

Analisis Prediksi Tingkat Kesejahteraan Masyarakat Nelayan Lombok Timur Dengan Algoritma Naïve Bayes

Arnila Sandi^{1*}, Yahya², Muhammad Qusyairi³

^{1,3}Program Studi Teknik Komputer, Universitas Hamzanwadi

²Program Studi Sistem Informasi, Universitas Hamzanwadi

*Arnilasandi1878@gmail.com

Abstrak

Wilayah perairan di Indonesia lebih luas dibandingkan dengan luas daratan, oleh karena itu Indonesia dikenal sebagai negara maritim. Indonesia juga memiliki banyak sumberdaya alam yang berasal dari lautan, contohnya adalah ikan. Masyarakat pesisir pantai memanfaatkan ikan sebagai bahan matapencaharian mereka, dimana sebagian besar mereka menjadi seorang Nelayan. Mencari ikan adalah pekerjaan sehari-hari dalam memenuhi kebutuhan ekonomi keluarga. Masyarakat Nelayan dapat disebut sebagai kelompok masyarakat yang sebagian besar bekerja sebagai nelayan. Kumpulan masyarakat ini berbeda dengan masyarakat yang tinggal di daerah dataran rendah dan tinggi, mereka memiliki karakteristik yang berbeda, faktor alam banyak merubah gaya hidup mereka dan juga perekonomian mereka. Karena keunikan dari gaya hidup mereka yang membuat penulis tertarik untuk melakukan penelitian. Penelitian kali ini menggunakan data yang sudah ada untuk mendapatkan informasi baru dari data tersebut. Peneliti akan memprediksi tingkat kesejahteraan masyarakat nelayan di Lombok Timur menggunakan algoritma Naive Bayes. Untuk menunjang hasil keputusan yang akan di evaluasi peneliti menggunakan lima variabel yaitu : Pendidikan, Anggota keluarga, sumur, pekerjaan dan perumahan. Dari lima variabel tersebut peneliti berharap mendapatkan hasil akhir yang baik yang dapat digunakan untuk memprediksi tingkat kesejahteraan masyarakat nelayan di Lombok Timur. Evaluasi dilakukan menggunakan algoritma Naive Bayes, algoritma ini dipilih karena sesuai dengan karakteristik data yang kita miliki. Naive Bayes merupakan algoritma yang sederhana. Hasil akhir dari penelitian ini adalah 99% untuk nilai akurasi, 81% untuk nilai presisi dan 100% untuk recall

Kata kunci : Bayes, Klasifikasi, Nelayan, Prediksi

Abstract

The water area in Indonesia is wider than the land area, therefore Indonesia is known as a maritime country. Indonesia also has many natural resources that come from the ocean, for example fish. Coastal communities use fish as their livelihood, where most of them become fishermen. Fishing is a daily job to meet the family's economic needs. The Fishermen's Community can be called a group of people who mostly work as fishermen. This group of people is different from people who live in lowland and highland areas, they have different characteristics, natural factors have changed their lifestyle and also their economy. It is because of the uniqueness of their lifestyle that makes the author interested in conducting research. This research uses existing data to obtain new information from that data. Researchers will predict the level of welfare of fishing communities in East Lombok using the Naive Bayes algorithm. To support the results of the decisions that will be evaluated, researchers use five variables, namely: Education, family members, wells, employment and housing. From these five variables, researchers hope to get good final results that can be used to predict the level of welfare of the fishing community in East Lombok. Evaluation is carried out using the Naive Bayes algorithm, this algorithm was chosen because it suits the characteristics of the data we have. Naive Bayes is a simple algorithm. The final results of this research were 99% for accuracy value, 81% for precision value and 100% for recall.

Keywords : Bayes, Classification, Fisherman, Prediction

1. Pendahuluan

Indonesia dikenal sebagai negara maritim, karena Indonesia memiliki wilayah perairan yang cukup luas, arti kata maritim sendiri adalah seluruh kegiatan atau aktivitas yang berhubungan dengan kelautan, baik dari sumber daya alam dan juga yang bersifat teknis seperti pelabuhan dan wisata bahari (laut) [1]. Wilayah laut Indonesia memiliki luas 5,8 juta km² dan garis pantai sepanjang 81.000 km, dengan jumlah pulau 17.000, sehingga menempatkan negara Indonesia sebagai negara yang memiliki jumlah nelayan yang cukup banyak serta sumber daya alam kelautan dan perikanan yang luar biasa [2]. Kehidupan sosial masyarakat nelayan sangatlah beragam dan kompleks. Mereka memiliki karakteristik yang sangat berbeda dengan Masyarakat pada umumnya, Sebagian besar Masyarakat pesisir Pantai bermata pencaharian sebagai nelayan atau petani tambak, hingga bisa dikatakan sebagai masyarakat yang jauh dari kata sejahtera. Kesejahteraan Masyarakat harus dilakukan secara adil dan merata agar perekonomian Masyarakat dapat tumbuh dengan pesat [3].

Salah satu Provinsi di Indonesia yang memiliki wilayah perairan dan Pantai yang cukup Panjang adalah pulau Lombok yang terletak di Provinsi NTB. Lombok Timur adalah salah satu kota/kabupaten yang juga memiliki Panjang pantai 220 km dan luas perairan laut kurang

lebih sekitar 1.074 km². Dapat disimpulkan potensi pengembangan sumberdaya perikanan dan kelautan di wilayah Lombok Timur cukup tinggi, walaupun penggunaan sumberdaya alam tersebut masih tergolong tradisional [4].

Ini yang menyebabkan Tingkat kesejahteraan masyarakat nelayan masih dibawah kata sejahtera atau dapat dikatakan sebagai masyarakat miskin. Dalam penilaian Tingkat kesejahteraan masyarakat kita memiliki beberapa faktor pendukung penilaian atau dapat disebut sebagai atribut/variabel pendukung, diantaranya adalah Pendidikan, pekerjaan, perumahan, anggota keluarga, dan sumur (air bersih), dari faktor pendukung kita dapat melihat berhasilkah Masyarakat atau Sejahtera dan tidaknya Masyarakat tersebut. Pada penelitian kali ini algoritma Naïve Bayes menjadi pilihan yang digunakan untuk mengolah data.

Algoritma *Naïve Bayes* adalah sebuah teknik *machine learning* yang didasari oleh teori probabilitas yang berhubungan dengan probabilitas sebelumnya [5].

Algoritma Naïve Bayes digunakan bertujuan untuk menganalisis tingkat kesejahteraan masyarakat nelayan di Lombok Timur Nusa Tenggara Barat, dan juga untuk mengetahui performance algoritma Naïve Bayes dalam memprediksi sebuah dataset..

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Dalam menjalankan penelitian ini, penulis merujuk pada penelitian terdahulu yang berkaitan dengan topik penelitian ini antar lain :

- Amri Muliawan Nur dan rekan-rekan yang berjudul "Penerapan Metode Naïve Bayes Untuk penentuan Penerima Beasiswa Program Indonesia Pintar". Pada penelitian ini dilakukan klasifikasi data siswa yang layak dan tidak layak untuk menerima beasiswa PIP di SMAN 1 Sukamulia. Dari penelitian tersebut dihasilkan nilai akurasi sebesar 92,81% dengan menggunakan cross validation, dan hasil ini di dapat pada pengujian ke 9. Selain cross validation penelitian ini juga menggunakan Area Under Curve (AUC) sebesar 0,979%, AUC adalah salah satu parameter yang digunakan untuk menganalisis klasifikasi dalam menentukan model terbaik untuk memprediksi suatu kelas [6].
- Muhammad Wasil dan Mahfuz dalam penelitiannya yang berjudul "Analisis Persepsi Masyarakat Terhadap Kinerja Aparatur Desa Dalam Meningkatkan Pelayanan Publik Di Desa Lenek Lauk Menggunakan Metode Naive Bayes". Penelitian ini memanfaatkan angket yang diisi oleh 50 responden, mencakup 10% dari total populasi, yakni 5.539 penduduk tetap di Desa

Lenek Lauk. Hasil pengukuran menunjukkan akurasi sebesar 92.26% dengan nilai AUC pada Kurva ROC mencapai 0.924 [7].

- Pengaruh Pertumbuhan Ekonomi Terhadap Tingkat Kesejahteraan Masyarakat Kecamatan Selong Kabupaten Lombok Timur – Nusa Tenggara Barat Menggunakan Algoritma Naive Bayes, penelitian ini dilakukan oleh Yahya dan Hariman Bachtiar. Fokus penelitian terpusat pada aspek ekonomi, pada Kecamatan Selong. Hasil pengolahan data dengan algoritma tersebut menghasilkan tingkat akurasi sebesar 93,45%. Dari hasil ini, dapat disimpulkan bahwa sebanyak 1056 keluarga masyarakat menunjukkan tingkat kesejahteraan, sementara 74 keluarga lainnya masih berada dalam kondisi kurang sejahtera. [8].
- Selanjutnya adalah penelitian yang berjudul "Implementasi Algoritma Naive Bayes Untuk Memprediksi Predikat Ketuntasan Belajar Siswa Pasca Pandemi Covid 19", yang dilakukan oleh Muhammad Saiful dan kawan-kawan. Penelitian ini bertujuan untuk mengidentifikasi indikator masalah dalam pencapaian ketuntasan belajar siswa kelas XII SMAN 3 Selong selama periode pasca pandemi COVID-19. Metode yang diterapkan untuk mengatasi tantangan dan dapat memberikan solusi yang efektif dalam

- mengatasi kendala-kendala dalam pembelajaran daring di masa mendatang [9].
- Berikutnya adalah penelitian yang dilakukan oleh Bq. Andricka Candra Permana dan Intan Komala Dewi dengan judul “Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naïve Bayes Untuk Prediksi Penyakit Diabetes”. Dalam penelitian ini, disimpulkan bahwa algoritma decision tree memiliki kinerja yang lebih baik dalam memprediksi penyakit diabetes. Algoritma ini mencapai akurasi sebesar 95,58% dan nilai AUC 0,981, yang lebih tinggi daripada algoritma Naïve Bayes yang hanya mencapai akurasi 87,69% dan nilai AUC 0,947 [10].
 - Penerapan Naïve Bayes dalam Mengklasifikasikan Masyarakat Miskin di Desa Lepak adalah judul penelitian yang dilakukan oleh Wiwit Pura Nurmayanti dan kawan-kawan. Metode Naïve Bayes Classifier digunakan dalam analisis data pada penelitian ini. Dari hasil pengujian menggunakan confusion matrix, didapati bahwa, tingkat akurasi data testing dalam memprediksi hasil klasifikasi mencapai 96.63%, yang menunjukkan tingkat keakuratan yang baik. Hasil penelitian ini menunjukkan bahwa mayoritas masyarakat desa Lepak tergolong dalam kategori miskin berdasarkan klasifikasi yang dilakukan [12].

2.2. Landasan Teori

1. Data Mining

Data mining adalah serangkaian proses yang digunakan untuk mengeksplorasi nilai tambah berupa informasi yang sebelumnya tidak terdeteksi secara manual dari suatu kumpulan data dan sebagai metode yang efektif untuk mengidentifikasi pola dan informasi yang dapat digunakan untuk mengungkap hubungan antar data [14].

2. Naïve Bayes

Naïve Bayes adalah salah satu teknik yang berguna dalam mengelompokkan data ke dalam kategori yang sesuai. Dan sebagai metode statistik yang membantu memprediksi seberapa mungkin suatu data akan termasuk dalam kelas tertentu [15].

Algoritma sederhana yang menggunakan teorema Bayes untuk probabilitas bersyarat digunakan untuk mengklasifikasi data dengan mempertimbangkan frekuensi munculnya deskriptor data dari set pelatihan. [16].

Untuk menghitung probabilitas pada sebuah sampel menggunakan teorema Bayes, kita dapat merujuk pada persamaan 1 dan merumuskannya dengan kata-kata yang berbeda.

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

Keterangan

$P(H)$: Nilai probabilitas prior dari hipotesis pada sebuah sampel, biasa disebut dengan priori.

$P(X)$: merupakan evidence dari probabilitas data pelatihan. $P(H|X)$ adalah nilai probabilitas H yang mempengaruhi X (posterior density)

$P(X|H)$: merupakan probabilitas x kepada h yang disebut dengan likelihood [17].

3. Klasifikasi

Klasifikasi adalah sebuah teknik dalam data mining yang digunakan dalam berbagai bidang pembelajaran. Ini melibatkan analisis data untuk membuat model yang menggambarkan kelas-kelas yang ada dalam data yang sedang diteliti [18].

4. K-Fold Cross Validation

K-Fold Cross Validation biasa digunakan untuk melatih dan menguji model pada dataset dengan pembagian yang efisien menjadi K subset yang sama besar, untuk mendapatkan estimasi akurasi yang lebih stabil. Dalam prosesnya, setiap dataset dibagi menjadi sepuluh partisi yang sama dan diulangi sepuluh kali untuk mendapatkan nilai akurasi, sehingga memungkinkan evaluasi yang lebih konsisten terhadap kinerja model yang digunakan [19]

5. Confussion Matrix

Confusion Matrix adalah alat yang digunakan untuk mengevaluasi kinerja dari suatu metode klasifikasi. Biasanya, metode ini menggunakan

empat istilah hasil klasifikasi, yaitu True Positif (TP), True Negatif (TN), False Positif (FP), dan False Negatif (FN) [20].

6. AUC

Area Di Bawah Kurva (Area Under the Curve/AUC) merujuk pada daerah di bawah kurva Receiver Operating Characteristic (ROC). ROC adalah kurva yang menggambarkan hubungan antara sensitivitas dan spesifisitas dari suatu model atau algoritma pada berbagai titik potong. Nilai AUC secara teoritis berada di antara 0 dan 1, dengan nilai yang lebih tinggi menunjukkan kinerja yang lebih baik dari model tersebut dalam membedakan antara kelas positif dan negatif. Dengan kata lain, semakin besar nilai AUC, semakin baik kemampuan model dalam memprediksi variabel yang diteliti [21].

2.1. Tahapan Penelitian

Tahapan-tahapan penelitian dapat dilihat dari alur proses pada gambar 1 dibawah ini :



Gambar 1 Alur proses penelitian Tahapan dari alur proses penelitian diatas yaitu:

1. Diawali dengan dataset, dataset di peroleh dari aplikasi nelayan dalam bentuk .xls. dan .csv
2. Selanjutnya adalah data preparation yaitu merubah dataset yang masih mentah menjadi sebuah dataset yang siap untuk dianalisis
3. Berikutnya adalah pembagian data atau split data, data dibagi menjadi data testing dan data training
4. Data yang sudah siap dapat diproses menggunakan bahas pemrograman python dengan software Google Colab dan Dengan algoritma Naïve Bayes
5. Hasil dari pemodelan berupa nilai akurasi, precision, recall dan AUC
6. Proses selesai sampai mendapatkan hasil atau nilai akurasi tertinggi.

3. Metode Penelitian

Penelitian ini menggunakan metode eksperimental, yaitu mengumpulkan data atau informasi untuk diuji, berikut adalah metode yang dilakukan :

3.1. Teknik Pengumpulan Data

Data yang diambil dari aplikasi nelayan termasuk data yang bersifat pribadi, sementara metode pengumpulan data pada aplikasi nelayan melibatkan observasi, wawancara, dan distribusi kuesioner kepada penduduk..

1. Teknik Observasi

Teknik observasi ini dilakukan dengan melakukan pengamatan langsung ke lokasi

penelitian yaitu di 3 kecamatan yang berada di wilayah Kabupaten Lombok Timur, antarlain: Labuan Haji, Kecamatan Sakra Timur dan Kecamatan Keruak.

2. Teknik Wawancara

Teknik wawancara ini digunakan untuk mendapatkan data langsung dari penduduk Nelayan di lokasi penelitian.

3. Angket

Angket ini disusun dengan tujuan memudahkan proses wawancara dan juga sebagai dokumen pendukung untuk pengumpulan data. Angket diberikan kepada Masyarakat sekitar pesisir tempat penelitian yaitu tiga kecamatan di Lombok Timur.

3.2. Teknik Analisis Data

Metode analisis data yang digunakan dalam penelitian ini adalah pendekatan kuantitatif, dengan menggunakan data privat yang belum pernah dimanfaatkan dalam penelitian.

3.3. Teknik Pengolahan Data

Dalam pengolahan data perangkat lunak yang digunakan adalah Google Colab dengan metode Klasifikasi dan menggunakan algoritma Naïve Bayes.

3.4. Model Yang diusulkan

Penelitian ini mengusulkan menggunakan metode klasifikasi dengan algoritma Naïve Bayes. Pengujian dilakukan dengan k-fold cross validation dan untuk evaluasi menggunakan

confussion matrix dengan hasil akhir nilai akurasi dan nilai AUC serta dalam bentuk grafik

4. Hasil dan Pembahasan

Algoritma Naïve Bayes diterapkan untuk menganalisis dan memprediksi tingkat kesejahteraan masyarakat nelayan di Lombok Timur. Pengolahan data dilakukan menggunakan Google Colab dengan metode klasifikasi dan algoritma Naïve Bayes. Total data yang terkumpul adalah 1855, dengan 6 atribut yang digunakan, yaitu Nomor Anggota, Pendidikan, Anggota Keluarga, Sumur, Pekerjaan, dan Perumahan.

4.1. Pemodelan data

Menentukan Atribut/variabel pendukung, Atribut-atribut ini menjadi faktor penentu dalam menentukan kondisi akhir kesejahteraan masyarakat nelayan. Berikut adalah sample data set yang akan digunakan

NO.	NO ANGGOTA	NAMA	DUSUN
0	A5ECC2932644909702	SUBEC	DASAN BAHTEK
1	A5ECC2932644909702	MUR JOA	1 BERUK BERUK
2	A5ECC2932644909705	MUR HADAT	TEBA LINDUR
3	A5ECC2932644909705	ANSAH	TEBA LINDUR
4	A5ECC2932644909705	KAKI SI IYATI	1 BERUK BERUK
1856	A5ECC2932644909280	LALU SYAHERUL	MONTONG BELAE TIRUR
1857	A5ECC2932644909485	RAHMAN	KONKOR BERUK
1858	A5ECC2932644909485	RAHMAN	KONKOR BERUK
1859	A5ECC2932644909463	ARMAN FAHMARZ	MONTONG REASGE
1860	A5ECC2932644909485	RAHMAN	KONKOR BERUK

Pendidikan	Anggota Keluarga	Sumur	Pekerjaan
1 Tidak Sekolah	0	NonTill	Kategori Berstasi
1 Tidak Sekolah	0	NonTill	Kategori Berstasi
2 Tidak Sekolah	2	NonTill	Kategori Berstasi
1 Tidak Sekolah	1	NonTill	Kategori Berstasi
1 Tidak Sekolah	1	NonTill	Kategori Berstasi

Pendidikan	PT	NonTill	Kategori	Sambutan	Tambahan
1856	PT	0	NonTill	Kategori	Sambutan Tambahan
1857	PT	2	NonTill	Kategori	Sambutan Tambahan
1858	PT	0	NonTill	Kategori	Sambutan Tambahan
1859	PT	0	NonTill	Kategori	Sambutan Tambahan

Pendidikan	Status Masyarakat
1	Anggota Tidak Sejahtera
1	Anggota Tidak Sejahtera

Gambar 2. Sample data masyarakat nelayan di Lombok Timur
Selanjutnya adalah menentukan parameter untuk kelas, Adapun code dalam menentukan kelas adalah sebagai berikut

#input parameter untuk prediksi class

```
xTarget = df.drop(['Status Masyarakat'],axis = 1)
print(xTarget)
```

Setelah itu, kita melakukan pemodelan data, kami menggunakan algoritma Naïve Bayes.

Berikut adalah contoh kode Python untuk pemodelan Naïve Bayes:

```
#Melakukan pembuatan model
training dengan NaiveBayes
NaiveBayes =
MultinomialNB().fit(X_train,np.
ravel(y_train,order='C'))
print(NaiveBayes)
MultinomialNB()
```

Prosesnya melibatkan pembagian data menjadi data latih = 70% dan data uji = 30%, inialisasi model, pelatihan model dengan data latih, prediksi pada data uji, dan pengukuran akurasi model, berikut kodingannya

```
#Membuat data training dan
data testing dari dataset
dengan data test = 30% dan
training 70%
#dari jumlah instance X_train,
X_test,y_train,y_test=train_test_sp
lit(X,Y,test_size = 0.25,
random_state=4)
print(X_train.shape)
print(y_train.shape)
print(X_test.shape)
print(y_test.shape)
```

```
output
(1391, 5129)
(1391, 1)
(464, 5129)
(464, 1)
```

4.2. Pemodelan metode

Model ini dipilih agar dapat membuat deskripsi kelas pada dataset, yang menggambarkan status sejahtera atau tidak sejahtera. Berikut adalah code dalam menentukan target yang diinginkan

```
#target classification dari
accept / rejected LoanyTarget
= df['Status Masyarakat']
print (yTarget)
output
0 Tidak Sejahtera
1 Tidak Sejahtera
2 Tidak Sejahtera
3 Tidak Sejahtera
4 Tidak Sejahtera
5 ...
6 1850 Tidak Sejahtera
7 1851 Sejahtera
8 1852 Sejahtera
9 1853 Tidak Sejahtera
10 1854 Tidak Sejahtera
11 Name: Status Masyarakat, Length:
1855, dtype: object
```

4.3. Evaluasi

Langkah selanjutnya adalah mengevaluasi performanya menggunakan Confusion Matrix dan penghitungan menggunakan rumus bayes,

1. Confusion Matriks

Dalam kode di bawah ini, kita akan melihat bagaimana kita dapat menggunakan Confusion Matrix untuk melakukan evaluasi tersebut, serta visualisasi hasilnya.

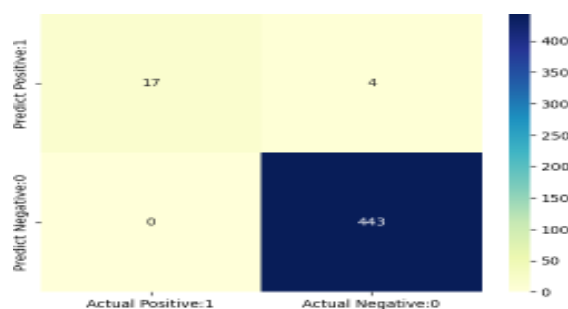
```
# Print the Confusion Matrix
and slice it into four pieces
from sklearn.metrics import
confusion_matrix
cm = confusion_matrix(y_test,
y_pred)
```

```
print('Confusion matrix\n\n', cm)
print('\nTrue Positives (TP) = ',
cm[0,0])
print('\nTrue Negatives (TN) = ',
cm[1,1])
print('\nFalse Positives (FP) = ', cm[0,1])
print('\nFalse Negatives (FN) = ',
cm[1,0])
```

output

```
Confusion matrix
[[ 17  4]
 [ 0 443]]
True Positives (TP) = 17 True
Negatives (TN) = 443 False
Positives (FP) = 4 False
Negatives (FN) = 0
```

Visualisasi dari confusion matriksnya dapat dilihat pada gambar 3 berikut ini



Gambar 3 Visualisasi confusion matriks algoritma Naïve Bayes

Dari gambar tersebut menyatakan bahwa terdapat 17 kasus True Positif, 4 kasus False Positif, tidak ada kasus False Negatif, dan 443 kasus True Negative.

2. Rumus Bayes

Untuk menggunakan rumus bayes langkah awal adalah mencari nilai probabilitas pada setiap atribut yang digunakan.

- Pada tabel 2 adalah nilai probabilitas untuk atribut Pendidikan

Tabel 1 Nilai probabilitas atribut pendidikan

Atribut	Status		Probabilitas	
	Sejahtera	Tidak Sejahtera	Sejahtera	Tidak Sejahtera
Tidak sekolah	6	218	0.05	0.13
TK	1	6	0.01	0.00
SD	74	1113	0.60	0.64
SMP	24	255	0.20	0.15
SMA	10	117	0.08	0.07
Perguruan Tinggi	8	23	0.07	0.01
Total	123	1732	1.00	1.00

- Berikutnya adalah nilai probabilitas untuk atribut anggota keluarga, hasilnya dapat dilihat pada tabel 3 berikut :

Tabel 2 Nilai probabilitas atribut anggota keluarga

Atribut	Status		Probabilitas	
	Sejahtera	Tidak Sejahtera	Sejahtera	Tidak Sejahtera
1	37	235	0.30	0.14
2	85	252	0.69	0.15
3	1	527	0.008	0.30
4	0	430	0.00	0.25
5	0	160	0.00	0.09
6	0	85	0.00	0.05
7	0	25	0.00	0.01
8	0	15	0.00	0.01
9	0	3	0.00	0.00
Total	123	1732	1.00	1.00

- Atribut selanjutnya adalah atribut perumahan, dan hasilnya dapat dilihat pada tabel 4 dibawah ini

Tabel 3 Nilai probabilitas atribut perumahan

- Selanjutnya Tabel 5 adalah tabel hasil dari probabilitas atribut sumur

Tabel 4 Nilai probabilitas atribut sumur

Atribut	Status		Probabilitas	
	sejahtera	Tidak Sejahtera	sejahtera	Tidak Sejahtera
memiliki	123	889	1	0.51
tidak memiliki	0	843	0	0.49
Total	123	1732	1	1

- Dan tabel 6, adalah tabel nilai probabilitas dari pekerjaan

Tabel 5 Nilai probabilitas atribut pekerjaan

Atribut	Status		Probabilitas	
	sejahtera	Tidak Sejahtera	sejahtera	Tidak Sejahtera
Nelayan Penuh	123	1312	1	0.76
Nelayan Sambilan	0	277	0	0.16
Nelayan Sambilan tambahan	0	143	0	0.08
Total	123	1732	1	1

- Dan yang terakhir adalah tabel 7, yang berisi tentang nilai probabilitas atribut status

Tabel 6 Nilai probabilitas atribut status

Status	Status		Probabilitas	
	Sejahtera	Tidak sejahtera	Sejahtera	tidak sejahtera
Jumlah	123	1732	1	1

Langkah berikutnya adalah mengukur kinerja model dengan menghitung akurasi, presisi, recall, dan AUC. Untuk ini, kita akan menggunakan confusion matrix, kita memiliki 17 kasus True Positif (TP), 4 kasus False Positif (FP), 0 kasus False Negatif (FN), dan 443 kasus True Negatif.

Atribut	Status		Probabilitas	
	Sejahtera	Tidak Sejahtera	Sejahtera	Tidak Sejahtera
Rumah Panggung	0	716	0	0.41
Rumah duduk semi Permanen	0	434	0	0.25
Rumah duduk permanen	123	556	1	0.32
Rumah Darurat	0	26	0	0.02
Total	123	1732	1	1

Berikut penghitungannya

1. Akurasi

$$akurasi = \frac{TP + TN}{FP + FN + TP + TN} \times 100\%$$

$$akurasi = \frac{17 + 443}{4 + 0 + 17 + 443} \times 100\%$$

$$akurasi = \frac{460}{464} \times 100\%$$

$$akurasi = 0,9913 \times 100\%$$

$$akurasi = 99,13\%$$

2. Presisi

$$presisi = \frac{TP}{FP + TP} \times 100\%$$

$$presisi = \frac{17}{4 + 17} \times 100\%$$

$$presisi = \frac{17}{21} \times 100\%$$

$$presisi = 0,8095 \times 100\%$$

$$presisi = 81\%$$

3. Recall

$$recall = \frac{TP}{FN + TP} \times 100\%$$

$$recall = \frac{17}{0 + 17} \times 100\%$$

$$recall = \frac{17}{17} \times 100\%$$

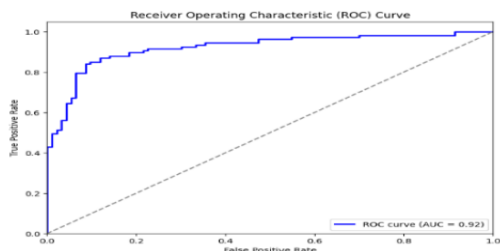
$$recall = 1 \times 100\%$$

$$recall = 100\%$$

Adapun code untuk mendapatkan nilai ROC AUC

adalah sebagai berikut:

```
#compute ROC AUC
from sklearn.metrics import roc_auc_score
ROC_AUC = roc_auc_score(y_test, y_pred) print('ROC
AUC: {:.4f}'.format(ROC_AUC))
Output
ROC AUC : 0.9048
```



Gambar 4 Grafik ROC AUC

Dari hasil penghitungan menggunakan python adalah sebagai:

	precision	recall	f1-score	support
	0	1.00	0.81	21
	1	0.99	1.00	443
accuracy				0.99 464
macro avg		1.00	0.90	0.95 464
weighted avg		0.99	0.99	0.99 464

Untuk memastikan hasil validasi yang akurat dalam penelitian ini, kami menggunakan metode

k-fold validation. Berikut code untuk menghitung k-fold validationnya.

```
From sklearn.model_selection import KFold,
StratifiedKFold,cross_val_score
from sklearn import linear_model, tree, ensemble
kf = StratifiedKFold(n_splits=5, shuffle=True,
random_state=42)
cnt = 1
# split() method generate indices to split data into
trainingand test set.
for train_index, test_index in kf.split(X, y): print('#Fold:{cnt},
Train set: {len(train_index)}, Test
set: {len(test_index)}')
cnt+=1
output
```

Fold:1, Train set: 1484, Test set:371

Fold:2, Train set: 1484, Test set:371

Fold:3, Train set: 1484, Test set:371

Fold:4, Train set: 1484, Test set:371

Fold:5, Train set: 1484, Test set:371.

5. Kesimpulan

Dari hasil penelitian yang telah dilakukan, metode klasifikasi dengan algoritma Naïve Bayes dalam memprediksi tingkat kesejahteraan masyarakat nelayan di Lombok Timur NTB, mendapatkan nilai akurasi sebesar 99%, presisi 81% dan recall 100%. Sedangkan untuk nilai AUC diperoleh 90%. Dari nilai akurasi yang diperoleh dapat disimpulkan bahwa masyarakat nelayan di Lombok Timur adalah tidak sejahtera, dan performance atau kinerja untuk algoritma Naïve Bayes sangat akurat digunakan untuk menganalisis serta memprediksi tingkat

kesejahteraan masyarakat nelayan di Lombok Timur Nusa Tenggara Barat

6. Daftar Pustaka

- [1] S. Wiranto, "Membangun kembali budaya maritim Indonesia melalui kebijakan kelautan Indonesia dengan strategi pertahanan maritim indonesia: perspektif pertahanan maritim," *J. Marit. Indones.*, vol. 8, no. 2, pp. 110–126, 2020.
- [2] I. M. Ali, L. Yudo, and D. Sianturi, "Sea Defense Strategy in Facing Maritime Security Threat in Indonesia'S Sea," *J. Prodi Strateg. Pertahanan Laut*, vol. 6, no. 2, pp. 169–188, 2021.
- [3] M. Ridho Matondang *et al.*, "Analisis Data mining dengan Metode C.45 pada Klasifikasi Kenaikan Rata-Rata Volume Perikanan Tangkap," *BRAHMANA J. Penerapan Kecerdasan Buatan*, vol. 2, no. 2, pp. 74–81, 2021.
- [4] O. Nurbayu and S. Rahmi, "HUBUNGAN PATRON-CLIENT DAN RITUAL PETIK LAUT-Studi Kasus Masyarakat Desa Tanjung Luar, Kabupaten Lombok Timur, Nusa Tenggara Barat." [Online]. Available: www.google.com
- [5] S. Nam and J. Hur, "Probabilistic Forecasting Model of Solar Power Outputs Based on the Na ve Bayes Classifier and Kriging Models," *Energies*, vol. 11, no. 11, 2018, doi: 10.3390/en11112982.
- [6] A. M. Nur, N. Nurhidayati, and I. Fathurrahman, "Penerapan Metode Naive Bayes Untuk Penentuan Penerima Beasiswa Program Indonesia Pintar (PIP).," *Infotek J. Inform. dan Teknol.*, vol. 7, no. 1, pp. 93–102, 2024, doi: 10.29408/jit.v7i1.23995.
- [7] M. Wasil, "Infotek : Jurnal Informatika dan Teknologi Analisis Persepsi Masyarakat Terhadap Kinerja Aparatur Desa Dalam Meningkatkan Pelayanan Publik Di Desa Lenek Lauk Menggunakan Metode Naive Bayes Pelayanan publik adalah kegiatan atau rangkaian kegiatan dalam ran," *Infotek J. Inform. dan Teknol.*, vol. 3, no. 2, pp. 140–146, 2020.
- [8] Y. Yahya and H. Bahtiar, "Pengaruh Pertumbuhan Ekonomi Terhadap Tingkat Kesejahteraan Masyarakat Kecamatan Selong Kabupaten Lombok Timur – Nusa Tenggara Barat Menggunakan Algoritma Naive Bayes," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 20–28, 2021, doi: 10.29408/jit.v4i1.2981.
- [9] M. Saiful1*, Syamsuddin2, and Moh. Farid Wajdi3, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Predikat Ketuntasan Belajar Siswa Pasca Pandemi Covid 19," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 96–104, 2021.
- [10] B. A. Candra Permana and I. K. Dewi Patwari, "Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naive Bayes Untuk Prediksi Penyakit Diabetes," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 63–69, 2021, doi: 10.29408/jit.v4i1.2994.
- [11] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [12] W. P. Nurmayanti, "Penerapan Naive Bayes dalam Mengklasifikasikan Masyarakat Miskin di Desa Lepak," *Geodika J. Kaji. Ilmu dan Pendidik. Geogr.*, vol. 5, no. 1, pp. 123–132, 2021, doi: 10.29408/geodika.v5i1.3430.
- [13] M. Hamzah, "Analisa dan Komparasi 5 Algoritma Klasifikasi untuk Penduduk Miskin berdasarkan Usia dan Jenis Kelamin," *J. Cakrawala Inf.*, vol. 1, no. 1, pp. 49–57, 2021, doi: 10.54066/jci.v1i1.212.
- [14] S. Ucha Putri, E. Irawan, F. Rizky, S. Tunas Bangsa, P. A. -Indonesia Jln Sudirman Blok No, and S. Utara, "Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5," *Januari*, vol. 2,

- no. 1, pp. 39–46, 2021.
- [15] M. Ridho Handoko, “SISTEM PAKAR DIAGNOSA PENYAKIT SELAMA KEHAMILAN MENGGUNAKAN METODE NAIVE BAYES BERBASIS WEB,” *J. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 50–58, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [16] C. I. Agustyaningrum, W. Gata, R. Nurfalah, U. Radiyah, and M. Maulidah, “KOMPARASI ALGORITMA NAIVE BAYES, RANDOM FOREST DAN SVM UNTUK MEMPREDIKSI NIAT PEMBELANJA ONLINE,” 2020.
- [17] N. B. Putri and A. W. Wijayanto, “Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing,” *Komputika J. Sist. Komput.*, vol. 11, no. 1, pp. 59–66, 2022, doi: 10.34010/komputika.v11i1.4350.
- [18] Y. A. Wijaya, A. Bahtiar, Kaslani, and N. R., “Analisa Klasifikasi menggunakan Algoritma Decision Tree pada Data Log Firewall,” *J. Sist. Inf. dan Manaj.*, vol. 9, no. 3, pp. 256–264, 2021, [Online]. Available: <https://ejournal.stmikgici.ac.id/>
- [19] M. Asrol, P. Papilo, and F. E. Gunawan, “Support Vector Machine with K-fold Validation to Improve the Industry’s Sustainability Performance Classification,” *Procedia Comput. Sci.*, vol. 179, no. 2020, pp. 854–862, 2021, doi: 10.1016/j.procs.2021.01.074.
- [20] Karsito and S. Susanti, “Klasifikasi Kelayakan Peserta Pengajuan Kredit Rumah Dengan Algoritma Naïve Bayes Di Perumahan Azzura Residencia,” *J. Teknol. Pelita Bangsa*, vol. 9, pp. 43–48, 2019.
- [21] *et al.*, “Implementasi Algoritma Naive Bayes Untuk Klasifikasi Penerima Beasiswa (Studi Kasus Universitas Hamzanwadi),” *Infotek J. Inform. dan Teknol.*, vol. 6, no. 1, pp. 177–188, 2023, doi: 10.29408/jit.v6i1.7529.