PERBANDINGAN ALGORITMA *NAIVE BAYES, LOGISTIC REGRESSION* DAN *ARTIFICIAL NEURAL NETWORK* UNTUK KLASIFIKASI TINGKAT KESEJAHTERAAN KELUARGA

Natalia R. Simon[1], Alter Lasarudin[2], Wahyudin Hasyim[3]*
[1,2,3,]Sistem Informasi, Universitas Muhammadiyah Gorontalo
Coresponding: nataliasimon769@gmail.com*
Submitted: 03-10-2025; Accepted: 03-11-2025; Published: 03-11-2025

ABSTRACT

Batu Hijau Village is one of the coastal villages located in the Bonepantai District, Bone Bolango Regency. The village's economic livelihood depends directly on the utilization of marine and coastal resources. Most of the people living in coastal areas are classified as poor because they lack access to capital, markets, and participation in natural resource management. Currently, the classification of family welfare levels in this area is not entirely accurate, leading to misdirected subsidy distribution. Therefore, it is essential to understand the level of family welfare in the village. In this study, the author used the Naive Bayes, Logistic Regression, and Artificial Neural Network algorithms. Based on the experimental results using RapidMiner and Google Colab tools, the Naive Bayes, Logistic Regression, and Artificial Neural Network algorithms achieved the same performance with 100% accuracy, 100% precision, and 100% recall. The testing results indicate that these algorithms are capable of perfectly classifying family welfare level data, demonstrating that the models can 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. Kehidupan ekonomi desa Batu Hijau 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. 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 network* 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 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. Maka diperlukan Analisa data menggunakan klasifikasi agar dapat membantu dalam pengambilan keputusan. Klasifikasi merupakan 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[1].

Data mining menjadi metode utama untuk menganalisis pola dan pengetahuan dari data medis guna meningkatkan ketepatan dan efisiensi dalam pengambilan keputusan klinis. Sejumlah algoritma telah diterapkan dalam konteks ini, termasuk *Naïve Bayes (NB), Logistic Regression (LR)* dan *Artificial Neural Network (ANN),* yang dikenal memiliki tingkat akurasi tinggi dalam proses klasifikasi. Data mining 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[2].

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. Dalam pengklasifikasian, algoritma *Naive Bayes* menggunakan konsep probabilitas sederhana untuk menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan[3]. *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[4]. *Artificial Neural Network* 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[5]. Dengan menggunakan aplikasi *RapidMiner* dan *Google Collab* sebagai Solusi untuk menganalisis penambangan data, *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[6]. *Google Collab* merupakan platform berbasis *cloud* yang digunakan untuk menulis dan menjalankan kode Python secara daring melalui peramban (*browser*) tanpa memerlukan instalasi perangkat lunak tambahan[7].

Pada penelitian sebelumnya melakukan klasifikasi data profil nasabah yang memiliki peluang untuk mengajukan kredit pinjaman atau tidak menggunakan data mining dengan bantuan tiga algoritma yaitu Naive Bayes, Decision Tree, Artificial Neural Network (ANN). Hasil yang didapatkan menunjukkan bahwa algoritma ANN menghasilkan nilai akurasi tertinggi dengan akurasi sebesar 99.61% dan AUC = 0.983. Disusul dengan decision tree menghasilkan nilai akurasi sebesar 99.36% dan AUC = 0.999. Terakhir, algoritma naïve bayes dengan nilai akurasi = 90.79% dan AUC sebesar 0.935 [8]. Dan peneliti [9] menyimpulkan bahwa dalam mengklasifikasikan kualitas air diperoleh dari metode yang digunakan akan dibandingkan dan mengambil nilai akurasi tertinggi. Hasil akurasi dari metode KNN, Naive Bayes, dan Logistic Regression masing-masing sebesar 89.62%, 78.69%, dan 89.81%. Kemudian peneliti [10] menyimpulkan bahwa untuk mengklasifikasi penyakit diabetes. Dengan menggunakan ketiga algoritma yaitu K-Nearest Neighbors (KNN), Naive Bayes, dan Regresi Logistik. 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.

2. 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*

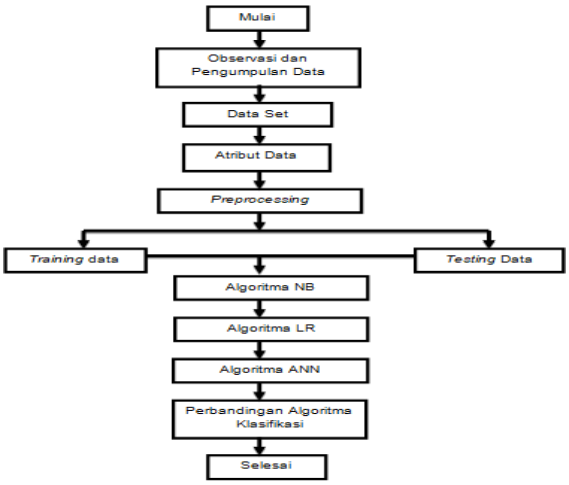
Untuk mengumpulkan informasi dan pengambilan data, maka peneliti turun langsung untuk mengamati kondisi yang ada di desa tersebut.

d. *Wawancara*

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

Tahapan *naive bayes* adalah sebagai berikut :

1. Menghitung jumlah kelas atau label
2. Menghitung jumlah kasus per kelas
3. Kalikan semua variabel
4. Bandingkan hasil perkelas



Gambar 1. Tahapan Penelitian

e. 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

Jadikan Variabel Tak Terbatas Menjadi Non-Negatif: Hapus centang

1. Pilih Metode Penyelesaian: GRG Nonlinie. Nilai koefisien regresinya otomatis berubah

2. Masukan data baru untuk melihat peluang suatu kejadian

3. *Tahapan Perhitungan Algoritma Logistic Regression*

4. Tahapan Logistic Regression adalah sebagai berikut:

5. Tentukan koefisien regresi b0,b1,b2

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

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

7. Buat nilai e logit yang dipangkatkan dengan nilai di kolom

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

8. 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

9. 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.

10. Dalam excel perhitungan koefisien regresi adalah sebagai berikut:

a. Instal add-in Solver Excel dengan mengeklik terlebih dahulu menu Beranda, lalu menu *Add-In*. Cari dan instal Solver dengan mengikuti petunjuknya.

b. 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

c. Tetapkan Tujuan: pilih sel dengan jumlah kemungkinan logaritma

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

c. Jadikan Variabel Tak Terbatas Menjadi Non-Negatif: Hapus centang

d. Pilih Metode Penyelesaian: *GRG e. Nonlinier*. Nilai koefisien regresinya otomatis berubah

e. Masukan data baru untuk melihat peluang suatu kejadian

f. Dengan Mengubah Sel Variabel: Pilih sel yang berisi koefisien regresi Anda Jadikan Variabel Tak Terbatas Menjadi Non-Negatif: Hapus centang

g. Tahapan Perhitungan Algoritma *Artificial Neural Network*

Pada tahap ini dilakukan perancangan model ANN dengan menentukan: Normalisasikan Data dengan rumus

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

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

1. Menentukan Bobot dan Bias ANN 2. 2. 2.

2. Melakukan tahapan f*eedforward*

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 :

$$z\_in_j = v_{oj} + \sum^n \; x_i \; v_{ij} \; \ldots\ldots\ldots\ldots\ldots\ldots \textbf{(3.7)}$$

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

3. Pengaktifan bobot pada hidden layer dengan menggunakan persamaan

$$y = f(x) = \left(\frac{1}{1+e^{-zj}}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textbf{(3.8)}$$

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

$$y\_in_k = w_{ok} + \sum^p \ldots\ldots\ldots\ldots z_j \; w_{jk} \ldots\ldots\ldots\ldots\ldots\ldots\ldots \textbf{(3.9)}$$

Hitung sinyal *output* pada *output layer* dengan menggunakan fungsi aktivasi berikut :

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

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.

4. 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 :

$\delta\_in_{kk} \equiv_1 \sum^m s_k\ w_{jk}$……**(3.13)**

5. 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)**

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

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

7. Selanjutnya hitung perubahan bobot *input layer* ke *hidden layer.*

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

8. 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)**

9. 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)**

10. 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 o*utput*.

$z\_in_j = v_{oj} + \sum_{i=1}^{n} x_i\ v_{ij}$ ……**(3.21)**

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

$z_j = \left(\frac{1}{1+e^{-zinj}}\right) v_{ij}$ ……**(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.

$y = \left(\frac{1}{\quad}\right) v_{ij}$ ……**(3.24)**

k $1+e^{-y_{ink}}$

3. HASIL DAN PEMBAHASAN

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.

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.

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

Sumber : Dinas sosial bone bolango 2024

Atribut data

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

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

Data *Preprocessing*

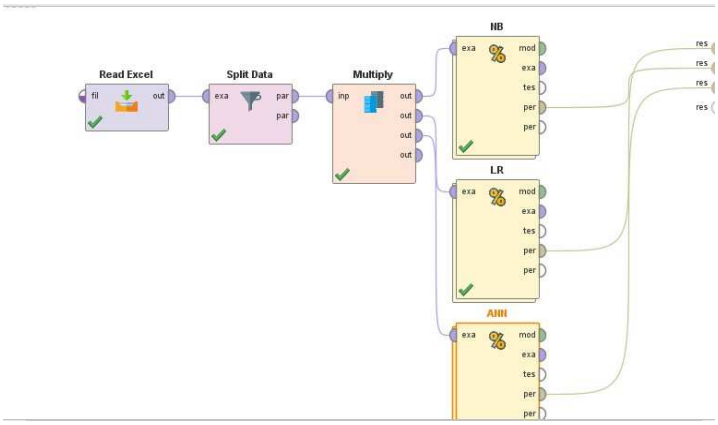
Preprocessing bertujuan merubah data awal menjadi data yang lebih sederhana atau mengubah data yang berupa kategori menjadi angka agar lebih mudah diolah.

Training Testing

Prosedur standar adalah dengan memilih 80% data *training* dan 20% data *testing* yang digunakan sebagai data latih dan data uji. Rasio ini memungkinkan model untuk dilatih pada jumlah data yang besar dan diuji pada jumlah data yg kecil, yang cukup dalam menetukan akurasi model dalam klasifikasi hasil dengan benar.

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.

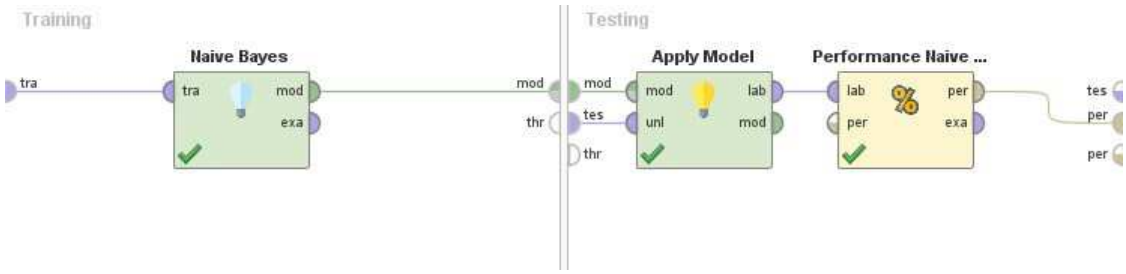


Gambar 2. Tampilan pola pada rapid miner

Pada gambar 2 di atas 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 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

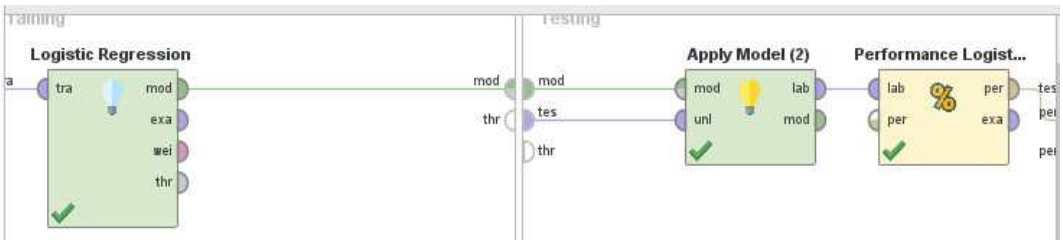
accuracy: 100.00% +/- 0.00% (micro average: 100.00%)

	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

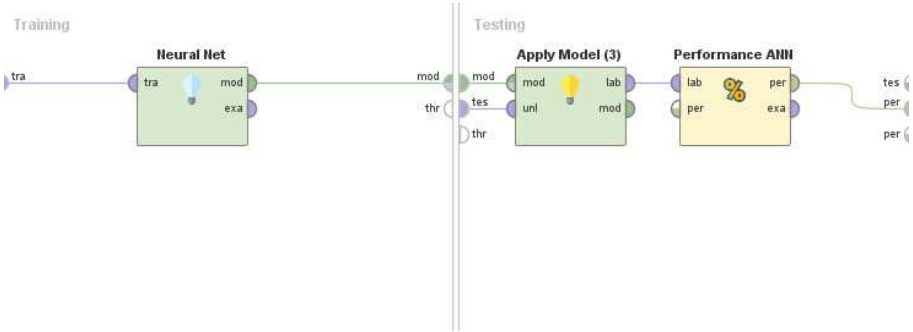
accuracy: 100.00% +/- 0.00% (micro average: 100.00%)

	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 Logistic 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 dengan algoritma Artificial Neural Network dapat di lihat pada gambar 7.dan nilai performance dapat di lihat pada gambar 8.



Gambar 5. Tampilan X -Validation Artificial Neural Network

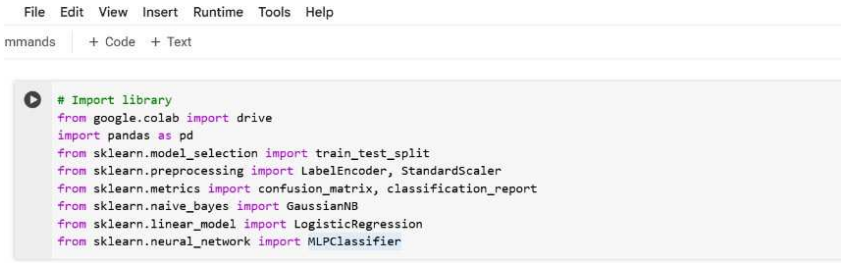
accuracy: 100.00% +/- 0.00% (micro average: 100.00%)

	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 6. Accuracy Artificial Neural Network

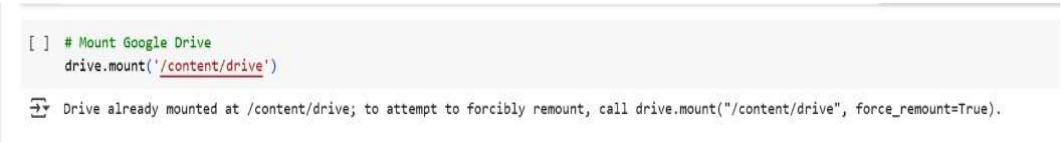
Berdasarkan data pada gambar 8. disimpulkan bahwa hasil akurasi algoritma Artificial Neural Network yaitu sebesar 100%, dengan jumlah data yang di prediksi sejahtera dan true sejahtera sebanyak 94, prediksi sejahtera dan true prsejahtera sebanyak 0 dan prediksi prsejahtera true sejahtera sebanyak 0 dan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

Pengujian Menggunakan Collab



Gambar 7. 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.

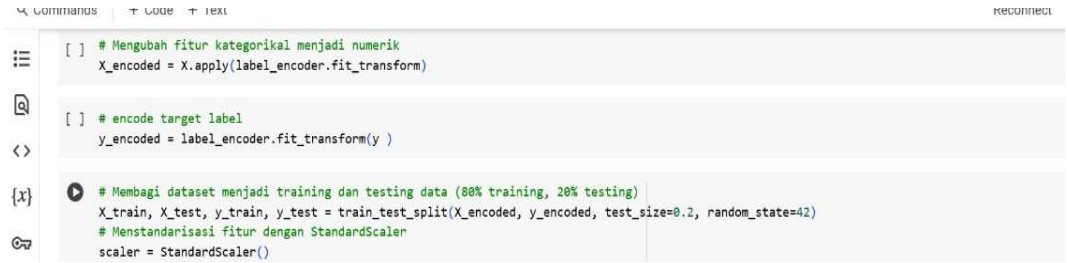


Gambar 8. Script yang menghubungkan Excel ke drive

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.

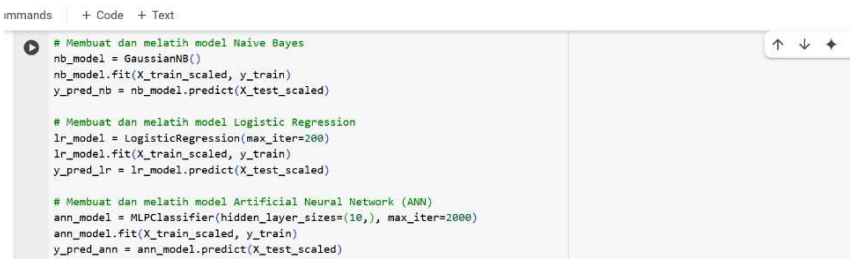
Gambar 9. Potongan Kode python untuk data frame, label encorder dan encoder fitur kategorikal

Pada gambar 11 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 10. Mengubah fitur kategorikal, ecode target label, membagi data training dan testing, dan menstandarisasi fitur

Pada gambar 12 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 11. Membuat dan melatih model naïve bayes, logistic regression

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

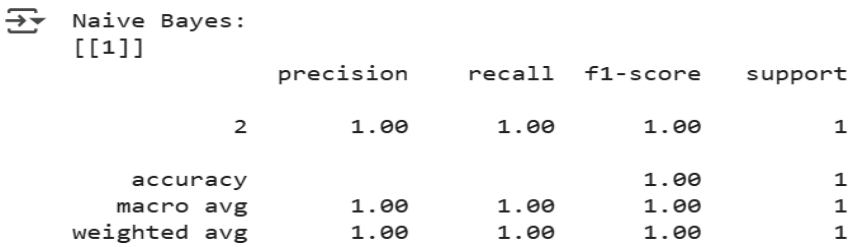
untuk melatih sebuah model regresi logistik pada data pelatihan (X_train_scaled) dan label target (y_train). predict(X_test_scaled): menghasilkan prediksi berdasarkan data uji yang sudah distandarisasi (X_test_scaled). y_pred_lr: Menyimpan hasil prediksi dari model regresi logistik. MLPClassifier (hidden_layer_sizes= (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).

```
# Menghitung confusion matrix dan classification report untuk masing-masing model
print("Naive Bayes:")
print(confusion_matrix(y_test, y_pred_nb))
print(classification_report(y_test, y_pred_nb))

print("\nLogistic Regression:")
print(confusion_matrix(y_test, y_pred_lr))
print(classification_report(y_test, y_pred_lr))

print("\nArtificial Neural Network (ANN):")
print(confusion_matrix(y_test, y_pred_ann))
print(classification_report(y_test, y_pred_ann))
```

Gambar 12. Menghitung confusion matrix dan clasification report untuk masing-masing model

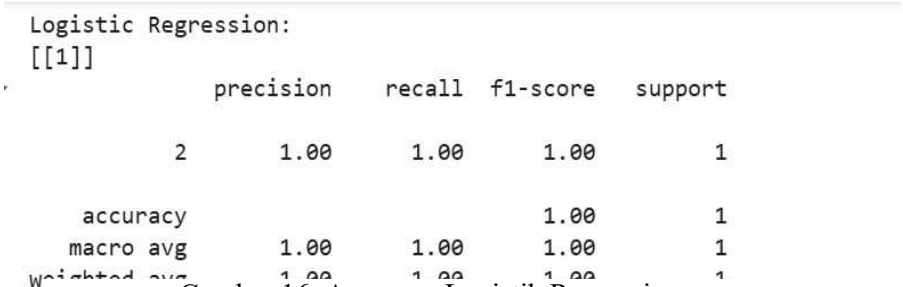
Pada gambar 14 ini confusion_matrix(y_test, y_pred_nb): Menghitung confusion matrix antara nilai sebenarnya (y_test) dan hasil prediksi (y_pred_nb). Confusion Matrix memberikan sebuah gambaran visual mengenai jumlah prediksi yang benar dan salah dari setiap kelas, termasuk true positives, true negative, 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.

Hasil accuracy menggunakan google collab pada algoritma naive bayes dapat dilihat pada gambar 15. Logistic Regression gambar 16 dan *Artificial neural network* pada gambar 17.



```
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



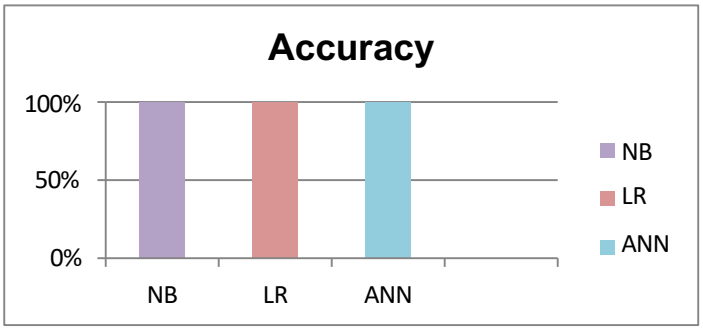
```
Logistic Regression:
[[1]]
              precision    recall  f1-score   support

           2       1.00      1.00      1.00         1

    accuracy                           1.00         1
   macro avg       1.00      1.00      1.00         1
```

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
```

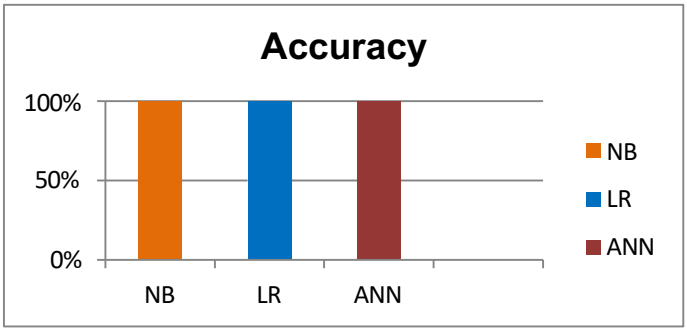
Gambar 13. Accuracy Artificial Neural Network

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



Gambar 14. Perbandingan Algoritma Naive Bayes, Logistic Regression
dan Artificial Neural Network Menggunakan Rapidminer

Pada gambar 18 di atas dapat di lihat 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 15. Hasil Perbandingan Algoritma Naive Bayes, Logistic Regression
dan Artificial Neural Network Menggunakan Colab

Pada gambar 19 di atas dapat di lihat 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.

4. KESIMPULAN

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 rapidminer 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.

DAFTAR PUSTAKA

[1] B. Delvika, S. Nurhidayarnis, and P. D. Rinada, "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," vol. 2, no. October, pp. 68–75, 2022.

[2] M. I. Abas, R. Lamusu, W. E. Pranata, S. Syahrial, and I. Ibrahim, "Penerapan Algoritma Naive Bayes Untuk Sistem Klasifikasi Status Gizi Bayi Balita," vol. 5, no. 5, pp. 1249–1260, 2025, doi: 10.47065/bulletincsr.v5i5.508.

[3] B. T. R. Doni, S. Susanti, and A. Mubarok, "Penerapan Data Mining Untuk Klasifikasi Penyakit Hepatocellular Carcinoma Menggunakan Algoritma Naïve Bayes," *J. Responsif Ris. Sains dan Inform.*, vol. 3, no. 1, pp. 12–19, 2021, doi: 10.51977/jti.v3i1.403.

[4] D. Najwa Ardelia, H. Desfianty Arifin, S. Daniswara, and A. Puspita Sari, "Klasifikasi Harga Ponsel Menggunakan Algoritma Logistic Regression," *J. Informatics Electron. Eng.*, vol. 04, no. 01, pp. 37–43, 2024.

[5] E. Sutoyo and M. A. Fadlurrahman, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, p. 379, 2020, doi: 10.26418/jp.v6i3.42896.

[6] Waldi Setiawan, Dedy Hartama, Muhammad Ridwan Lubis, Ihsan Syajidan, and Agus Perdana Windarto, "Klasifikasi Peminatan Topik Keilmuan Dalam Penyelesaian Studi Menggunakan Algoritma Naive Bayes," *J. Comput. Informatics Res.*, vol. 3, no. 2, pp. 188–198, 2024, doi: 10.47065/comforch.v3i2.1200.

[7] G. I. E. Soen, M. Marlina, and R. Renny, "Implementasi Cloud Computing dengan Google Colaboratory pada Aplikasi Pengolah Data Zoom Participants," *JITU J. Inform. Technol. Commun.*, vol. 6, no. 1, pp. 24–30, 2022, doi: 10.36596/jitu.v6i1.781.

[8] P. Rahmawati, A. Larasati, and M. Marsono, "Pengembangan Model Persetujuan Kredit Nasabah Bank Dengan Algoritma Klasifikasi Naïve Bayes, Decision Tree, Dan Artificial Neural Network," *J@ti Undip J. Tek. Ind.*, vol. 17, no. 1, pp. 1–12, 2022, doi: 10.14710/jati.1.1.1-12.

[9] M. S. Denny and D. E. Herwindiati, "Klasifikasi Kualitas Air Menggunakan K-Nearest Neighbors, Naïve Bayes, Dan Logistic Regression," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 4, pp. 851–858, 2024, doi: 10.47233/jteksis.v6i4.1649.

[10] Dewi Nasien *et al.*, "Perbandingan Implementasi Machine Learning Menggunakan Metode KNN, Naive Bayes, dan Logistik Regression Untuk Mengklasifikasi Penyakit Diabetes," *JEKIN - J. Tek. Inform.*, vol. 4, no. 1, pp. 10–17, 2024, doi: 10.58794/jekin.v4i1.640.