Dedy Hartama, Muhammad Ridwan Lubis, Ihsan Syajidan, & Agus Perdana Windarto. (2024). Klasifikasi Peminatan Topik Keilmuan Dalam Penyelesaian Studi Menggunakan Algoritma Naive Bayes. Journal of Computing and Informatics Research, 3(2), 188–198. https://doi.org/10.47065/comforch.v3i2.1200
21. Waldi Setiawan, Dedy Hartama, Muhammad Ridwan Lubis, Ihsan Syajidan, & Agus Perdana Windarto. (2024). Klasifikasi Peminatan Topik Keilmuan Dalam Penyelesaian Studi Menggunakan Algoritma Naive Bayes. Journal of Computing and Informatics Research, 3(2), 188–198. https://doi.org/10.47065/comforch.v3i2.1200