

TESIS
KOMPARASI ALGORITMA NAIVE BAYES DAN RANDOM FOREST UNTUK
MEMREDIKSI TINGKAT KESEHATAN LINGKUNGAN MASYARAKAT
DALAM PROGRAM SUSTAINABLE DEVELOPMENT GOALS DI
KABUPATEN LOMBOK TIMUR NUSA TENGGARA BARAT



Disusun oleh:

Nama : ZULKIPLI
NIM : 22.55.1209
Konsentrasi : Business Intelligence

PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2024

TESIS
KOMPARASI ALGORITMA NAIVE BAYES DAN RANDOM FOREST UNTUK
MEMREDIKSI TINGKAT KESEHATAN LINGKUNGAN MASYARAKAT
DALAM PROGRAM SUSTAINABLE DEVELOPMENT GOALS DI
KABUPATEN LOMBOK TIMUR NUSA TENGGARA BARAT

COMPARISON OF NAIVE BAYES AND RANDOM FOREST ALGORITHMS FOR
PREDICTING COMMUNITY ENVIRONMENTAL HEALTH LEVELS IN
SUSTAINABLE DEVELOPMENT GOALS PROGRAM IN EAST
LOMBOK DISTRICT WEST NUSA TENGGARA

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

Nama : ZULKIPLI
NIM : 22.55.1209
Konsentrasi : Business Intelligence

PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2024

HALAMAN PENGESAHAN

**KOMPARASI ALGORITMA NAIVE BAYES DAN RANDOM FOREST UNTUK
MEMPREDIKSI TINGKAT KESEHATAN LINGKUNGAN MASYARAKAT
DALAM PROGRAM SUSTAINABLE DEVELOPMENT GOALS DI
KABUPATEN LOMBOK TIMUR NUSA TENGGARA BARAT**

**COMPARISON OF NAIVE BAYES AND RANDOM FOREST ALGORITHMS FOR
PREDICTING COMMUNITY ENVIRONMENTAL HEALTH LEVELS IN
SUSTAINABLE DEVELOPMENT GOALS PROGRAM IN EAST
LOMBOK DISTRICT, WEST NUSA TENGGARA**

Dipersiapkan dan Disusun oleh

ZULKIPLI

22.55.1205

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Kamis, 01 Februari 2024

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 01 Februari 2024

Rektor

Prof. Dr. M. Suvanto, M.M.
NIK. 190302001

HALAMAN PERSETUJUAN

**KOMPARASI ALGORITMA NAIVE BAYES DAN RANDOM FOREST UNTUK
MEMPREDIKSI TINGKAT KESEHATAN LINGKUNGAN MASYARAKAT
DALAM PROGRAM SUSTAINABLE DEVELOPMENT GOALS DI
KABUPATEN LOMBOK TIMUR NUSA TENGGARA BARAT**

**COMPARISON OF NAIVE BAYES AND RANDOM FOREST ALGORITHMS FOR
PREDICTING COMMUNITY ENVIRONMENTAL HEALTH LEVELS IN
SUSTAINABLE DEVELOPMENT GOALS PROGRAM IN EAST
LOMBOK DISTRICT, WEST NUSA TENGGARA**

Dipersiapkan dan Disusun oleh

ZULKIPLI

22.55.1209

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Kamis, Tanggal 01 Februari 2024

Pembimbing Utama

Anggota Tim Penguji

Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

Tonny Hidayat, M.Kom., Ph.D.
NIK. 190302106

Pembimbing Pendamping

Hanif Al Fatta, M.Kom., Ph.D.
NIK. 190302096

Sudarmawan., MT.
NIK. 190302035

Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 01 Februari 2024
Direktur Program Pascasarjana

Prof. Dr. Kusriani., M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : ZULKIPLI
NIM : 22.55.1209
Konsentrasi : Business Intelligence

Menyatakan bahwa Tesis dengan judul berikut:

KOMPARASI ALGORITMA NAIVE BAYES DAN RANDOM FOREST UNTUK
MEMPREDIKSI TINGKAT KESEHATAN LINGKUNGAN MASYARAKAT
DALAM PROGRAM SUSTAINABLE DEVELOPMENT GOALS DI
KABUPATEN LOMBOK TIMUR NUSA TENGGARA BARAT

Dosen Pembimbing Utama : Prof. Dr. Kusriani, M.Kom
Dosen Pembimbing Pendamping : Sudarmawan., MT.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 01 Februari 2024

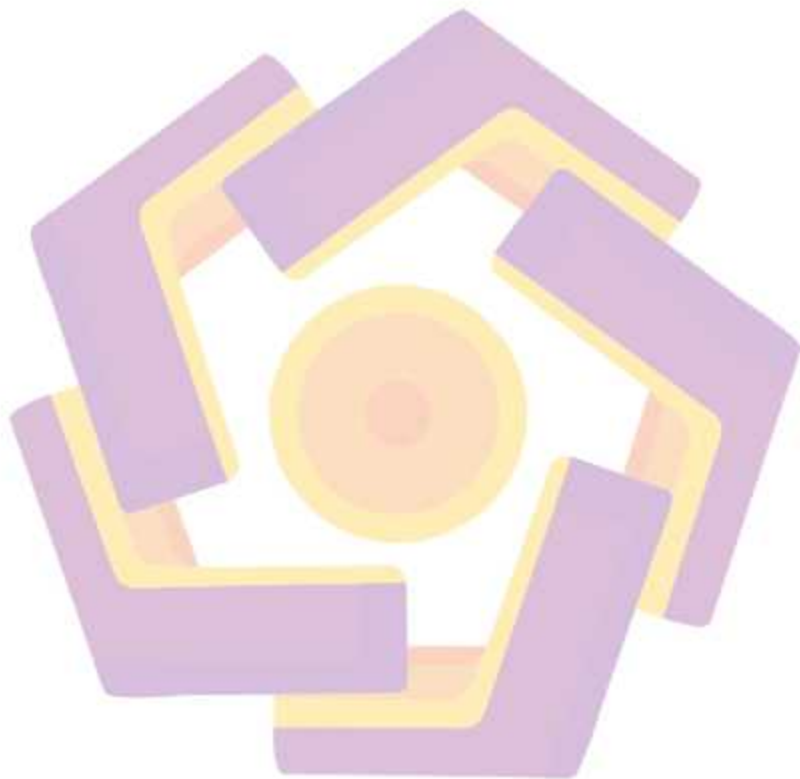
Yang Menyatakan,



ZULKIPLI

HALAMAN PERSEMBAHAN

Thesis ini ku Persembahkan Buat Ibu dan Ayahku Tercinta, Istri dan Anak-anakku. dan Untuk semua penuntut Ilmu.



HALAMAN MOTTO

" Tuntutlah ilmu sebanyak mungkin Sampai mendapat gelar muflihın, Gelar dunia perlu dijalin Dengan ajaran Rabbul 'Alamin.

"Kalau anakku ingin mendapat, Ilmu berguna ilmu yang berkat, Ibu bapakmu dan gurumu ingat, Wajib dihormati wajib di tho'ati "

TGKH. Muhammad Zainuddin Abdul Majid
Pahlawan Nasional Asal NTB



KATA PENGANTAR

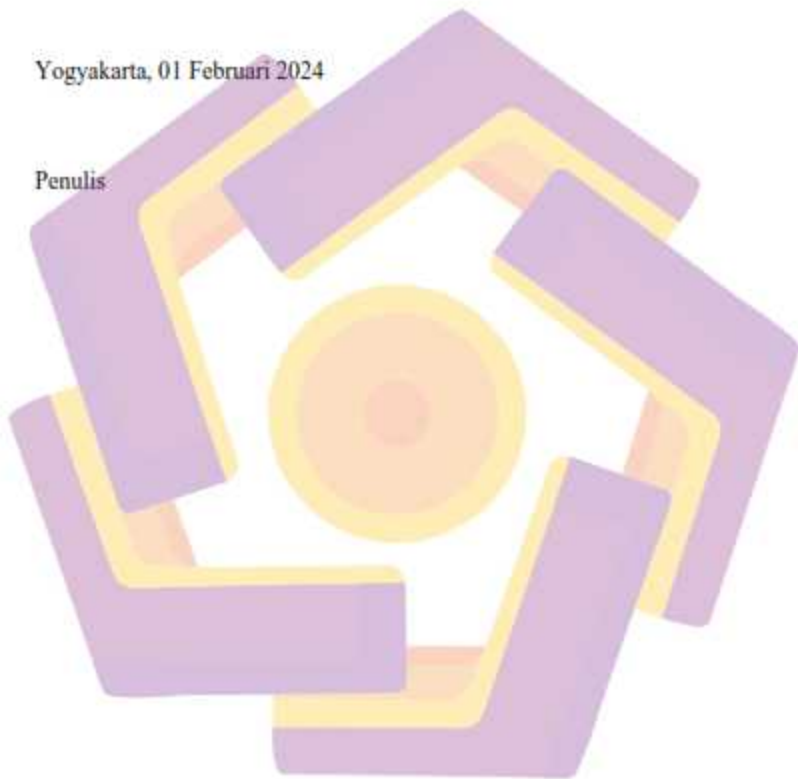
Puji Syukur kehadiran Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga saya dapat menyelesaikan tesis yang merupakan syarat untuk menyelesaikan jenjang Pendidikan S2 di Universitas AMIKOM Yogyakarta yang berjudul Perbandingan Komparasi Algoritma Naive Bayes Dan Random Forest Untuk Memprediksi Tingkat Kesehatan Lingkungan Masyarakat Dalam Program Sustainable Development Goals di sepuluh kecamatan Kabupaten Lombok Timur Nusa Tenggara Barat. Pada kesempatan ini penulis ingin mengucapkan terimakasih kepada :

1. Kedua Orang Tua dan Istri tercinta yang selalu mendoakan dan selalu mendukung selama menempuh Pendidikan Magister ini.
2. Prof. Dr. M. Suyanto, M.M. Selaku Rektor Universitas AMIKOM Yogyakarta
3. Prof. Dr. Kusriani, M.Kom Selaku Direktur Pasca Sarjana Universitas AMIKOM Yogyakarta dan juga sebagai Pembimbing Utama yang telah banyak memberikan ilmu baru dalam menulis dan memberikan masukan yang membangun dalam penelitian ini.
4. Bapak Sudarmawan, M.T. selaku dosen pembimbing pendamping yang telah banyak mengarahkan secara teknis dan memberikan saran yang membangun pada penelitian ini.
5. Bapak Tonny Hidayat, M.Kom., Ph.D. Selaku Penguji I
6. Bapak Hanif Al Fatta, M.Kom., Ph.D. Selaku Penguji II
7. Bapak Ibu Dosen dan Staf Akademika Universitas AMIKOM Yogyakarta
8. Rekan-Rekan Seperjuangan Angkatan 7 PJJ Informatika Amikom Yogyakarta

Penulis menyadari banyak sekali kesalahan dan kekeliruan yang terdapat dalam penulisan dan penyusunan tesis ini. Oleh karena itu penulis dengan senang hati menerima kritik dan saran yang bersifat membangun dari pembaca. Akhirnya penulis berharap semoga tesis ini dapat bermanfaat bagi yang membacanya.

Yogyakarta, 01 Februari 2024

Penulis



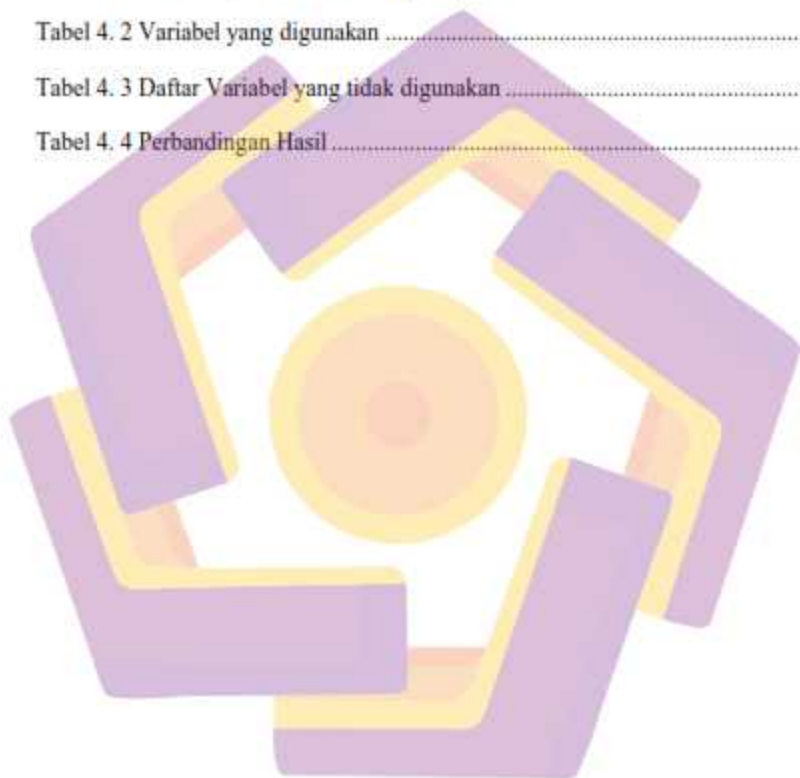
DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PERNYATAAN KEASLIAN.....	v
HALAMAN MOTTO.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
INTISARI.....	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	7
1.3. Batasan Masalah.....	8
1.4. Tujuan Penelitian.....	8
1.5. Manfaat Penelitian.....	9
BAB II TINJAUAN PUSTAKA	10
2.1. Tinjauan Pustaka.....	10
2.2. Keaslian Penelitian.....	12
2.3. Landasan Teori.....	23
BAB III METODE PENELITIAN	30
3.1. Jenis, Sifat, dan Pendekatan Penelitian.....	30

3.2. Metode Pengumpulan Data.....	30
3.3. Metode Analisis Data.....	32
3.4. Alur Penelitian.....	36
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	40
4.1. Diagnosa Penelitian.....	40
4.2. Penyusunan Perencanaan.....	45
4.3. Menentukan parameter untuk kelas.....	45
4.4. Pemodelan Metode.....	46
4.5. PraProses Data.....	47
4.6. Pemodelan Data.....	55
4.7. Algoritma Naïve Bayes.....	56
4.8. Algoritma Random Forest.....	69
4.9. Perbandingan Hasil.....	73
BAB V PENUTUP.....	76
5.1. Kesimpulan.....	76
5.2. Saran.....	76
DAFTAR PUSTAKA.....	78

DAFTAR TABEL

Tabel 2. 1 Matriks literatur review dan posisi penelitian.....	13
Tabel 4. 1 Data Masyarakat SDGs Lingkungan 10 Teratas.....	41
Tabel 4. 2 Variabel yang digunakan	49
Tabel 4. 3 Daftar Variabel yang tidak digunakan	51
Tabel 4. 4 Perbandingan Hasil	74



DAFTAR GAMBAR

Gambar 4. 1 Sample Dataset Masyarakat Yang Memperoleh Angket SDGs Lingkungan	42
Gambar 4. 2 Dataset Sebelum dilakukan Normalisasi	44
Gambar 4. 3 Kode data traintest 80 : 20 dengan random state 42.	46
Gambar 4. 4 Pencarian data Missing value	47
Gambar 4. 5 Hasil Proses Pengecekan Missing Value	48
Gambar 4. 6 Dataset setelah dilakukan normalisasi	48
Gambar 4. 7 Sample Variabel yang digunakan	49
Gambar 4. 8 Target Klasifikasi Label	52
Gambar 4.9 tf-idf_transformer menggunakan OneHotEncoder	53
Gambar 4. 10 Hasil Proses OneHotEncoder	53
Gambar 4. 11 Merubah encode nilai y target menjadi dua class yaitu 1 dan 0.....	53
Gambar 4. 12 Hasil Label Binarizer	54
Gambar 4. 13 Merubah encode nilai atribut menjadi index nilai	54
Gambar 4. 14 Hasil Merubah encode nilai atribut menjadi index nilai	54
Gambar 4. 15 Kode training dan data testing dari dataset dengan data test = 80% dan training 20%.....	56
Gambar 4. 16 Model Training Dengan Naïve Bayes	57
Gambar 4. 17 Kode Program prediksi terhadap model yang dibuat	57
Gambar 4. 18 Source Code Menampilkan Confusion Matrix	58
Gambar 4. 19 Source Code Menampilkan Confusion Matrix (Lanjutan).....	58
Gambar 4. 20 Pengujian Pertama Traintest 90:10	59

Gambar 4. 21 Confusion Matrix Tanpa Normalisasi.....	60
Gambar 4. 22 Confusion matrix setelah normalisasi.....	60
Gambar 4. 23 Pengujian Kedua TrainTest 80 : 20.....	61
Gambar 4. 24 Confusion matrix tanpa normalisasi traintest 80:20.....	61
Gambar 4. 25 Confusion matrix tanpa normalisasi traintest 80:20.....	62
Gambar 4. 26 Pengujian Ketiga TrainTest 70 : 30.....	62
Gambar 4. 27 Confusion matrix tanpa normalisasi traintest 70:30.....	63
Gambar 4. 28 Confusion matrix normalisasi traintest 70:30.....	63
Gambar 4. 29 Pengujian Keempat TrainTest 60 : 40.....	64
Gambar 4. 30 Confusion matrix tanpa normalisasi traintest 60:40.....	65
Gambar 4. 31 Confusion matrix normalisasi traintest 60:40.....	65
Gambar 4. 32 Pengujian Kelima TrainTest 50 : 50.....	66
Gambar 4. 33 Confusion matrix tanpa normalisasi traintest 50:50.....	66
Gambar 4. 34 Confusion matrix normalisasi traintest 50:50.....	67
Gambar 4. 35 Kode Membuat Model Random Forest.....	70
Gambar 4. 36 Kode Memprediksi Hasil Tes.....	70
Gambar 4. 37 Kode Membuat Confusion Matrix.....	71
Gambar 4. 38 Hasil Random Forest Untuk TrainTest 80:20.....	71
Gambar 4. 39 Perbandingan Hasil.....	74

INTISARI

Indikasi penurunan tingkat kesadaran masyarakat tentang menjaga lingkungan berpengaruh langsung terhadap kebutuhan mendesak dalam perencanaan dan pengelolaan lingkungan secara terpadu, sehingga dampaknya berpengaruh pada aspek lainnya, seperti lingkungan fisik dan sosial ekonomi. Faktanya rusaknya lingkungan berkaitan erat dengan masalah kemiskinan dan tingkat pertumbuhan ekonomi. Dalam menjaga kesehatan lingkungan masyarakat dan menerapkan Program Sustainable Development Goals di Kabupaten Lombok Timur Nusa Tenggara Barat.

Universitas Hamzanwadi bekerja sama dengan pemerintah Kabupaten Lombok Timur untuk melakukan sampling kepada masyarakat dengan jumlah data 4624 penduduk di sepuluh kecamatan di wilayah Kabupaten Lombok Timur. Tujuan penelitian ini adalah Untuk membantu memprediksi tingkat kesehatan lingkungan masyarakat di wilayah Lombok Timur, data yang dimiliki akan diklasifikasikan lalu diolah dengan algoritma Naïve Bayes dengan metode multinomial naïve bayes dan algoritma random forest yang output yang di hasilkan adalah sehat dan tidak sehat.

Hasil akurasi terbaik pada algoritma naïve bayes dengan data traintest 80 : 20 dengan nilai akurasinya 89 %, Presisi 90 %, recall 96 % dan f1- score 93 % dan pada algoritma random forest pada data traintest 80:20 nilai akurasi 99 %, presisi 99 %, recall, 100 % dan f1-score 100 %. Dari hasil tersebut algoritma random forest memiliki hasil tertinggi. Untuk penelitian selanjutnya dapat dilakukan komparasi dengan beberapa algoritma klasifikasi lainnya yang biasa digunakan pada data mining dan Machine Learning.

Kata Kunci : Naïve Bayes, Random Forest, Lingkungan Sehat, SDGs.

ABSTRACT

Indications of a decline in the level of public awareness about protecting the environment have a direct effect on the urgent need for integrated environmental planning and management, so that the impact affects other aspects, such as the physical and socio-economic environment. In fact, environmental degradation is closely related to poverty and economic growth. In maintaining the health of the community environment and implementing the Sustainable Development Goals Program in East Lombok Regency, West Nusa Tenggara.

Hamzanwadi University collaborated with the East Lombok Regency government to conduct sampling to the community with a total data of 4624 residents in ten sub-districts in the East Lombok Regency area. The purpose of this research is to help predict the level of public environmental health in the East Lombok region, the data owned will be classified and then processed with the Naïve Bayes algorithm with the multinomial naïve bayes method and the random forest algorithm whose output is healthy and unhealthy.

The best accuracy results in the naïve bayes algorithm with 80 traintest data: 20 with an accuracy value of 89%, precision 90%, recall 96% and f1-score 93% and in the random forest algorithm on 80:20 traintest data the accuracy value is 99%, precision 99%, recall, 100% and f1-score 100%. From these results the random forest algorithm has the highest results. For further research, comparisons can be made with several other classification algorithms commonly used in data mining and Machine Learning.

Keywords: *Naïve Bayes, Random Forest, Healthy Environment, SDGs.*

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Menurunnya tingkat kesadaran masyarakat tentang menjaga lingkungan dan rusaknya fungsi ekosistem menyebabkan masalah yang serius dan ancaman pada kehidupan manusia. Faktanya sejumlah aktivitas manusia memberikan tekanan terhadap lingkungan, sehingga berpotensi terjadinya degradasi lingkungan dan rusaknya fungsi ekosistem, baik di daratan, laut yang berakibat pada kerugian yang sangat besar dalam kehidupan (Khairina et al., 2020).

Indikasi penurunan tingkat kesadaran masyarakat tentang menjaga lingkungan berpengaruh langsung terhadap kebutuhan mendesak dalam perencanaan dan pengelolaan lingkungan secara terpadu, sehingga dampaknya berpengaruh pada aspek lainnya, seperti lingkungan fisik dan sosial ekonomi. Faktanya rusaknya lingkungan berkaitan erat dengan masalah kemiskinan dan tingkat pertumbuhan ekonomi.

Dalam menjaga kesehatan lingkungan masyarakat dan menerapkan program sustainable development goals di Kabupaten Lombok Timur Nusa Tenggara Barat. Universitas Hamzanwadi bekerja sama dengan pemerintah kabupaten lombok timur untuk melakukan sampling kepada masyarakat dengan jumlah data 4624 penduduk di sepuluh kecamatan di wilayah kabupaten lombok timur

Untuk membantu memprediksi tingkat kesehatan lingkungan masyarakat di wilayah Lombok Timur, data yang dimiliki akan diklasifikasikan lalu diolah

dengan membandingkan algoritma Naïve Bayes dan Random Forest. Perbandingan dilakukan agar peneliti mendapatkan nilai akurasi tertinggi dengan mengambil variabel yang sudah ditentukan di dalam angket. Mengapa Naïve Bayes dan Random Forest? Naïve Bayes sendiri adalah sebuah metode pengklasifikasian data yang digunakan untuk memprediksi probabilitas suatu class, algoritma ini sangat sederhana karena tidak memerlukan banyak data proses pengklasifikasian (Khairina et al., 2020)

Algoritma Random Forest memiliki konsep pohon keputusan dimana pada pohon keputusan tersebut memiliki prediksi kelas dan output akhir dari sebuah keputusan dengan hasil yang maksimum (Agustyaningrum et al., 2020)

Data yang digunakan dalam penelitian ini adalah data *private* yang diperoleh dari wawancara kepada masyarakat di sepuluh wilayah di kabupaten Lombok Timur di antaranya Kecamatan Selong, Kecamatan Labuhan Haji, Kecamatan Suralaga, Kecamatan Masbagik, Kecamatan Terara, Kecamatan Aikmel, Kecamatan Lenek, Kecamatan Sukamulia, Kecamatan Sikur dan Kecamatan Pringgasela. Dalam penelitian ini penulis sadar akan kelemahan dari data *private* itu sendiri diantaranya sebagai berikut : Data *private* tidak selalu dapat dijamin keakuratannya. Terkadang data pribadi dapat mengandung kesalahan atau tidak lengkap, yang dapat mengakibatkan ketidakakuratan dalam penggunaan atau analisis data tersebut. Jika data tidak dikelola dengan baik atau tidak valid, dapat mengarah pada kesimpulan yang salah atau pengambilan keputusan yang buruk.

Adapun beberapa penelitian sebelumnya yang terkait dengan metode dan algoritma yang digunakan pada penelitian ini antara lain yang telah dilakukan

mengenai perbandingan algoritma machine learning dalam pengklasifikasian tingkat kemiskinan di Indonesia dalam konteks SDGs. Sebagai contoh penelitian oleh (Nugroho & Wijayanto, n.d.). Dalam analisis ini diterapkan beberapa algoritma machine learning yaitu Decision Tree, K-Nearest Neighbor (KNN), Logistic Regression, Naïve Bayes, Random Forest (RF), dan Support Vector Machine (SVM). Hasilnya diperoleh bahwa model Naïve Bayes memiliki akurasi ketepatan dalam melakukan klasifikasi lebih tinggi sebesar 71,79% dibandingkan model yang lain. Sehingga model Naïve Bayes direkomendasikan untuk digunakan oleh para pembuat kebijakan dalam mengambil keputusan. Penelitian yang dilakukan oleh (Pranckevičius & Marcinkevičius, 2017) membandingkan pengklasifikasian menggunakan lima algoritma yaitu naïve bayes, random forest, Decisn tree, Support Vektor Machine dan Logistic Regression dalam mengevaluasi akurasi klasifikasi berdasarkan ukuran kumpulan data teks singkat untuk data ulasan produk dari amazon mendapatkan kesimpulan bahwa metode klasifikasi multi-kelas Regresi Logistik untuk ulasan produk telah mencapai akurasi klasifikasi tertinggi (min 32,43%, maks 58,50%) dibandingkan dengan metode klasifikasi Naïve Bayes, Random Forest, Decision Tree, dan Support Vector Machines. Sebaliknya, Decision Tree mendapatkan nilai akurasi rata-rata terendah (min dalam trigram: 24,10%, maks dalam uni/bi/tri-gram: 34,58%).

Penelitian yang dilakukan oleh (Kuspandi Putra, 2021) menganalisis pengaruh pertumbuhan ekonomi terhadap tingkat kesejahteraan masyarakat di desa suralaga dengan menggunakan algoritma naïve bayes dan support vector machine (svm) memberikan hasil Nilai akurasi yang tertinggi ditunjukkan oleh eksperimen

menggunakan K-Vold Validation 8 dan K-Vold Validation 10. Sedangkan toleransi yang diberikan pada K-Vold Validation 8 (0.49%) lebih kecil dari K=Vold Validation 10 sebesar (0.58%). Hal ini berarti K-Vold Validation 8 lebih ketat dari K-Vold Validation 10. Sehingga yang paling baik digunakan dalam pengambilan keputusan adalah K-Vold Validation 8 sebesar 99.62% dengan toleransi 0.49%. Hasil pengolahan data menggunakan algoritma Naive Bayes dan Support Vector Machine, sama-sama memberikan gambaran bahwa pengaruh ekonomi terhadap tingkat kesejahteraan masyarakat Desa Suralaga sangat besar dan dapat diambil kesimpulan bahwa rata-rata masyarakat Desa Suralaga termasuk dalam kategori masyarakat yang tidak sejahtera. Hal ini ditunjukkan dengan masih banyaknya masyarakat yang menggantungkan hidupnya dari bekerja sebagai buruh, tenaga kerja asing.

Dalam Penelitiannya (Purwanto et al., n.d.). Dalam mengidentifikasi habitat bentik menggunakan pendekatan segmentasi object-based image analysis (obia) dan algoritma machine learning menggunakan tiga algoritma yaitu *Support Vector Machine (SVM)*, *Decision Tree (DT)*, dan *Random Forest (RF)*. Ketiga algoritma tersebut diterapkan pada hasil segmentasi citra berbasis objek untuk menguji akurasi dari hasil klasifikasi habitat bentik. Data yang digunakan adalah citra satelit resolusi tinggi SPOT 6 yang diakuisisi pada wilayah Gugus Pulau Pari pada 20 Mei 2020. Hasil penelitian menunjukkan pendekatan segmentasi berbasis objek membentuk pola segmen yang lebih rapat pada obyek heterogen dibandingkan dengan obyek yang relatif homogen (misalnya obyek daratan dan perairan). Algoritma *machine learning* yang diterapkan pada hasil segmentasi berbasis objek

menghasilkan nilai akurasi keseluruhan (*overall accuracy*) lebih dari 70% untuk setiap algoritma, dengan rincian hasil adalah 75,83% untuk SVM, 74,17% untuk DT, dan 83,33% untuk RF. Dari ketiga algoritma *machine learning* yang diterapkan pada hasil segmentasi berbasis objek terlihat algoritma RF memiliki nilai akurasi yang paling baik dibandingkan dengan algoritma lainnya.

Penelitian yang dilakukan oleh (Samosir et al., 2019) dengan judul Komparasi Algoritma Random Forest, Naïve Bayes dan K- Nearest Neighbor Dalam klasifikasi Data Penyakit Jantung menyebutkan bahwa Metode klasifikasi yang umum digunakan antara lain Naïve Bayes, K-Nearest Neighbor, dan Random Forest. Metode Naïve Bayes menggunakan probabilitas disetiap data, metode K-Nearest Neighbor menggunakan perhitungan jarak, sedangkan metode Random Forest menggunakan beberapa pohon keputusan yang disatukan. Penulisan ini bertujuan untuk membandingkan ketiga algoritma tersebut dalam mengklasifikasikan data penyakit jantung. Perbandingan algoritma akan dilihat berdasarkan performance measure yang terdiri dari tingkatan akurasi, recall disetiap kelas, dan presisi disetiap kelas. Pada setiap algoritma diuji menggunakan cross validation. Berdasarkan hasil perbandingan terhadap 304 dataset penyakit jantung, algoritma Naïve Bayes lebih baik dan optimal dibanding dengan Algoritma, K-Nearest Neighbor dan Random Forest untuk mengklasifikasikan penyakit jantung. Hasil klasifikasi dengan algoritma Naïve Bayes memiliki rerata hasil akurasi sebesar 0,91 AUC, 0,84 CA, 0,84 F1, 0,839 Precision dan 0,84 Recall.

Penelitian tentang SDGs sebelumnya dilakukan oleh (Khairina et al., 2020) (Kuspandi Putra, 2021) (Yahya & Bahtiar, 2021) tetapi belum ada yang secara

spesifik yang mengangkat masalah SDGs Lingkungan khususnya lingkungan yang berada di Kabupaten Lombok Timur adapun kompleksitas dataset yang digunakan pada penelitian ini sangat kompleks seperti variabel yang digunakan atribut yang bervariasi, yang membahas tentang masalah Kesehatan lingkungan yang notabahnya kegiatan yang kita lakukan sehari-hari, sehingga ketika penelitian ini selesai diharapkan bukan hanya mendapatkan hasil perbandingan nilai akurasi tetapi mendapatkan nilai lebih dari itu yaitu mengubah pola hidup kita untuk menjaga lingkungan kita masing-masing.

Dari penelitian-penelitian sebelumnya algoritma Naïve Bayes dan Random Forest memiliki hasil akurasi yang bermacam-macam atau naik turun, sehingga pada penelitian yang dilakukan ini peneliti melakukan perbandingan metode klasifikasi antara algoritma Naïve Bayes dan Random Forest dengan dataset yang berbentuk kategorikal agar dapat memperoleh informasi tentang tingkat akurasi terbaik dengan menggunakan metode evaluasi yang digunakan, sehingga hasil yang didapatkan digunakan untuk memprediksi tingkat kesehatan lingkungan masyarakat di Lombok Timur, dan memberikan wawasan baru dalam bidang pengolahan data terutama data mining. Selain itu juga dari dataset yang berbentuk data kategorikal tersebut kita mendapatkan informasi tentang performance dari algoritma Naïve Bayes dan Random Forest dalam mengevaluasi sebuah dataset kepada para pembaca. Penelitian ini juga dapat memberikan kontribusi kepada Masyarakat atau Lembaga yang membutuhkan informasi.

Algoritma Naïve Bayes lebih banyak dan lebih tepat diterapkan pada data yang jumlahnya lebih besar dan dapat menangani data yang tidak lengkap (*missing*

value) serta dapat menangani noise pada data dan kuat terhadap atribut yang tidak sesuai atau tidak relevan. Akan tetapi, Algoritma *Naïve Bayes* juga memiliki kelemahan dimana sebuah probabilitas tidak bisa mengukur seberapa besar tingkat keakuratan sebuah prediksi. Selain itu, Algoritma *Naïve Bayes* juga memiliki kelemahan pada seleksi atribut sehingga dapat mempengaruhi nilai akurasi. Oleh karena itu, Algoritma *Naïve Bayes* akan dilakukan klasifikasi dengan beberapa jumlah data uji dan data training yang berbeda untuk mengukur pengaruh data uji yang berpengaruh dalam mendapatkan nilai akurasi. Oleh karena itu untuk mengetahui tingkat nilai akurasi dari *Naïve Bayes* perlu di lakukan komparasi dengan random forest untuk menemukan algoritma terbaik dalam melakukan prediksi tingkat kesehatan lingkungan masyarakat di Kabupaten Lombok Timur. Pada penelitian yang akan dilakukan adalah melakukan komparasi atau perbandingan tingkat akurasi dari algoritma *naïve Bayes* dan algoritma random forest dalam mencari algoritma yang terbaik dalam melakukan prediksi tingkat kesehatan lingkungan di sepuluh Kecamatan Kabupaten Lombok Timur dengan nilai akurasi yang tinggi.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang di atas maka rumusan masalah ini adalah :

1. Bagaimana performance algoritma *naïve Bayes* dan random forest dalam memprediksi tingkat kesehatan lingkungan masyarakat dalam program

sustainable development goals di Kabupaten Lombok Timur Nusa Tenggara Barat ?

2. Berapa hasil dari performance komparasi algoritma naïve bayes dan random forest dalam memprediksi data tingkat kesehatan lingkungan masyarakat dalam program sustainable development goals di Kabupaten Lombok Timur Nusa Tenggara Barat ?
3. Faktor apa saja yang mempengaruhi komparasi antara algoritma naïve bayes dan random forest dalam memprediksi data tingkat kesehatan lingkungan masyarakat dalam program sustainable development goals di Kabupaten Lombok Timur Nusa Tenggara Barat ?

1.3. Batasan Masalah

Agar penelitian ini tidak keluar dari pembahasan, maka pembahasan yang di bahas pada penelitian ini adalah sebagai berikut :

1. Proses pengolahan data hanya menggunakan dua algoritma yaitu naïve bayes dan random forest
2. Parameter yang digunakan dalam penelitian ini adalah nilai akurasi
3. Pengolahan data menggunakan Bahasa pemrograman python.

1.4. Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut :

1. Untuk mengetahui kinerja algoritma naïve bayes dan random forest dalam memprediksi tingkat kesehatan lingkungan masyarakat dalam program

sustainable development goals di Kabupaten Lombok Timur Nusa Tenggara Barat ?

2. Mengetahui tingkat keakuratan dari algoritma naïve bayes dan random forest dalam memprediksi tingkat Kesehatan lingkungan masyarakat dalam program sustainable development goals di Kabupaten Lombok Timur Nusa Tenggara Barat ?
3. Menghasilkan pengetahuan bagi pembaca tentang performance dari algoritma naïve bayes dan random forest dalam memprediksi tingkat Kesehatan lingkungan masyarakat dalam program sustainable development goals di Kabupaten Lombok Timur Nusa Tenggara Barat ?

1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut :

1. Diharapkan penelitian ini dapat menambah wawasan dan pengembangan ilmu pengetahuan secara teoritis dan praktek.
2. Hasil penelitian juga diharapkan dapat menjadi pengalaman dan pengetahuan serta rujukan tentang manfaat dari pengolahan data menggunakan metode klasifikasi dan algoritma Naïve Bayes dan Random Forest.
3. Memberikan inovasi dalam pengolahan dataset menggunakan metode klasifikasi dengan dua algoritma Naïve Bayes dan Random Forest untuk mendapatkan keakuratan data.

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Beberapa penelitian sebelumnya dijadikan sebagai bahan acuan dan tinjauan pustaka untuk penelitian yang dilakukan saat ini diantaranya adalah:

Penelitian yang dilakukan oleh (van der Heide et al., 2019) membandingkan regresi logistik berganda, sebuah metode linier, dengan naive Bayes dan random forest, 2 metode pembelajaran mesin nonlinier. ketiga metode tersebut untuk memprediksi kelangsungan hidup individu hingga laktasi kedua pada sapi perah dara. Kumpulan data yang digunakan untuk prediksi terdiri dari 6.847 sapi dara yang lahir antara Januari 2012 dan Juni 2013, dan memiliki hasil kelangsungan hidup yang telah diketahui. Setiap hewan memiliki 50 nilai estimasi pemuliaan genom yang tersedia saat lahir dan hingga 65 variabel fenotipik yang terakumulasi dari waktu ke waktu. Kelangsungan hidup diprediksi pada 5 momen dalam kehidupan: saat lahir, pada usia 18 bulan, saat pertama kali beranak, pada 6 minggu setelah beranak, dan pada 200 hari setelah beranak. Kumpulan data dibagi secara acak menjadi 70% set pelatihan dan 30% set pengujian untuk mengevaluasi kinerja model untuk validasi 20 kali lipat. Metode-metode tersebut dibandingkan untuk akurasi, sensitivitas, spesifisitas, nilai area di bawah kurva (AUC), kontras antara kelompok untuk hasil prediksi, dan peningkatan hewan yang bertahan hidup dalam skenario praktis. Pada saat lahir dan 18 bulan, semua metode memiliki kinerja yang

tumpang tindih; tidak ada metode yang secara signifikan mengungguli metode lainnya.

Penelitian sebelumnya yang dilakukan oleh (Thaufik Rizaldi, 2020) dalam memprediksi dan mengklasifikasikan Perbandingan Teknik Pembagian Data untuk Klasifikasi Sarana Akses Air menggunakan dua metode yaitu Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier. Penelitian ini membandingkan teknik pembagian data K-Means Clustering, K-Medoid dan Hold-out pada algoritma klasifikasi K-Nearest Neighbor dan Naïve Bayes Classifier menghasilkan K-Medoid sebagai pemodelan pembagian data terbaik dengan nilai akurasi 89,39% pada algoritma K-Nearest Neighbor dengan parameter $K = 10$.

Dalam penelitian yang di lakukan oleh (Azhari et al., 2021) dengan melakukan perbandingan akurasi, recall dan presisi kalsifikasi pada algoritma C4.5, Random Forest, SVM, dan Naïve Bayes. Pada penelitian ini algoritma random Forest memiliki hasil akurasi terkecil dibandingkan tiga algoritma lainnya. Hal ini yang membuat peneliti ingin melakukan perbandingan hasil dari performance algoritma Random Forest dan Naïve Bayes menggunakan dataset yang dimiliki. Penelitian sebelumnya dilakukan dengan menggunakan 200 data dengan perbandingan 70% data training dan 30% data testing, dan penelitian yang akan dilakukan memiliki data lebih banyak dari ini. Pada penelitian sebelumnya hasil akurasi yang didapat 95% untuk SVM, C4.5 dan Naïve Bayes memiliki akurasi yang sama yaitu 86,67% dan yang terakhir adalah random Forest dengan nilai akurasi 83,33%.

Komparasi akurasi pada Naïve Bayes dan Random Forest dalam klasifikasi penyakit liver pernah dilakukan oleh (Lubis et al., n.d.) Penelitian tersebut dilakukan untuk mengidentifikasi penyakit liver melalui diagnosis sebagai proses untuk menentukan sifat yang dapat membedakannya dari keadaan yang terjadi. Data diambil dari Indian Liver Patient Dataset yang diperoleh dari UCI Machine Learning Repository dan memiliki 583 record data. Hasil dari pengujian didapat algoritma Random Forest memiliki nilai akurasi terbaik yaitu sebesar 70,60%, sedangkan Naïve Bayes memperoleh akurasi sebesar 55,80%.

Penelitian yang dilakukan oleh (Abdul Aziz, n.d.) dalam mengimplementasi algoritma rough set dan naïve bayes untuk mendapatkan rule dalam menyeleksi pemohon bantuan fasilitas rumah ibadah, hasil dari penelitian ini naïve bayes mendapatkan akurasi 77 % lebih rendah dari algoritma rough set yang memperoleh nilai akurasi 92 %.

2.2. Keaslian Penelitian

Pada Penelitian yang akan dilakukan ini akan mencoba melakukan perbandingan nilai akurasi dari NBC dan Random Forest untuk mendeteksi tingkat Kesehatan lingkungan Masyarakat dalam program Sustainable Developments Goals. Komparasi Algoritma dilakukan oleh banyak peneliti. Untuk mengetahui posisi penelitian yang akan dilakukan dengan beberapa penelitian yang dilakukan oleh para peneliti yang di jadikan sebagai acuan dalam melakukan penelitian ini, perbandingannya dapat dilihat di tabel 2.1 di bawah ini.

Tabel 2. 1 Matriks literatur review dan posisi penelitian

KOMPARASI ALGORITMA NAIVE BAYES DAN RANDOM FOREST UNTUK MEMPREDIKSI TINGKAT KESEHATAN LINGKUNGAN MASYARAKAT DALAM PROGRAM SUSTAINABLE DEVELOPMENT GOALS DI KABUPATEN LOMBOK TIMUR NUSA TENGGARA BARAT

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	Perbandingan Teknik Pembagian Data untuk Klasifikasi Sarana Akses Air pada Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier	Thaufik Rizaldi S, SNTIKI, 2020	Untuk mengetahui indikator kelayakan dalam menentukan rekomendasi akses air. Penelitian ini membandingkan teknik pembagian data K-Means Clustering	Perbandingan teknik pembagian data terbaik menghasilkan K-Medoid secara berurutan pada algoritma K-Nearest Neighbor(KNN) pada parameter K=10 dengan nilai akurasi 89,39%, nilai recallsebesar 79,44% dan nilai presisi sebesar 81,77% dan Naïve Bayes Classifier dengan nilai akurasi 40,91%, nilai recallsebesar 55,19% dan nilai presisi sebesar 45,69%.	Berdasarkan hasil penelitian Maka percobaan yang dilakukan menggunakan data Sarana Akses Air dapat digunakan K-Medoid sebagai teknik pembagian data untuk klasifikasi Sarana Akses air pada algoritma klasifikasi	Penelitian yang dilakukan oleh Thaufik rizaldi S. menggunakan metode klasifikasi menggunakan algoritma KNN dan naïve bayes classifier Sedangkan penelitian yang akan dilakukan menggunakan dua algoritma yaitu Naïve Bayes dan Random Forest untuk mencari akurasi terbaik yang dapat digunakan dalam pengambilan keputusan.

Tabel 2. 1 Matriks literatur review dan posisi penelitian (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
2	Komparasi Algoritma Naive Bayes , Random Forest Dan Svm Untuk Memprediksi Niat Pembelanja Online	Cucu Ika Agustyaningrum, Windu Gata, Ridan Nurfalab, Ummu Radiyah, Mawadatul Maulidah, Jurnal Informatika, 2020	Mengklasifikasikan tiga algoritma terbaik dalam menentukan minat konsumen dalam menentukan pilihan barang mana yang akan di beli	Algoritma <i>Naive Bayes</i> dan <i>Support Vector Machine</i> merupakan algoritma yang tidak direkomendasikan untuk penilaian data <i>Online Shoppers Purchasing Intention Dataset</i>	Untuk mendapatkan hasil akurasi lebih baik lagi, penelitian selanjutnya dapat menggunakan metode lainnya sebagai bahan perbandingan	Penelitian yang dilakukan oleh cucu ika Agustyaningrum dan rekan rekan adalah dengan mengklasifikasikan minat konsumen dalam menentukan pilihan barang mana yang akan dibeli dan menggunakan tiga algoritma yaitu naive bayes, random forest dan svm untuk mencari tingkat akurasiya sedangkan pada penelitian ini algoritma Naive Bayes dan Random Forest digunakan untuk mendapatkan tingkat keakurasiyan data dalam mengklasifikasikan tingkat kebersihan lingkungan masyarakat dalam program SDGs di kabupaten Lombok Timur

Tabel 2. 1 Matriks literatur review dan posisi penelitian (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
3	Komparasi Akurasi Pada Naïve Bayes Dan Random Forest Dalam Klasifikasi Penyakit Liver	Ahmadi Irmansyah Lubis, Umri Erdiansyah, Rosma Siregar, Jurnal CESS, 2022	Melakukan komparasi terhadap metode Naïve Bayes dan Random Forest dalam klasifikasi data pasien penyakit liver	Berdasarkan hasil pengujian yang dilakukan dalam memperoleh nilai akurasi perhitungan klasifikasi menggunakan Confusion Matrix, maka metode Random Forest memperoleh hasil yang terbaik yaitu dengan peroleh akurasi sebesar 70.60 % bila dibandingkan dengan Naïve Bayes yang hanya memperoleh akurasi sebesar 55.80 %. Sehingga Random Forest memiliki performa kinerja yang lebih unggul dalam perolehan akurasi yang dihasilkan dalam klasifikasi penyakit liver.	Kelemahan atau kekurangan dari penelitian ini adalah jumlah dataset yang digunakan tidak banyak, untuk mendapatkan nilai akuarasi lebih baik lagi sebaiknya menggunakan dataset yang jumlahnya lebih banyak.	Perbandingan penelitian sebelumnya dengan penelitian yang akan dilakukan adalah memiliki metode yang sama yaitu klasifikasi data, selain itu ada perbedaan pada objek yang diteliti, serta penggunaan algoritma juga lebih bervariasi atau lebih banyak. Penelitian yang saya lakukan menggunakan dua algoritma sebagai pengambil keputusan untuk memprediksi

Tabel 2. 1 Matriks literatur review dan posisi penelitian (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
4	Analisis Perbandingan Pengaruh Pertumbuhan Ekonomi Terhadap Tingkat Kesejahteraan Masyarakat Pada Desa Suralaga Dengan Menggunakan Algoritma Naive Bayes Dan Support Vector Machine (Svm)	Fathurrahman, Yupi Kuspani Putra, Jurnal INFOTEK, 2021	Pengolahan data menggunakan algoritma Naive Bayes dan Support Vector Machine, untuk memberikan gambaran bahwa pengaruh ekonomi terhadap tingkat kesejahteraan masyarakat Desa Suralaga kabupaten lombok timur.	Berdasarkan hasil penelitian yang telah dilakukan, Nilai akurasi yang tertinggi ditunjukkan oleh eksperimen menggunakan K-Vold Validation 8 dan K-Vold Validation 10. Sedangkan toleransi yang diberikan pada K-Vold Validation 8 (0.49%) lebih kecil dari K-Vold Validation 10 sebesar (0.58%). Hal ini berarti K-Vold Validation 8 lebih ketat dari K-Vold Validation 10. Sehingga yang paling baik digunakan dalam pengambilan keputusan adalah K-Vold Validation 8 sebesar 99.62% dengan toleransi 0.49%.	Dalam melakukan perbandingan algoritma tidak disebutkan algoritma mana yang terbaik dalam melakukan analisis data.	Dalam penelitian yang dilakukan sebelumnya tidak dijelaskan berapa jumlah data yang di analisis dan pada penelitian berikutnya data yang akan diteliti sudah jelas berapa data yang akan diteliti yang berjumlah kurang lebih 4000 data yang di dapatkan dari kkn tematik desa gemilang.

Tabel 2. 1 Matriks literatur review dan posisi penelitian (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
5	Klasifikasi masyarakat miskin menggunakan metode <i>naïve bayes</i>	Haditsah Annur Jurnal Ilmiah ILKOM, 2018	Untuk membantu Mengklasifikasi Masyarakat miskin di kecamatan Tibawa kabupaten Gorontalo	Berdasarkan hasil pengujian <i>confusion matrix</i> dengan teknik <i>split validasi</i> , penggunaan metode klasifikasi <i>naïve bayes</i> terhadap dataset yang telah diambil pada objek penelitian diperoleh tingkat akurasi sebesar 73% atau termasuk dalam kategori <i>Good</i> . Sementara nilai <i>Precision</i> sebesar 92% dan <i>Recall</i> sebesar 86%. Berdasarkan hal tersebut dapat dinyatakan bahwa sistem klasifikasi yang dibangun dapat digunakan sebagai bahan masukan bagi pengambil keputusan	Pada penelitian lain diharapkan dapat digunakan dataset dalam jumlah yang lebih besar atau dengan sejumlah variabel lainnya guna meningkatkan performa dari metode yang digunakan.	Obyek penelitian yang berbeda dan algoritma yang digunakan. Penelitian sebelumnya menggunakan 1 algoritma dan penelitian yang dilakukan saat ini menggunakan dua algoritma.

Tabel 2. 1 Matriks literatur review dan posisi penelitian (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
6	Comparison of Naïve Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification	Tomas PRANCKEVIČIUS, Virginijus MARCINKEVIČIUS, Baltic J. Modern Computing, 2017	Membandingkan pengklasifikasi ini dengan mengevaluasi akurasi klasifikasi, berdasarkan ukuran kumpulan data pelatihan, dan jumlah n-gram. Dalam eksperimen, teks singkat untuk data ulasan produk dari Amazon.	metode klasifikasi multi-kelas Regresi Logistik untuk ulasan produk telah mencapai akurasi klasifikasi tertinggi (min 32,43%, maks 58,50%) dibandingkan dengan metode klasifikasi Naïve Bayes, Random Forest, Decision Tree, dan Support Vector Machines. Sebaliknya, Decision Tree mendapatkan nilai akurasi rata-rata terendah (min dalam trigram: 24,10%, maks dalam uni/bi/trigram: 34,58%).	Investigasi menunjukkan bahwa meningkatkan ukuran kumpulan data pelatihan dari 5000 menjadi 75000 ulasan per kelas menyebabkan pertumbuhan akurasi klasifikasi yang tidak signifikan (1 - 2%) dari pengklasifikasi Naïve Bayes, Random Forest, dan Support Vector Machines. Hasil ini menunjukkan bahwa ukuran set pelatihan 3000 ulasan per kelas cukup untuk semua metode klasifikasi yang dianalisis, dan akurasi klasifikasi lebih terkait dengan properti n-gram.	Penelitian sebelumnya menggunakan beberapa algoritma dalam melakukan pengujian untuk hasil akurasi terbaiknya. Sedangkan penelitian ini menggunakan atau menerapkan perbandingan dua algoritma yaitu naïve bayes dan random forest dalam pengambilan keputusan untuk memprediksi tingkat Kesehatan lingkungan dalam program SDGs di Kabupaten Lombok Timur.

Tabel 2. 1 Matriks literatur review dan posisi penelitian (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
7	Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes	Mulkan Azhari, Zakaria Situmorang, Rika Rosnelly, Jurnal Media Informatika Budidarma, 2021	Membandingkan performance beberapa algoritma klasifikasi yaitu C4.5, Random Forest, SVM, dan naive bayes. Data penelitian berupa data peserta JISC yang berjumlah sebanyak 200 data	Akurasi algoritma C4.5, Random Forest, SVM dan Naive Bayes pada dataset kegiatan pramuka Jogja International Scout Camp (JISC) Provinsi Sumatera Utara tahun 2020 terdapat perbedaan. Pada algoritma C4.5 didapatkan akurasi sebesar 86,67%. Pada algoritma Random Forest didapatkan akurasi sebesar 83,33%. Pada algoritma SVM didapatkan akurasi sebesar 95%. Pada algoritma Naive Bayes didapatkan akurasi sebesar 86,67%.	Untuk mendapatkan nilai akurasi yang lebih baik maka jumlah dataset juga mempengaruhi hasil.	Penelitian sebelumnya hasil akurasi hanya digunakan untuk mengklasifikasikan kegiatan Jogja International Scout Camp (JISC) , sedangkan untuk penelitian ini mencoba mengklasifikasi tingkat Kesehatan lingkungan masyarakat dengan algoritma naive bayes dan random forest.

Tabel 2. 1 Matriks literatur review dan posisi penelitian (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
8	Klasifikasi Penerima Program Indonesia Pintar Menggunakan Algoritma Naive Bayes dan Random Forest	Andi Suandi, Gifhera Dwi Iestari, Nining R.	Untuk melakukan klasifikasi penerima Program Indonesia Pintar menggunakan teknik machine Learning dengan metode naive bayes dan random forest.	Dari penerapan dua algoritma yang digunakan pada data penerima program Indonesia pintar ternyata naive bayes memiliki performance terbaik dengan hasil akurasi 99,96%		Metode klasifikasi pada penelitian sebelumnya digunakan untuk data penerima program Indonesia Pintar dengan hasil akurasi tertinggi dimiliki oleh naive bayes, sedangkan pada penelitian ini dua algoritma tersebut digunakan untuk memprediksi tingkat kesejahteraan masyarakat nelayan dengan hasil akurasi tertinggi.
9	Perbandingan Klasifikasi SMS berbasis Support Vector Machine, Naive Bayes Classifier, Random Forest dan Begging Classifier	Devi Irawan, Eza Budi Perkasa, Yurindra, Delpiah Wahyuningsih, Ellya Helmud	Untuk menyelesaikan masalah SMS yang teridentifikasi spam yang sering terjadi saat ini. Sehingga perbandingan metode dapat menyaring dan memisahkan sms spam dan non spam.	Setelah dilakukan perbandingan metode maka metode Bagging Classifier mendapatkan performance tertinggi	Saran untuk penelitian selanjutnya dilakukan pengembangan untuk algoritma dan data.	Perbandingan klasifikasi digunakan untuk menyelesaikan masalah SMS yang teridentifikasi Spam dengan metode Bagging Classifier, untuk perbandingan atau komparasi algoritma pada penelitian ini dilakukan untuk mendapatkan hasil akurasi tertinggi yang dapat digunakan sebagai keputusan dalam memprediksi tingkat kesehatan lingkungan.

Tabel 2. 1 Matriks literatur review dan posisi penelitian (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
10	Perbandingan Naïve Bayes Dan Random Forest Dalam Klasifikasi Bahasa Dacrah	Gabriela Militia Momole, Evanga Mailoa	Tujuan peneliti melakukan penelitian untuk mengidentifikasi bahasa menggunakan metode naïve bayes dan random forest dari hasil identifikasi bahasa sesuai dengan teks dari bahasa Toraja, Kalimantan Dan Halmahera dengan menggunakan teknologi komputer machine learning untuk menghitung nilai accuracy dari kedua metode tersebut dan melakukan perbandingan metode yang paling baik untuk melakukan identifikasi bahasa	Hasil dari metode Naïve Bayes saat melakukan identifikasi bahasa sangat Baik karena menghasilkan nilai accuracy diatas 0,90 dibandingkan Random Forest hanya mendapatkan nilai accuracy dibawa 0,70. Dengan perhitungan confusion matrix, metode Naïve bayes lebih baik dengan nilai accuracy 0,9922 dibandingkan dengan hasil nilai accuracy random Forest 0,6544.		Perbandingan algoritma pada penelitian ini dilakukan untuk mengidentifikasi Bahasa dan ingin mendapatkan nilai terbaik berupa accuracy terbaik.

Tabel 2. 1 Matriks literatur review dan posisi penelitian (Lanjutan)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
11	Komparasi Algoritma Random Forest, Naïve Bayes dan K- Nearest Neighbor Dalam klasifikasi Data Penyakit Jantung	Amril Samosir, Ms Hasibuan, Wahyu Eko Justino, Tri Hariyono, Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat 2021 Institut Informatika dan Bisnis Darmajaya, 19 Agustus 2021 ISSN: 2598-0256, E-ISSN: 2598-0238	Tujuan penelitian ini melakukan perbandingan tiga algoritma dalam melakukan klasifikasi penyakit jantung	Hasil klasifikasi dengan algoritma Naïve Bayes memiliki rerata hasil akurasi sebesar 0,91 AUC, 0,84 CA, 0,84 F1, 0,839 Precision dan 0,84 Recall.	Penelitian lebih lanjut dapat mengembangkan klasifikasi kelas yang digunakan, agar menjadi rinci dalam mengklasifikasi penyakit jantung koroner berdasarkan level atau tingkatan tertentu.	Penelitian ini menggunakan tiga algoritma dan penelitian lebih lanjut juga dapat menambahkan algoritma klasifikasi lainnya sehingga mendapat hasil perbandingan yang lebih beragam.

2.3. Landasan Teori

Pada penelitian kali ini peneliti menggunakan metode klasifikasi yang digunakan untuk memprediksi tingkat kesehatan lingkungan di Kabupaten Lombok Timur. Metode klasifikasi sendiri adalah bagian dari Data Mining. Naive Bayes dan Random Forest adalah dua algoritma pembelajaran mesin yang berbeda dengan pendekatan yang berbeda dalam melakukan klasifikasi data. Karena itu, perbandingan antara kedua algoritma ini perlu dilakukan untuk menentukan algoritma mana yang paling cocok untuk digunakan dalam suatu kasus atau situasi tertentu. Dalam penelitian ini ada beberapa alasan kenapa perlu membandingkan Naive Bayes dan Random Forest adalah sebagai berikut:

Performa Kedua algoritma ini dapat memberikan performa klasifikasi yang baik pada data tertentu, namun salah satu algoritma dapat memberikan performa yang lebih baik tergantung pada data dan tujuan spesifik. Oleh karena itu, perlu dilakukan perbandingan untuk menentukan algoritma mana yang memberikan performa terbaik pada data yang diberikan.

Kebutuhan Komputasi Kedua algoritma memiliki perbedaan dalam kebutuhan komputasi. Naive Bayes cenderung lebih cepat dan lebih sederhana dalam perhitungan, sedangkan Random Forest membutuhkan waktu yang lebih lama dan lebih kompleks. Oleh karena itu, perlu dilakukan perbandingan dalam hal kebutuhan komputasi untuk menentukan algoritma mana yang lebih cocok untuk digunakan pada data yang tersedia.

Sifat Data Naive Bayes mengasumsikan bahwa semua fitur data adalah independen, sedangkan Random Forest tidak memiliki asumsi ini. Oleh karena itu, perlu dilakukan perbandingan untuk menentukan algoritma mana yang lebih cocok digunakan pada jenis data tertentu.

Dalam penelitian ini proses komparasi algoritma antara naïve bayes dan random forest akan memperhatikan beberapa aspek antara lain performa kedua algoritma dapat dibandingkan dalam hal performa, yaitu akurasi, presisi, recall, F1-score, dan lainnya. Perbandingan ini dilakukan dengan membandingkan performa algoritma pada data training dan data testing yang berbeda-beda.

Data mining adalah suatu rangkaian atau tindakan untuk menemukan hubungan yang memiliki arti melalui pola dan kecenderungan dalam himpunan besar data yang tersimpan dengan menggunakan metode.(Ucha Putri et al., 2021)

Data mining memiliki teknik dan sifat diantaranya :

- a. Classification (Predictive)
- b. Clustering (Descriptive)
- c. AssociationRule Discovery (Descriptive)
- d. Regression (Predictive)
- e. Deviation Detection (Predictive) (Eskal, n.d.)

Klasifikasi merupakan salah satu tugas yang penting data mining. Sebuah pengklasifikasian dibuat dari sekumpulan data latih dengan kelas yang telah di tentukan (Wibawa et al., 2018) untuk menciptakan model yang dapat mendeskripsikan dan dijadikan sebagai pembelajaran.

Proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui merupakan suatu klasifikasi. (Annur, 2018)

Algoritma pemodelan klasifikasi random forest adalah pohon klasifikasi (*classification tree*) yang dapat digunakan untuk mengklasifikasi data dengan variabel respon multikelas. (Ramadhan et al., 2019). Dimana kumpulan pohon keputusan dilatih menggunakan metode bagging untuk mendapatkan hasil keseluruhan.

Dalam algoritma machine learning seperti Random Forest output yang dihasilkan dari metode klasifikasi dibagi menjadi empat yaitu : *True Positive (TP)*, *True Negative (TN)*, *False Negative (FN)*, dan yang terakhir adalah *False Positive (FP)*. Dalam machine learning tingkat kesalahan dari pengklasifikasian diberi simbol E yang merupakan frekuensi dari tingkat *error*. Tingkat kesalahan tersebut dapat dihitung menggunakan persamaan (1).

$$E = \frac{FP+FN}{FP+FN+TP+TN} \quad (1)$$

Selain tingkat kesalahan, hasil akurasi yang merupakan frekuensi dari klasifikasi dataset juga dapat dihitung menggunakan persamaan (2).

$$E = \frac{TP+TN}{FP+FN+TP+TN} \quad (2)$$

Dan persamaan yang ke (3) adalah presisi yang merupakan hasil persentase dari *True Positif* dari semua dataset yang mendapatkan label positif

$$Pr = \frac{TP}{FP+TP} \quad (3)$$

Karakteristik dari random forest adalah sebagai berikut :

1. Ensemble Learning: Random Forest adalah contoh dari ensemble learning, di mana model dibuat dari beberapa model yang lebih kecil. Random Forest menggabungkan beberapa pohon keputusan yang dibuat pada subset acak dari data pelatihan untuk menghasilkan prediksi yang lebih akurat.
2. Pohon Keputusan: Random Forest terdiri dari sejumlah pohon keputusan. Setiap pohon dibuat dengan cara yang acak dan independen dari pohon lainnya. Setiap pohon melakukan klasifikasi dengan cara yang sama dengan pohon keputusan biasa, yaitu dengan membagi ruang fitur menjadi beberapa bagian hingga terbentuk daerah yang homogen.
3. Sampling Acak: Random Forest melakukan sampling acak dengan penggantian dari data pelatihan untuk setiap pohon keputusan yang dibuat. Hal ini memastikan bahwa setiap pohon keputusan dibuat dengan data yang berbeda dari data latihan yang tersedia.
4. Feature Sampling Acak: Random Forest melakukan feature sampling acak pada setiap pohon keputusan dengan memilih subset acak dari fitur-fitur yang tersedia. Ini memastikan bahwa setiap pohon keputusan melihat sejumlah fitur yang berbeda dan mencegah overfitting pada satu fitur.
5. Scalability: Random Forest dapat digunakan untuk mengolah data besar, karena proses pelatihan dapat dijalankan secara parallel pada beberapa CPU atau komputer.
6. Tidak Sensitif terhadap Outlier: Random Forest tidak sensitif terhadap outlier dalam data, karena pada setiap pohon keputusan hanya subset acak dari data yang dipertimbangkan.

7. Interpretability: Hasil dari Random Forest cukup mudah untuk diinterpretasikan. Kita dapat mengekstrak fitur-fitur yang paling penting dan melihat bagaimana setiap pohon keputusan berkontribusi pada prediksi akhir.

Selain Random Forest algoritma yang kedua adalah Naïve Bayes. Algoritma Naïve Bayes adalah sebuah teknik machine learning yang didasari oleh teori probabilitas yang berhubungan dengan probabilitas sebelumnya, sehingga dapat memberikan keputusan secara matematis dan menciptakan sebuah model yang sederhana dan efisien (Models et al., 2018).

Nilai probabilitas dalam teorema Bayes seperti pada persamaan 1:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (1)$$

Dimana

$P(H)$: Nilai probabilitas dari prior dari hipotesis sampel

$P(X)$: Evidence dari probabilitas data latih

$P(H|X)$: Nilai probabilitas H yang mempengaruhi X

$P(X|H)$: Probabilitas X kepada H atau dapat disebut dengan likelihood.

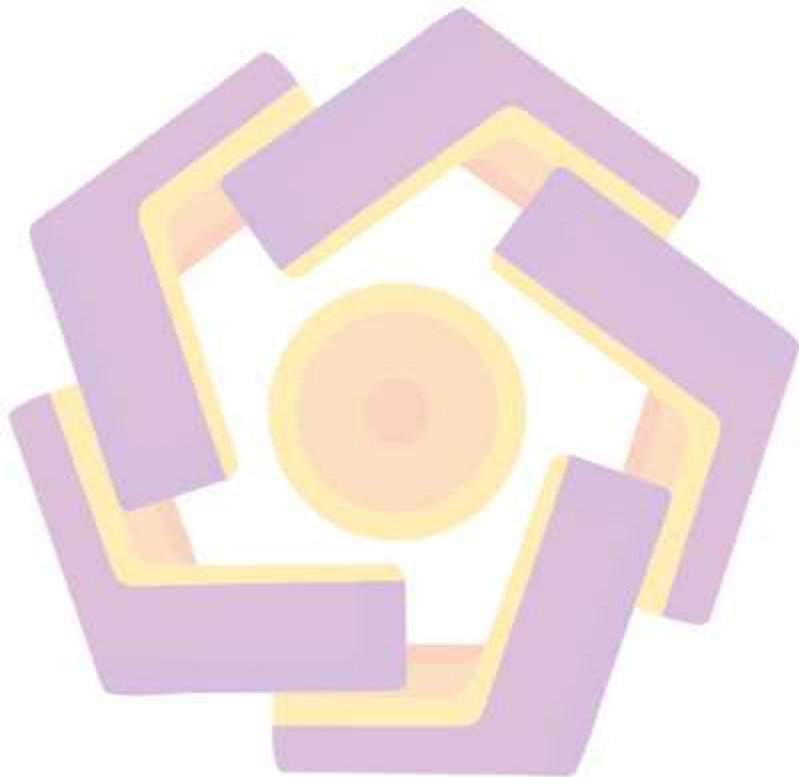
Teorema Naïve Bayes diatas digunakan untuk menentukan kelas yang sesuai dengan sampel data yang dianalisis (Putri & Wijayanto, 2022)

Dalam metode klasifikasi ada sebuah metode yang digunakan untuk mengukur kinerja metode klasifikasi itu yaitu Confusion matrix. Dimana Confusion matrix adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah (Normawati & Prayogi, 2021)

Adapun karakteristik dari algoritma naïve bayes adalah sebagai berikut :

1. Probabilistik: Naive Bayes adalah algoritma probabilistik yang menggunakan teori probabilitas untuk mempelajari hubungan antara fitur-fitur dalam data dan kelas yang diinginkan. Algoritma ini menghitung probabilitas kemunculan setiap fitur untuk setiap kelas, dan menggunakannya untuk memprediksi kelas yang paling mungkin.
2. Naive Assumption: Naive Bayes mengasumsikan bahwa setiap fitur dalam data independen satu sama lain, artinya setiap fitur hanya dipengaruhi oleh kelas yang diinginkan. Meskipun asumsi ini sering tidak benar dalam praktiknya, Naive Bayes masih sering digunakan karena kesederhanaan dan kemampuan untuk menghasilkan prediksi dengan cepat.
3. Mudah diimplementasikan: Algoritma Naive Bayes relatif mudah diimplementasikan dan membutuhkan sedikit waktu pelatihan. Ini membuatnya menjadi algoritma yang cocok digunakan pada data yang sangat besar dan aplikasi real-time.
4. Klasifikasi Multikelas: Algoritma Naive Bayes dapat dengan mudah diadaptasi untuk mengklasifikasikan data dengan lebih dari dua kelas. Hal ini memungkinkan Naive Bayes untuk digunakan pada berbagai macam masalah klasifikasi, termasuk teks klasifikasi, analisis sentimen, dan klasifikasi gambar.
5. Tahan terhadap data yang hilang: Naive Bayes dapat bekerja dengan baik meskipun terdapat data yang hilang. Ini dikarenakan algoritma ini dapat menangani data yang tidak lengkap dengan mempertimbangkan semua data yang tersedia.

6. Tidak membutuhkan parameter yang kompleks: Naive Bayes memiliki sedikit parameter yang perlu disetel. Karena sifatnya yang sederhana, Naive Bayes mudah diimplementasikan dan tidak membutuhkan banyak waktu untuk dilatih.



BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

Jenis penelitian yang digunakan oleh peneliti adalah penelitian kuantitatif, dimana penelitian yang dilakukan adalah sebuah proses perhitungan matematis agar mendapatkan atau menemukan hasil yang diinginkan.

Sifat dari penelitian yang dilakukan adalah eksperimental dimana melakukan sebuah eksperimen penelitian untuk mendapatkan hasil akurasi atau parameter dengan melakukan komparasi algoritma Naïve Bayes dan Random Forest pada dataset yang memiliki data yang cukup besar.

Pendekatan penelitian ini adalah pendekatan kuantitatif dimana peneliti akan melakukan penelitian sesuai dengan tahap-tahap atau alur penelitian yang telah dibuat.

3.2. Metode Pengumpulan Data

Data didapat dari program KKN Tematik Desa Gemilang yang dilakukan oleh universitas hamzanwadi yang dibuat khusus, dan data termasuk data private yang belum digunakan dalam penelitian-penelitian sebelumnya. Data diperoleh dengan cara observasi, wawancara dengan penyebaran angket kepada penduduk.

1. Observasi

Salah satu teknik pengumpulan data yang dilakukan dengan pengamatan langsung ke lokasi penelitian yang berada di beberapa wilayah Kecamatan di Kabupaten Lombok Timur

2. Wawancara

Wawancara dilakukan di lokasi penelitian dengan masyarakat di beberapa wilayah Kecamatan di Kabupaten Lombok Timur.

3. Angket

Pengisian angket dilakukan untuk mempermudah saat dilakukan wawancara kepada masyarakat di sepuluh kecamatan di kabupaten Lombok Timur dengan total 4624 responden dengan rincian form angket yang diambil dari aplikasi khusus SDGs terdiri dari nama lengkap, jenis kelamin, umur, Pendidikan terakhir, tempat sekolah, putus sekolah, provinsi, kabupaten, kecamatan, desa, tempat membuang sampah, pengelolaan sampah, mengelola sampah menjadi bermanfaat, pemanfaatan sampah, dampak sampah, dampak yang dirasakan, mempunyai jamban, kondisi jamban, MCK umum, kondisi MCK umum, Pembuangan air rumah tangga, sumber mata air, kondisi sumber mata air, menebang pohon illegal, pekarangan rumah, pemanfaatan pekarangan rumah, penghijauan, hewan ternak, jenis hewan ternak, status hewan ternak, tempat menaruh ternak, kondisi kandang kolektif, tidak menaruh di kandang kolektif, tidak memiliki kandang kolektif, penambangan galian c, penambangan dapat merusak lingkungan.

3.3. Metode Analisis Data

Metode analisis data untuk penelitian ini adalah kuantitatif, metode ini digunakan karena sesuai dengan penelitian yang dilakukan. Data diperoleh dari kegiatan KKN Tematik desa gemilang Universitas Hamzanwadi.

Adapun instrumen penelitian yang digunakan adalah angket dan wawancara yang dilakukan di lokasi penelitian. Karakteristik dari dataset yang diperoleh adalah sebagai berikut : Dataset SDGs lingkungan terdiri dari banyak variabel atau fitur yang menggambarkan aspek-aspek berbeda dari masalah lingkungan yang relevan dengan tujuan pembangunan berkelanjutan. Variabel ini bisa mencakup data tentang kualitas udara, ketersediaan air bersih, keanekaragaman hayati, atau perubahan lingkungan dari tahun ketahun dan lain sebagainya.

Algoritma Naïve Bayes digunakan dalam penelitian karena ini adalah salah satu metode klasifikasi data yang dapat digunakan untuk memprediksi probabilitas suatu class. Keuntungan dari metode pengklasifikasian menggunakan naïve bayes adalah tidak memerlukan banyak data pelatihan untuk memprediksi parameter. Dengan kata lain nilai output diberikan berdasarkan kepada asumsi penyederhanaan nilai atribut secara kondisional.

Sedangkan Random Forest digunakan dalam penelitian ini dikarenakan algoritma ini dapat menganalisis data yang cukup banyak dengan melakukan pengambilan acak contoh yang dianalisis.

Untuk menentukan status sebelum menggunakan model klasifikasi, ada beberapa tahapan yang perlu dijelaskan di antaranya sebagai berikut :

1. Persiapan data

Pengumpulan data yang relevan dan berkualitas tinggi untuk masalah yang ingin diselesaikan atau model yang ingin dibangun. Data ini mencakup berbagai fitur atau atribut yang relevan untuk memprediksi atau mengklasifikasikan status yang ingin ditentukan pada tahap ini dilakukan pengumpulan data dalam bentuk format excel. Data yang didapat dari kegiatan KKN Tematik desa gemilang Universitas Hamzanwadi dipisahkan menurut Desa, Kecamatan, untuk memudahkan dan membuat lebih sederhana proses penelitian semua data digabungkan terlebih dahulu lalu data yang awalnya menggunakan ekstensi XLS dikonversikan ke bentuk CSV.

2. Penanganan data yang hilang

Dalam tahap ini missing value dilakukan dalam tahap preprocessing, karena nilai yang hilang dalam suatu data dapat mempengaruhi hasil akurasi, penanganan ini dilakukan di bahasa pemrograman python, proses tersebut dilakukan apabila ada yang kosong maka di isi dengan angka 0 (nol).

3. Menormalisasikan data

Normalisasi data adalah proses mengatur atau mentransformasikan data dalam sebuah database atau dataset sehingga data tersebut memenuhi kriteria tertentu, seperti menghilangkan redundansi, meminimalkan kerentanan terhadap kesalahan, atau memastikan keakuratan data. Tujuan utama normalisasi data adalah untuk memastikan basis data memiliki struktur yang efisien dan konsisten, sehingga memungkinkan untuk melakukan operasi penyimpanan, pencarian, dan manipulasi data dengan lebih efektif dalam penelitian Data SDGs Lingkungan di sepuluh

kecamatan di Kabupaten Lombok Timur data yang bersifat karakter di ubah menjadi numerik agar data dapat dibaca oleh model yang di usulkan.

4. Seleksi fitur.

Tahap ini dilakukan untuk pemilihan fitur dan atribut target yang akan digunakan, pemilihan fitur sangat penting karena dapat mempengaruhi hasil akurasi. Dalam melakukan seleksi fitur data SDGs Lingkungan dilakukan menjadi 26 variabel.

5. Klasifikasi data

Pada tahap ini dilakukan proses klasifikasi atau pengelompokan data agar memudahkan dalam melihat laporan

6. Distribusi data tes dan data latih

Tahap selanjutnya data dibagi menjadi data latih dan data uji. Data latih digunakan untuk pemodelan dan data uji dilakukan untuk menguji kinerja dan kebenaran dalam model yang digunakan.

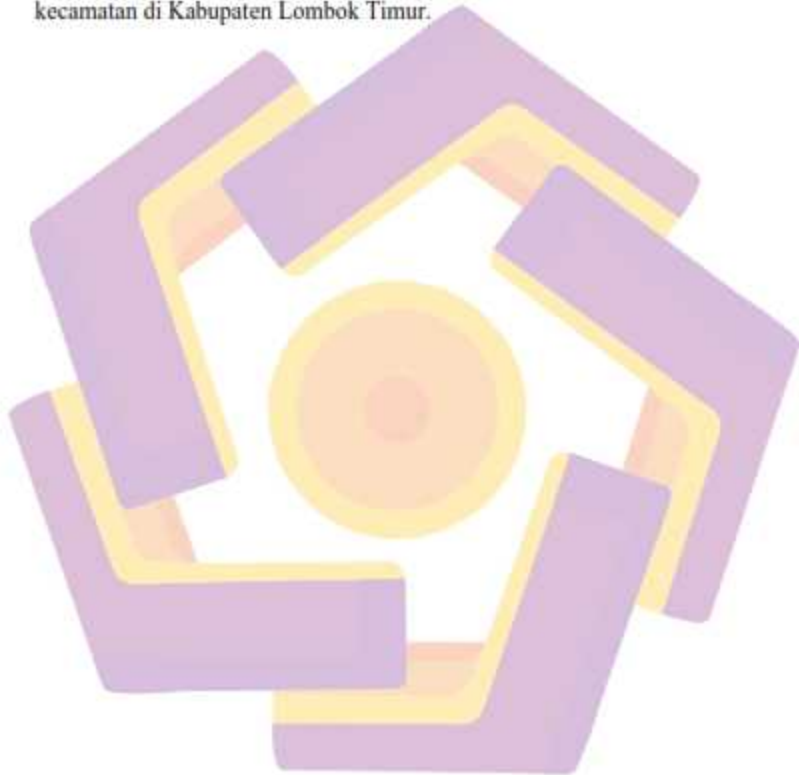
7. Naïve Bayes dan Random Forest Model

Dua algoritma ini digunakan karena menurut peneliti sesuai dengan dataset yang dimiliki dan juga metode terbaik dalam klasifikasi. Hasil pembagian dari dataset dan data uji yang telah dibuat akan dimodelkan menggunakan dua algoritma yaitu Naïve Bayes dan Random Forest.

8. Evaluasi

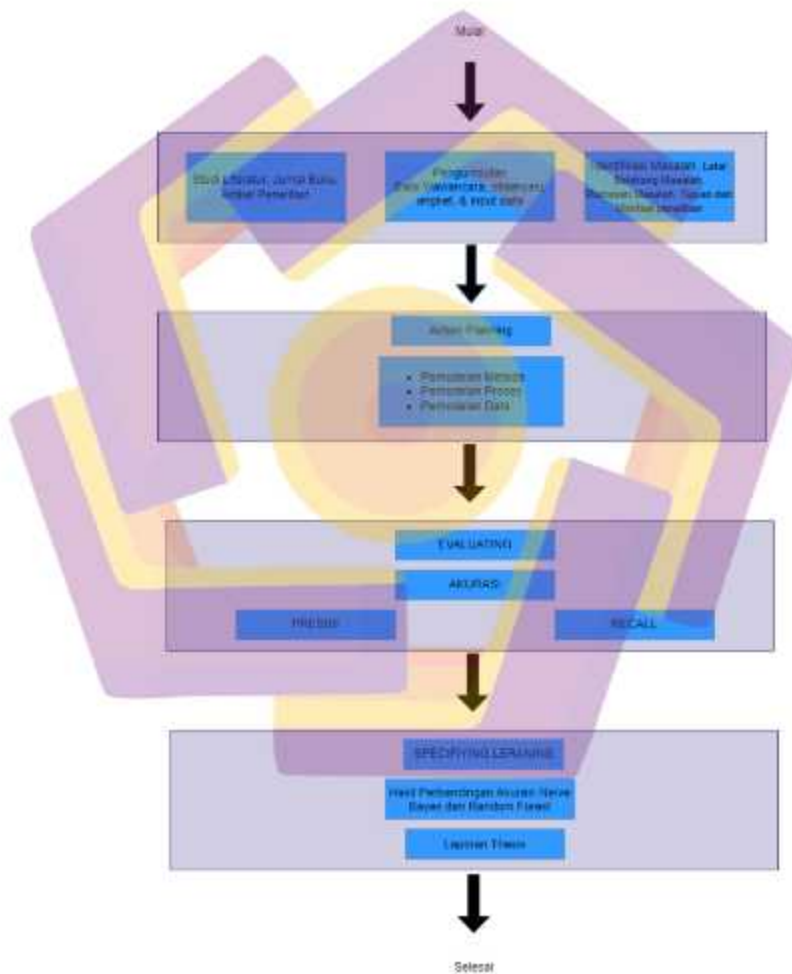
Hasil akhir adalah evaluasi, tahap ini dilakukan untuk mendapatkan hasil akurasi yang terbaik dari dua algoritma yang diterapkan.

Dengan melakukan langkah-langkah di atas sebelum menggunakan model yang diajukan untuk memastikan bahwa model yang dibangun dan dievaluasi dengan baik, memiliki pemahaman yang baik tentang bagaimana model tersebut seharusnya bekerja dalam mengevaluasi data SDGs Lingkungan di sepuluh kecamatan di Kabupaten Lombok Timur.



3.4. Alur Penelitian

Alur Penelitian ini menggunakan dua metode yaitu Naïve Bayes dan Random Forest. Adapun alur penelitian dapat dilihat pada Gambar 3.1.



Gambar 3.1 Alur Penelitian

Pada Gambar 3.1 dijelaskan bahwa pada penelitian dilakukan dalam empat tahapan yaitu:

1. Diagnosing

Pada tahap ini dilakukan studi literatur dengan berbagai macam artikel dari berbagai macam jurnal dan buku yang sesuai dengan permasalahan atau topik penelitian. Kemudian dilakukan pengumpulan data dengan cara wawancara dan pengisian angket ke masyarakat dilokasi penelitian. Setelah itu dilakukan identifikasi masalah, dengan menentukan latar belakang, rumusan masalah, manfaat dan tujuan dari penelitian yang akan dilakukan.

2. Action Planning

Pada tahap ini dilakukan proses metode data, pemodelan data menggunakan dua algoritma yaitu Naïve Bayes dan Random Forest, Pemodelan proses menggunakan python.

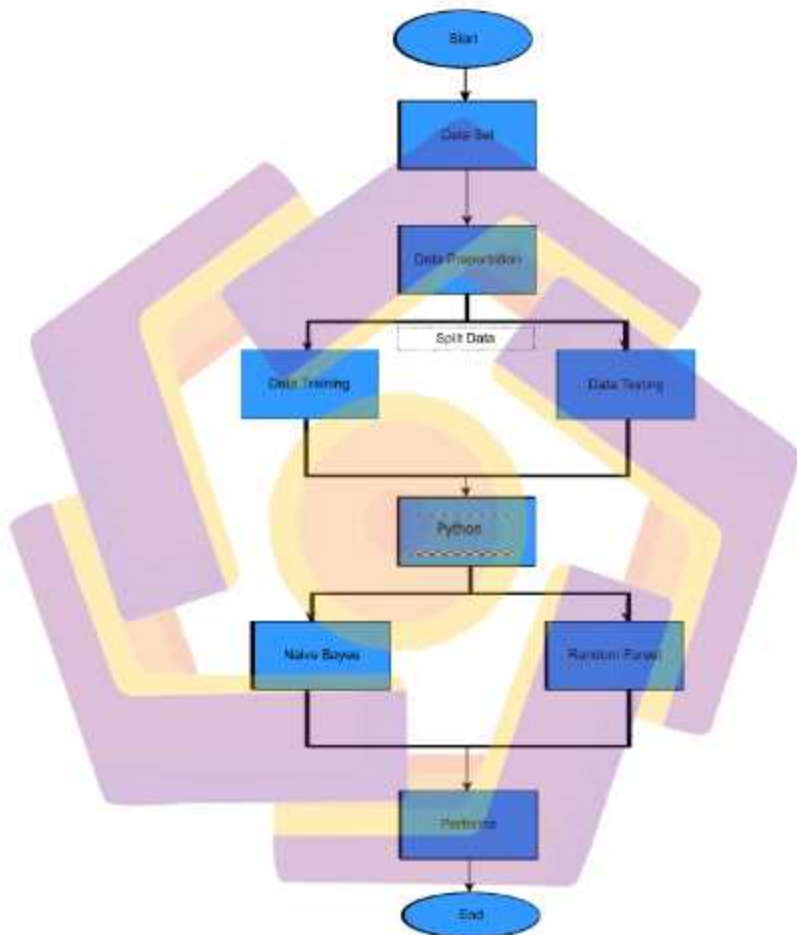
3. Evaluating

Pada tahap ini dilakukan penerapan untuk mendapatkan hasil dari penelitian dengan mengukur tingkat akurasi precision, recall dan F1 Score yang akan disajikan dalam bentuk tabel, nilai angka dan juga diagram.

4. Specifying Learning

Pada tahap ini dilakukan perbandingan hasil akurasi dari dua algoritma yang digunakan untuk pengambilan keputusan dalam memprediksi dan juga sebagai hasil dari penelitian yang sudah diterapkan (Journal et al., 2019).

Alur proses penelitian dengan dua algoritma yaitu naïve Bayes dan Random Forest dapat dilihat pada Gambar 2 dibawah ini.



Gambar 3. 2 Alur Proses Modeling data

Tahapan-tahapan prosesnya dimulai dengan dataset yang diambil dari wawancara ketika melakukan KKN Tematik desa gemilang yang dilakukan oleh

Universitas Hamzanwadi kepada masyarakat di sepuluh kecamatan di Lombok Timur dengan format .xls dan .csv, kemudian data preparation adalah merubah dataset yang belum diolah menjadi data yang dapat dianalisis disini dataset dibagi menjadi data testing dan data training atau splitdata.

Langkah selanjutnya adalah dengan penerapan algoritma yang digunakan yaitu Naïve Bayes dan Random Forest menggunakan rapidminer, dimana penginputan data dalam bentuk .xls, kemudian pemilihan atribut pendukung untuk target atau output yang diinginkan, lalu dilakukan validasi data dengan k-fold validation sehingga pengujian akan mendapatkan nilai akurasi, precision dan recall dan hasil klasifikasi dapat dilihat menggunakan confusion matrix.



BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

Algoritma yang digunakan untuk menentukan hasil dari penelitian ini adalah algoritma Naïve Bayes dan Random Forest. Dua algoritma ini digunakan untuk mengkomparasi atau membandingkan nilai prediksi, presisi, akurasi, recall, F1 score tertinggi agar dapat digunakan untuk memprediksi dataset yang digunakan dalam penelitian ini. Dataset yang digunakan adalah dataset SDGs Kesehatan Lingkungan di sepuluh Kecamatan yang berada di Kabupaten Lombok Timur Nusa Tenggara Barat.

4.1. Diagnosa Penelitian

Untuk tahap diagnosa penelitian ini dilakukan studi literatur yang diperoleh dari berbagai macam artikel, jurnal dan juga buku-buku yang sesuai dengan permasalahan atau topik penelitian yang dibahas. Kemudian dilakukan identifikasi masalah, dengan menentukan latar belakang penelitian yaitu memprediksi tingkat kesehatan lingkungan masyarakat dalam Program Sustainable Development Goals di Kabupaten Lombok Timur Nusa Tenggara Barat. Dataset yang belum pernah digunakan/diolah, berdasarkan dataset yang diolah menjadi sebuah informasi yang lebih bermanfaat. Kemudian melakukan perumusan masalah, serta menentukan manfaat dan tujuan dari penelitian.

Pada tahap ini juga dilakukan pengumpulan data dengan cara wawancara dan pengisian angket ke masyarakat dilokasi penelitian. Dengan jumlah data 4624 untuk sample dari dataset dapat dilihat pada Tabel 4.1.

Tabel 4. 1 Data Masyarakat SDGs Lingkungan 10 Teratas

ID	NAMA	JK	TMS	MJ	PART	KMA	PENAMBANGAN	STATUS
1	SUPARDI	Laki - Laki	Parit dekat rumah	Ya, saya punya	Pembuangan khusus yang dibuat sendiri	Jernih, tidak berbau dan rasanya enak	Tidak tahu	1
2	MUHLTAUFIK NASRULLAH	Laki - Laki	Sungai	Ya, saya punya	Parit dekat rumah	Jernih, tidak berbau dan rasanya enak	Ya saya tahu	0
3	SIDIN	Laki - Laki	Tempat pembuangan sampah (TPS) yang sudah disediakan	Ya, saya punya	Parit dekat rumah	Jernih, tidak berbau dan rasanya enak	Ya saya tahu	1
4	M.HUSNAI HADI S.H	Laki - Laki	Parit dekat rumah	Tidak punya	Parit dekat rumah	Jernih, tidak berbau dan rasanya enak	Ya saya tahu	0
5	BAHRUL FAHMI	Laki - Laki	Sungai	Ya, saya punya	Kali	Jernih, tidak berbau dan rasanya enak	Ya saya tahu	1
6	HASYIM	Laki - Laki	Parit dekat rumah	Ya, saya punya	Parit dekat rumah	Jernih, tidak berbau dan rasanya enak	Tidak tahu	1
7	HUMAIDI	Laki - Laki	Parit dekat rumah	Ya, saya punya	Parit dekat rumah	Jernih, tidak berbau dan rasanya enak	Ya saya tahu	1
8	RAFIQAH	Perempuan	Tempat pembuangan sampah (TPS) yang sudah disediakan	Ya, saya punya	Pembuangan khusus yang dibuat sendiri	Jernih, tidak berbau dan rasanya enak	Ya saya tahu	1
9	MAHSAN	Laki - Laki	Tempat pembuangan sampah (TPS) yang sudah disediakan	Ya, saya punya	Parit dekat rumah	Jernih, tidak berbau dan rasanya enak	Tidak tahu	1
10	ABDUL QADIR JAELANI	Laki - Laki	Tempat pembuangan sampah (TPS) yang sudah disediakan	Ya, saya punya	Pembuangan khusus yang dibuat sendiri	Jernih, tidak berbau dan rasanya enak	Tidak tahu	1

Penelitian menggunakan bahasa pemrograman python, untuk hasil dari dataset yang digunakan dapat dilihat pada Gambar 4.1

berupa Gambar sample dataset masyarakat Lombok Timur yang menerima angket SDGs Lingkungan.

id	nama	jk	umur	pendidikan	sekolah	prov	kab	kec
0	1	Supardi	Laki - Laki	55	SMP Swasta	KAB. LOMBOK TIMUR	Selong	...
1	2	Muh.taufik Nasrullah	Laki - Laki	47	SMP Negri	KAB. LOMBOK TIMUR	Selong	...
2	3	SIDIN	Laki - Laki	69	SD Negri	KAB. LOMBOK TIMUR	LABUAN HATI	...
3	4	M.husnai hadi S.H	Laki - Laki	53	SI Negri	KAB. LOMBOK TIMUR	Selong	...
4	5	Bahrul Fahmi	Laki - Laki	41	SI Negri	KAB. LOMBOK TIMUR	Selong	...
...
4619	4620	AQ DURRIATI	Laki - Laki	41	SMA Negri	KAB. LOMBOK TIMUR	MASBAGIK	...
4620	4621	YULIA ASTUTI	Perempuan	30	SMA Negri	KAB. LOMBOK TIMUR	MASBAGIK	...
4621	4622	HASHAMATI	Perempuan	48	SMP Negri	KAB. LOMBOK TIMUR	MASBAGIK	...
4622	4623	JAUMARIAH	Perempuan	31	OS Negri	KAB. LOMBOK TIMUR	MASBAGIK	...
4623	4624	REHAN	Perempuan	36	SMP Negri	KAB. LOMBOK TIMUR	MASBAGIK	...

id	putus sekolah	prov	kab	kec
0	Ya	Nusa Tenggara Barat	KAB. LOMBOK TIMUR	Selong
1	Ya	Nusa Tenggara Barat	KAB. LOMBOK TIMUR	Selong
2	Tidak	Nusa Tenggara Barat	KAB. LOMBOK TIMUR	LABUAN HATI
3	Tidak	Nusa Tenggara Barat	KAB. LOMBOK TIMUR	Selong
4	Tidak	Nusa Tenggara Barat	KAB. LOMBOK TIMUR	Selong
...
4619	Ya	Nusa Tenggara Barat	KAB. LOMBOK TIMUR	MASBAGIK
4620	Ya	Nusa Tenggara Barat	KAB. LOMBOK TIMUR	MASBAGIK
4621	Ya	Nusa Tenggara Barat	KAB. LOMBOK TIMUR	MASBAGIK
4622	Tidak	Nusa Tenggara Barat	KAB. LOMBOK TIMUR	MASBAGIK
4623	Ya	Nusa Tenggara Barat	KAB. LOMBOK TIMUR	MASBAGIK

id	hewan ternak	jenis hewan ternak	status hewan ternak
0	Tidak	-	-
1	Tidak	-	-
2	Tidak	-	-
3	Ya	Lainnya	Ya, itu milik saya
4	Tidak	-	-
...
4619	Tidak	-	-
4620	Tidak	-	-
4621	-	-	-
4622	Ya	Lainnya	Ya, itu milik saya
4623	Ya	Sapi	-

id	tempat menaruh hewan ternak	kondisi kandang kolektif
0	-	-
1	-	-
2	-	-
3	Kandang milik sendiri	Berfungsi dengan baik
4	-	-
...
4619	-	-
4620	-	-
4621	-	-
4622	Kandang milik sendiri	-
4623	-	-

Gambar 4. 1 Sample Dataset Masyarakat Yang Memperoleh Angket SDGs Lingkungan

tidak menaruh di kandang kolektif \		tidak memiliki kandang kolektif penambangan galian c \	
0	-	0	Tidak ada
1	-	1	-
2	-	2	Tidak ada
3	karena kami tidak memiliki kandang kolektif	3	Tidak diprogramkan oleh desa
4	-	4	Tidak ada
...
4619	-	4619	-
4620	-	4620	Tidak ada
4621	-	4621	-
4622	karena dikandang milik sendiri lebih nyaman da...	4622	Tidak diprogramkan oleh desa
4623	-	4623	-

penambangan dapat merusak lingkungan label		
0	tidak tahu	1
1	-	0
2	Ya saya tahu	1
3	Ya saya tahu	0
4	Ya saya tahu	1
...
4619	-	0
4620	Tidak tahu	1
4621	-	0
4622	Tidak tahu	1
4623	-	0

[4624 rows x 33 columns]

Gambar 4. 1 Sample Dataset Masyarakat Yang Memperoleh Angket SDGs Lingkungan (Lanjutan)

id	nama	jk	umur	pendidikan	sekolah	jenis kelamin	prov	kecamatan	kecamatan	kecamatan	jenis kelamin ternak	status ternak ternak	tempat ternak ternak ternak	memiliki kandang kolektif	tidak memiliki kandang kolektif	tidak memiliki kandang kolektif	penyakit ternak	penyakit ternak	penyakit ternak	infeksi
0	1	Supendi	Laki	65	SMP	Suwata	Ya	Tenggara	Nusa Tenggara Barat	KAB. LOMBOK TIMUR	Selang	Tidak	-	-	-	-	Tidak ada	Tidak tahu	1	
1	2	Muh tufts laaculan	Laki	47	SMP	Ngali	Ya	Tenggara	Nusa Tenggara Barat	KAB. LOMBOK TIMUR	Selang	Tidak	-	-	-	-	Tidak ada	Tidak tahu	0	
2	3	SDIN	Laki	69	SD	Ngali	Tidak	Tenggara	Nusa Tenggara Barat	KAB. LOMBOK TIMUR	LABUAN HARU	Tidak	-	-	-	-	Tidak ada	Ya saya tahu	1	
3	4	M huanu kat SH	Laki	62	ST	Ngali	Tidak	Tenggara	Nusa Tenggara Barat	KAB. LOMBOK TIMUR	Selang	Ya	Ya, lahnye mla saya	Kandang mla sendiri	Berfungsi sangat baik	Karena kami tidak memiliki kandang kolektif	Tidak diprogramkan oleh desa	Tidak ada	Ya saya tahu	0
4	5	Bahul Fahru	Laki	41	ST	Ngali	Tidak	Tenggara	Nusa Tenggara Barat	KAB. LOMBOK TIMUR	Selang	Tidak	-	-	-	-	Tidak ada	Ya saya tahu	1	

Gambar 4. 2 Dataset Sebelum dilakukan Normalisasi

4.2. Penyusunan Perencanaan

Pada tahapan ini dilakukan penyusunan rencana penelitian yang sesuai agar tepat dalam menyelesaikan masalah. Untuk penelitian ini yang dilakukan adalah dengan pemodelan data, pemodelan metode dan pemodelan proses. Untuk pemodelan proses penelitian menggunakan python.

4.3. Menentukan parameter untuk kelas

Tahap penelitian berikutnya adalah menentukan parameter untuk kelas. Dalam menentukan parameter untuk kelas pada pemrograman python dapat dilihat pada code di bawah ini.

```
# input parameter untuk prediksi class
xTarget = df.drop(['id', 'label'], axis = 1)
print(xTarget)
```

Dataset masyarakat yang sudah mendapatkan angket dalam Program Sustainable Development Goals (SDGs) dalam bidang Kesehatan Lingkungan Masyarakat berjumlah 4624 data dan data tersebut akan dibagi menjadi dua bagian yaitu data training dan data testing. Pembagian data tersebut terdiri dari 80 % data *training* dan 20 % data testing. Pembagian data atau *split* data ini digunakan untuk melatih model, data latih ini dijadikan model untuk dapat memperbaharui parameter ada saat *training*. Pembagian 80 - 20 ini dilakukan karena sesuai dengan jumlah dataset, Adapun code untuk *split* data sebagai berikut dapat di lihat pada Gambar 4.3 di bawah ini.

```

#membuat data training dan data testing dari dataset dengan data test = 20% dan training 80%
#dari jumlah instance
X_train, X_test, y_train, y_test=train_test_split(X,Y,test_size = 0.2, random_state=42)
print(X_train.shape)
print(y_train.shape)
print(X_test.shape)
print(y_test.shape)

(3699, 84)
(3699, 1)
(925, 84)
(925, 1)

```

Gambar 4. 3 Kode data traintest 80 : 20 dengan random state 42.

4.4. Pemodelan Metode

Untuk mendapatkan output yang diinginkan maka langkah berikutnya adalah pemodelan metode untuk penelitian, yaitu dengan menentukan atribut pendukung keputusan. Model metode untuk penelitian ini adalah klasifikasi, model ini dipilih agar pada saat proses menganalisa dapat tercipta model-model yang dapat mendeskripsikan kelas pada dataset selain itu juga metode ini dipilih karena sesuai dengan dataset. Kelas yang diinginkan adalah sehat dan tidak sehat lingkungan tersebut, berikut kode dari penentuan target atau output yang dihasilkan:

```

#target classification dari accept / rejected Loan
yTarget = df['label']

print (yTarget)

```

Output

```

0    1 Sehat
1    0 Tidak Sehat
2    1 Sehat
3    0 Tidak Sehat
4    1 Sehat
...
4619 0 Tidak Sehat

```

```

4620 1 Sehat
4621 0 Tidak Sehat
4622 1 Sehat
4623 0 Tidak Sehat
Name: label, Length: 4624, dtype: int64

```

4.5. PraProses Data

Jenis data yang digunakan dalam penelitian ini adalah jenis data kontinyu atau data numerik namun semua data atribut memiliki data bertipe data teks atau string. Supaya bisa dikenali oleh system atau baris kode yang di pakai untuk melakukan perhitungan pada perangkat lunak, dilakukannya perubahan dari berupa kategori atau teks menjadi numerik. Perubahan nilai dari sebuah atribut yang bernilai string atau teks diubah menjadi nilai angka.

a. Missing Value

Pada Tahap ini yang dilakukan adalah mencari atribut yang memiliki nilai kosong atau jauh dari data, data tersebut akan dilakukan penghapusan secara manual atau pemberian nilai 0. Hal ini bertujuan supaya dataset bisa dibaca oleh perangkat lunak yang di pakai pada proses perhitungan. Dalam penelitian ini proses missing value menggunakan perintah `df.isna().sum()`. gambar kode perintah tersebut dapat dilihat pada gambar 4.4

```
df.isna().sum()
```

Gambar 4. 4 Pencarian data Missing value

	tempat membuang sampah	pengelolaan sampah	mengelola sampah menjadi bermanfaat	pernikfaatan sampah	daurugt sampah	barang yang dihasilkan	mempunya jenbar	kondisi jenbar
Pencapaian rumah	Tidak tahu	Tidak pernah	Tidak tahu	Tidak tahu	Melakukan pemertanganan	Ya, saya punya	-	Darfung dengan baik
Surga	Tidak tahu	Tidak pernah	Tidak tahu	Ya, saya punya	Kemungkinan jumlah hasil daurugt ini tidak akan terlalu banyak	Ya, saya punya	-	Darfung dengan baik
Sebelum pemertanganan sampah (TMS) yang sudah dilakukan	Ya, saya tahu	Ya, saya pernah	Ya, saya tahu	Ya, saya punya	Melakukan pemertanganan	Ya, saya punya	-	Perfungs dengan baik
Apakah rumah	Tidak tahu	Tidak pernah	Tidak tahu	Ya, saya punya	Melakukan pemertanganan	Tidak punya	-	-
Surga	Ya, saya tahu	Tidak pernah	Tidak tahu	Ya, saya punya	Sampahnya jumlah hasil daurugt ini tidak akan terlalu banyak	Ya, saya punya	-	Perfungs dengan baik

	kondisi rumah	perencanaan air	sumber mata air	kondisi mata air	menembang gigitan legal	pekerjaan rumah	pernikfaatan	pekerjaan rumah	penghijauan
-	Perencanaan khusus yang dibuat sendiri	Sumur	Jernih, tidak berbau dan rasanya enak	Tidak pernah	Ya, saya punya	-	Tidak gunakan atau tidak dimanfaatkan	Tidak pernah	-
Tidak berfungsi dengan baik karena rusak	Pant darah rumah	Sumur	Jernih, tidak berbau dan rasanya enak	Tidak pernah	Ya, saya punya	-	Tidak gunakan atau tidak dimanfaatkan	Ya, pernah	-
Darfungs dengan baik	Pant darah rumah	Sumur	Jernih, tidak berbau dan rasanya enak	Tidak pernah	Ya, saya punya	-	Digunakan untuk menanam sayur-sayuran	Ya, pernah	-
Darfungs dengan baik	Pant darah rumah	Sumur	Jernih, tidak berbau dan rasanya enak	Tidak pernah	Ya, saya punya	-	Tidak gunakan atau tidak dimanfaatkan	Ya, pernah	-
Darfungs dengan baik	Kan	Kan	Jernih, tidak berbau dan rasanya enak	Tidak pernah	Tidak punya	-	-	Ya, pernah	-

hewan ternak tidak hewan ternak status hewan ternak tempat menanam hewan ternak

Tidak	-	-	-
Tidak	-	-	-
Tidak	-	-	-
Ya	Lempe	Ya, itu milik saya	Kandang milik sendiri
Tidak	-	-	-

Gambar 4. 7 Sample Variabel yang digunakan

Keterangan variabel yang digunakan dalam penelitian ini dapat di lihat pada tabel

4.2

Tabel 4. 2 Variabel yang digunakan

No	Singkatan	Keterangan
1	TMS	Tempat Membuang Sampah
2	PLS	Pengelolaan Sampah
3	MSMB	Mengelola Sampah Menjadi Bermanfaat
4	PMS	Pemanfaatan Sampah

Tabel 4. 2 Variabel yang digunakan (Lanjutan)

No	Singkatan	Keterangan
5	DS	Dampak Sampah
6	DSR	Dampak Sampah yang di rasakan
7	MJ	Mempunyai Jamban
8	KD	Kondisi Jamban
9	MCKU	MCK Umum
10	KMCKU	Kondisi MCK Umum
11	PART	Pembuangan Air Rumah Tangga
12	SMA	Sumber Mata Air
13	KMA	Kondisi Mata Air
14	MPI	Menebang Pohon Ilegal
15	PR	Pekarangan Rumah
16	PPR	Pemanfaatan Pekarangan Rumah
17	PHJ	Penghijauan
18	HT	Hewan Ternak
19	JHT	Jenis Hewan Ternak
20	SHT	Status Hewan Ternak
21	TMHT	Tempat Menaruh Hewan Ternak
22	KKK	Kondisi Kandang Kolektif
23	TMRKK	Tidak Menaruh di Kandang Kolektif
24	TMKK	Tidak Memiliki Kandang Kolektif

Tabel 4. 2 Variabel yang digunakan (Lanjutan)

No	Singkatan	Keterangan
25	PG C	Penambahan Galian C
26	PDML	Penambahan dapat Merusak Lingkungan

Dalam data tersebut ada label sebagai luaran yang di tandai dengan angka 1 dan 0, Dimana angka 1 adalah Sehat dan angka 0 adalah Tidak Sehat, berdasarkan atribut variabel yang digunakan.

Setelah dilakukan pemilihan variabel ada 12 variabel yang tidak digunakan karena tidak berpengaruh terhadap data ujicoba, data tersebut dapat dilihat pada tabel 4.3

Tabel 4. 3 Daftar Variabel yang tidak digunakan

No	Nama Variabel
1	Id
2	Nama
3	Jenis Kelamin
4	Umur
5	Pendidikan
6	Sekolah
7	Putus Sekolah
8	Provinsi
9	Kabupaten

Tabel 4. 3 Daftar Variabel yang tidak digunakan lanjutan

No	Nama Variabel
10	Kecamatan
11	Desa
12	Label

Pada tahap preprocessing selanjutnya target klasifikasi dari proses accept/rejected Loan yaitu target y yang di mana itu adalah label dapat dilihat hasilnya pada gambar 4.8

```

0      1
1      0
2      1
3      0
4      1
..
4619   0
4620   1
4621   0
4622   1
4623   0
Name: label, Length: 4624, dtype: int64

```

Gambar 4. 8 Target Klasifikasi Label

b. Inisialiasi

Data yang bersifat karakter atau kategori diubah menjadi bentuk numerik atau angka, hal tersebut dilakukan supaya data yang bersifat kategorikal atau teks bisa dibaca atau di proses pada perhitungan algoritma. Pada tahap ini dilakukan label encoding dengan Teknik Label Encoding dan One-Hot Encoding. Label encoding mengubah setiap nilai dalam kolom menjadi angka yang berurutan. One-Hot Encoding adalah proses Dimana pembuatan variabel baru yang kemudian menjadi

hasil yang akan di hitung pada proses klasifikasi. Proses tf-idf_transformes menggunakan OneHotEncoder dapat dilihat pada gambar 4.9

```
tfidf_transformer = OneHotEncoder()
X = tfidf_transformer.fit_transform(xTarget)
print(X)
```

Gambar 4.9 tf-idf_transformer menggunakan OneHotEncoder

(0, 3207)	1.0	(4623, 4641)	1.0
(0, 3917)	1.0	(4623, 4644)	1.0
(0, 3957)	1.0	(4623, 4647)	1.0
(0, 4005)	1.0	(4623, 4650)	1.0
(0, 4009)	1.0	(4623, 4653)	1.0
(0, 4012)	1.0	(4623, 4657)	1.0
(0, 4013)	1.0	(4623, 4660)	1.0
(0, 4018)	1.0	(4623, 4663)	1.0
(0, 4056)	1.0	(4623, 4666)	1.0
(0, 4303)	1.0	(4623, 4669)	1.0
(0, 4637)	1.0	(4623, 4675)	1.0
(0, 4642)	1.0	(4623, 4677)	1.0
(0, 4645)	1.0	(4623, 4681)	1.0
(0, 4648)	1.0	(4623, 4684)	1.0
(0, 4651)	1.0	(4623, 4687)	1.0
(0, 4656)	1.0	(4623, 4691)	1.0
(0, 4659)	1.0	(4623, 4696)	1.0
(0, 4662)	1.0	(4623, 4701)	1.0
		(4623, 4702)	1.0

Gambar 4. 10 Hasil Proses OneHotEncoder

```
#Merubah (encode) nilai yTarget menjadi dua class yaitu 1 dan 0
encoder = LabelBinarizer()
Y = encoder.fit_transform(yTarget)
print(Y)
```

Gambar 4. 11 Merubah encode nilai y target menjadi dua class yaitu 1 dan 0

```
[[1]
 [0]
 [1]
 ...
 [0]
 [1]
 [0]]
```

Gambar 4. 12 Hasil Label Binarizer

```
#Merubah (encode) nilai atribut menjadi index nilai
tfidf_transformer = OneHotEncoder()
X = tfidf_transformer.fit_transform(xTarget)
print(X)
print(X.shape)
```

Gambar 4. 13 Merubah encode nilai atribut menjadi index nilai

```
(0, 3207) 1.0 (4622, 4608) 1.0
(0, 2017) 1.0 (4622, 4603) 1.0
(0, 2057) 1.0 (4623, 4606) 1.0
(0, 4685) 1.0 (4622, 4600) 1.0
(0, 4699) 1.0 (4622, 4675) 1.0
(0, 4612) 1.0 (4623, 4677) 1.0
(0, 4613) 1.0 (4623, 4681) 1.0
(0, 4618) 1.0 (4623, 4684) 1.0
(0, 4656) 1.0 (4623, 4687) 1.0
(0, 4603) 1.0 (4623, 4601) 1.0
(0, 4627) 1.0 (4623, 4606) 1.0
(0, 4662) 1.0 (4623, 4701) 1.0
(0, 4645) 1.0 (4623, 4702) 1.0
(0, 4648) 1.0 (4623, 4705) 1.0
(0, 4651) 1.0 (4623, 4708) 1.0
(0, 4653) 1.0 (4623, 4711) 1.0
(0, 4659) 1.0 (4623, 4712) 1.0
(0, 4682) 1.0 (4623, 4734) 1.0
(0, 4685) 1.0 (4623, 4717) 1.0
(0, 4688) 1.0 (4623, 4738) 1.0
(0, 4672) 1.0 (4624, 4723)
```

Gambar 4. 14 Hasil Merubah encode nilai atribut menjadi index nilai

Untuk menentukan TF-IDF (Term Frequency Inverse Document Frequency) yang mana metode ini adalah metode yang digunakan untuk menentukan nilai frekuensi sebuah kata di dalam sebuah dokumen menggunakan One Hot Encoder, One Hot Encoder sendiri adalah sebuah Teknik untuk mengubah

data kategori (*categorical data*) menjadi representasi biner yang disebut vektor one-hot. Ini digunakan untuk mengatasi masalah ketika data memiliki fitur kategori yang tidak bisa langsung digunakan oleh sebagian besar algoritma pembelajaran mesin, yang biasanya membutuhkan input numerik.

4.6. Pemodelan Data

Setelah data diklasifikasikan maka akan melalui proses tahap selanjutnya yaitu melakukan pemodelan data dengan melakukan input data lingkungan Masyarakat dalam program sustainable development goals (SDGs). Pemodelan ini dilakukan untuk dapat memberikan hasil yang ingin dicapai secara optimal yaitu dua kelas (Sehat dan tidak sehat). Pemodelan data pada penelitian ini menggunakan dua model atau dua algoritma yaitu *Naïve Bayes* dan *Random Forest* yang masing-masing memiliki performance yang memiliki fungsi sebagai validasi untuk dapat dijadikan acuan dalam mencari keakuratan data. Untuk melakukan pemodelan data dipilih beberapa atribut yang mendukung untuk pengambilan keputusan pada output yang diinginkan. Penelitian kali ini menggunakan 24 atribut sebagai pendukung keputusan.

Setelah menentukan atribut sebagai komponen pendukung hasil yang diinginkan maka dilakukan pemodelan data, pemodelan pertama dilakukan menggunakan algoritma *Naïve Bayes* dan yang kedua adalah Algoritma *Random Forest*. Untuk pemodelan *Naïve Bayes* yang dilakukan beberapa percobaan pada pemrograman python dengan data training dan data testing 10:90, 20:80 30:70,

40:60 50:50, dengan random state = 42 dapat dilihat pada gambar 4.15 di bawah ini.

```
X_train, X_test, y_train, y_test=train_test_split(X,y,test_size = 0.2, random_state=42)
print(X_train.shape)
print(y_train.shape)
print(X_test.shape)
print(y_test.shape)

(3699, 84)
(3699, 1)
(925, 84)
(925, 1)
```

Gambar 4. 15 Kode training dan data testing dari dataset dengan data test = 80% dan training 20%

4.7. Algoritma Naïve Bayes

Algoritma *Naive Bayes* merupakan sebuah metode klasifikasi menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Algoritma *Naive Bayes* memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dari *Naive Bayes Classifier* ini adalah asumsi yg sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian.

Keuntungan penggunaan metode ini yaitu hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Karena yang diasumsikan sebagai variabel *independent*, maka hanya varians dari suatu variabel dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians.

Berikut adalah kode pembuatan model training dan kode prediksi yang dibuat dengan Naïve Bayes dalam melakukan pemodelan dataset Sustainable Development Goals menggunakan Multinomial Naïve Bayes, kode tersebut dapat di lihat pada gambar 4.16 dan gambar 4.17

```
NaiveBayes = MultinomialNB().fit(X_train, np.ravel(y_train,order='C'))
print(NaiveBayes)

MultinomialNB()
```

Gambar 4. 16 Model Training Dengan Naïve Bayes

```
prediction = naiveBayes.predict(X_test)
Precision = precision_score(y_test, prediction)
Accuracies = accuracy_score(y_test, prediction)
recall=recall_score(y_test, prediction)
F1 = F1_score(y_test, prediction)
print(prediction)
print("Precision = ", Precision)
print("Accuracies = ", accuracies)
print("Recall) = ", recall)
print("F1 score = ", F1)
```

Gambar 4. 17 Kode Program prediksi terhadap model yang dibuat

Untuk melakukan evaluasi pada algoritma *Naïve Bayes* dan *Random Forest* setelah dilakukan split data maka langkah selanjutnya adalah dengan menggunakan *Confusion Matrix*, dari setiap percobaan dibawah ini sudah ditampilkan pada gambar confusion matrix baik gambar tanpa normalisasi atau setelah dilakukan normalisasi. Berikut adalah source code untuk menampilkan confusion matrix dan prosedur untuk memetakan confusion matrix, dan visualisasi hasil dapat di lihat pada gambar 4.18 dan gambar 4.19.

```

# Menampilkan matriks confusion
# print(confusion_matrix(y_test, prediction))

# Prosedur untuk membuat matriks confusion
def plot_confusion_matrix(cm, title, y_true, y_pred, classes,
                          normalize=False,
                          titlehome,
                          cmap=plt.cm.Blues):
    if not title:
        if normalize:
            title = 'Normalized confusion matrix'
        else:
            title = 'Confusion matrix without normalization'

    # Compute confusion matrix
    cm = confusion_matrix(y_true, y_pred)
    # Only use the labels that appear in the data
    classes = classes[unique_labels(y_true, y_pred)]
    if normalize:
        cm = cm.astype('float') / cm.sum(axis=1) # row-normalize
        print('Normalized confusion matrix')
    else:
        print('Confusion matrix, without normalization')

    print(cm)

```

Gambar 4. 18 Source Code Menampilkan Confusion Matrix

```

fig, ax = plt.subplots()
im = ax.imshow(cm, interpolation='nearest', cmap=cmap)
ax.figure.colorbar(im, ax=ax)
# We want to show all ticks...
ax.set(xticks=np.arange(cm.shape[1]),
       yticks=np.arange(cm.shape[0]),
       # .. and label them with the respective list entries
       xticklabels=classes, yticklabels=classes,
       title=title,
       ylabel='True label',
       xlabel='Predicted label')

# Rotate the tick labels and set their alignment
plt.setp(ax.get_xticklabels(), rotation=45, ha="right",
         rotation_mode="anchor")

# Loop over data dimensions and create text annotations.
fmat = .2f if normalize else 'd'
thresh = cm.max() / 2.
for i in range(cm.shape[0]):
    for j in range(cm.shape[1]):
        ax.text(j, i, format(cm[i,j], fmat),
                ha="center", va="center",
                color="white" if cm[i,j] > thresh else "black")
fig.tight_layout()
return ax

# Visualisasi Hasil
class_names = y_target
np.set_printoptions(precision=1)

# Plot non-normalized confusion matrix
plot_confusion_matrix(y_test, y_prediction, classes=class_names,
                      title='Confusion matrix, without normalization')

# Plot normalized confusion matrix
plot_confusion_matrix(y_test, y_prediction, classes=class_names, normalize=True,
                      title='Normalized confusion matrix')

plt.show()

```

Gambar 4. 19 Source Code Menampilkan Confusion Matrix (Lanjutan)

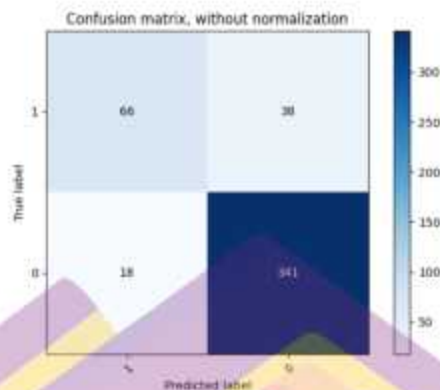
4.7.1 Algoritma Naïve Bayes Data Training dan Testing 90 : 10

Untuk melihat data training dengan data traintest 90 : 10 dalam dataset untuk memprediksi tingkat kesehatan lingkungan di kabupaten Lombok Timur di dapatkan hasil akurasi 0.88 % dengan nilai presisi untuk kelas 0 0.79 %, kelas 1 0.90%, nilai recall untuk kelas 0 0.69% dan kelas 1 0.95% untuk f1-score di dapatkan hasil untuk kelas 0 0.70% dan untuk kelas 1 didapatkan hasil 0.92%, Hasil Data Training dapat dilihat pada gambar 4. 20

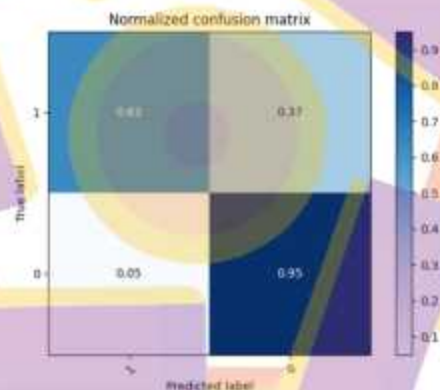
	precision	recall	f1-score	support
0	0.79	0.63	0.70	184
1	0.90	0.95	0.92	359
accuracy			0.88	463
macro avg	0.84	0.79	0.81	463
weighted avg	0.87	0.86	0.87	463

Gambar 4. 20 Pengujian Pertama Traintest 90:10

Tampilan *Confusion Matrix* tanpa normalisasi dan setelah dilakukan normalisasi untuk model *Naïve Bayes* dengan data *training* dan data *testing 90 : 10* , dapat dilihat pada gambar gambar 4.22 dan gambar 4.23 dibawah, Gambar tersebut menjelaskan bahwa Nilai *True Positif* adalah 66, *False Positif* adalah 38, untuk Nilai *False Negatif*nya adalah 18 dan untuk nilai *True Negative* nya adalah 341. Nilai-nilai tersebut dapat digunakan untuk mengevaluasi.



Gambar 4. 21 Confusion Matrix Tanpa Normalisasi



Gambar 4. 22 Confusion matrix setelah normalisasi

4.7.2 Algoritma Naïve Bayes Data Training dan Testing 80 : 20

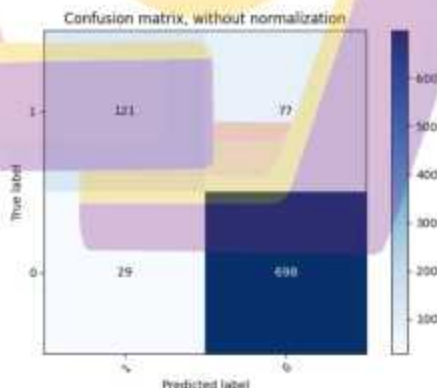
Pada pengujian ini dilakukan dengan data training dan data testing 80 : 20. Pengujian dilakukan sebanyak satu kali dengan menetapkan *random state* pada masing-masing pengujian yaitu 42. Dalam Pengujian ini menghasilkan nilai akurasi sebesar 0.89 atau 89 %, Presisi 0.81 atau 81 % di kelas 0, 0,90 atau 90 % di kelas 1

Recall 0.61 atau 61% di kelas 0, 0,96 atau 96 % di kelas 1, dan f1-Score 0.70 atau 70 % di kelas 0 dan 0,90 atau 90 % di kelas 1. Adapun hasil dari percobaan tersebut dapat dilihat pada gambar 4. 23

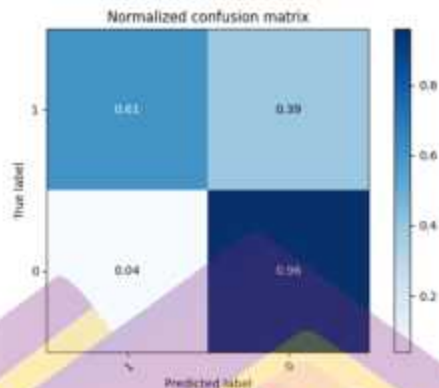
	precision	recall	f1-score	support
0	0.81	0.61	0.70	198
1	0.90	0.96	0.93	727
accuracy			0.89	925
macro avg	0.85	0.79	0.81	925
weighted avg	0.88	0.89	0.88	925

Gambar 4. 23 Pengujian Kedua TrainTest 80 : 20

Confusion Matrix tanpa normalisasi dan setelah dilakukan normalisasi untuk model *Naïve Bayes* dengan data *training* dan data *testing* 80 : 20 , dapat dilihat pada gambar gambar 4.24 dan gambar 4.25 dibawah, Gambar tersebut menjelaskan bahwa Nilai *True Positif* adalah 121, *False Positif* adalah 77, untuk Nilai *False Negatif*nya adalah 29 dan untuk nilai *True Negative* nya adalah 698. Nilai-nilai tersebut dapat digunakan untuk mengevaluasi.



Gambar 4. 24 Confusion matrix tanpa normalisi traintest 80:20



Gambar 4. 25 Confusion matrix tanpa normalisi traintest 80:20

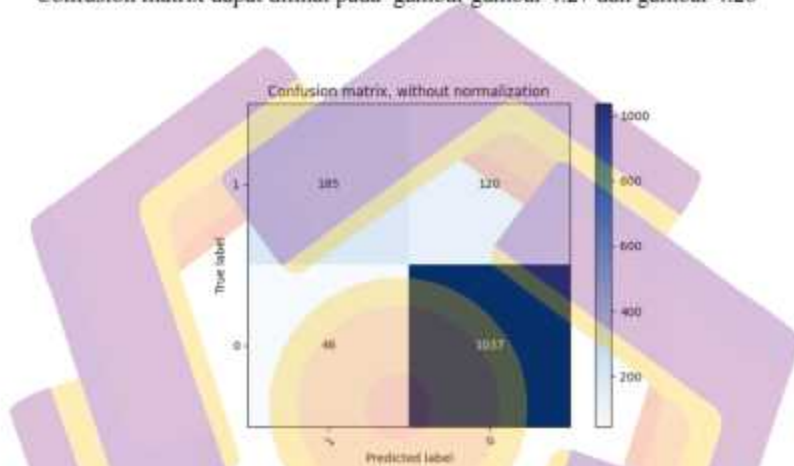
4.7.3 Algoritma Naïve Bayes Data Training dan Data Testing 70 : 30

Untuk menguji algoritma naïve bayes dengan data training 70 dan data testing 30, dengan penentuan random state = 42. Dalam pengujian ini menghasilkan nilai akurasi di 0.88 %. Nilai presisi di kelas 0 0.80% dan kelas 1 0.90 %. Sedangkan untuk nilai recall di kelas 0 menghasilkan 0.61% dan di kelas 1 mendapatkan hasil 0.96% Adapun untuk nilai f1- score mendapatkan nilai untuk kelas 0 mendapatkan nilai 0.69% dan untuk kelas 1 menghasilkan nilai 0.93%. Untuk melihat hasil dari prediksi dengan data training 30 : 70 dapat dilihat pada gambar 4.26

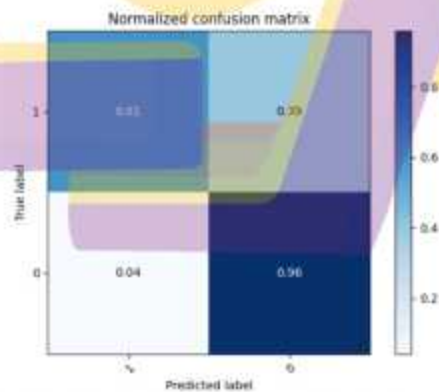
	precision	recall	f1-score	support
0	0.80	0.61	0.69	305
1	0.90	0.96	0.93	1023
accuracy			0.88	1388
macro avg	0.85	0.78	0.81	1388
weighted avg	0.88	0.88	0.87	1388

Gambar 4. 26 Pengujian Ketiga TrainTest 70 : 30

Pada gambar *Confusion Matrix* tanpa normalisasi dan setelah dilakukan normalisasi untuk model *Naïve Bayes* dengan data *training* dan data *testing* 70 : 30. Gambar tersebut menjelaskan bahwa Nilai *True Positif* adalah 185, *False Positif* adalah 120, untuk Nilai *False Negatif*nya adalah 46 dan untuk nilai *True Negative*nya adalah 1037. Nilai-nilai tersebut dapat digunakan untuk mengevaluasi. Gambar *Confusion matrix* dapat dilihat pada gambar gambar 4.27 dan gambar 4.28



Gambar 4. 27 Confusion matrix tanpa normalisasi traintest 70:30



Gambar 4. 28 Confusion matrix normalisasi traintest 70:30

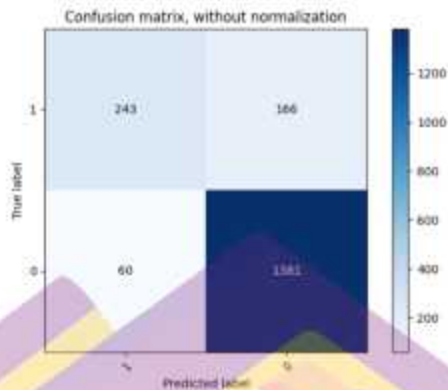
4.7.4 Algoritma Naïve Bayes dengan data training dan data testing 60 : 40

Dalam pengujian algoritma naïve bayes dengan data training dan data testing 60 : 40 di dapatkan hasil dengan akurasi 0.88 %, untuk nilai presisi untuk kelas 0 di dapatkan 0.80 % dan kelas 1 didapatkan 0.89, untuk nilai recall kelas 0 di dapatkan hasil 0.59% dan kelas 1 0.96% untuk hasil f1-score diperoleh kelas 0 0,68 % dan kelas 1 0.92 %. Untuk gambar confusion matrix dapat dilihat pada gambar 4.29

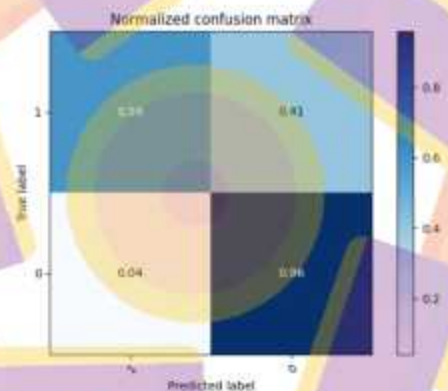
	precision	recall	f1-score	support
0	0.80	0.59	0.68	405
1	0.89	0.96	0.92	1441
accuracy			0.88	1856
macro avg	0.85	0.78	0.80	1856
weighted avg	0.87	0.88	0.87	1856

Gambar 4. 29 Pengujian Keempat TrainTest 60 : 40

Pada gambar Selanjutnya *Confusion Matrix* tanpa normalisasi dan setelah dilakukan normalisasi untuk model *Naïve Bayes* dengan data *training* dan data *testing 60 : 40*, Gambar tersebut menjelaskan bahwa Nilai *True Positif* adalah 243, *False Positif* adalah 166, untuk Nilai *False Negatif*nya adalah 60 dan untuk nilai *True Negative* nya adalah 1382. Nilai-nilai tersebut dapat digunakan untuk mengevaluasi. Gambar *Confusion matrix* dapat dilihat pada gambar gambar 4.30 dan gambar 4.31



Gambar 4. 30 Confusion matrix tanpa normalisasi traintest 60:40



Gambar 4. 31 Confusion matrix normalisasi traintest 60:40

4.7.5 Algoritma Naïve Bayes dengan data training dan data training 50 : 50

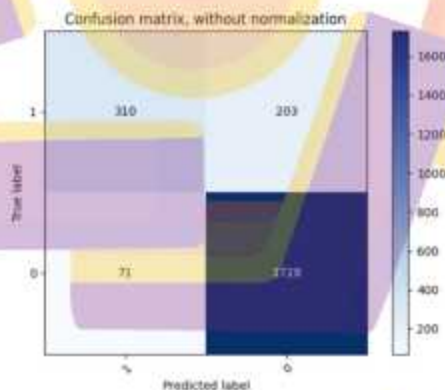
Pengujian terakhir dalam algoritma naive bayes dengan data training 50 dan data testing 50 di dapatkan hasil akurasi 0.88 %. Untuk nilai presisi di kelas 0 di dapatkan 0.81 % dan kelas 1 0.89 %, untuk hasil recall di kelas 0 di dapatkan 0.60 % dan kelas 1 di dapatkan hasil 0.96, untuk hasil f1-score di dapatkan hasil di kelas

0.69 % dan di kelas 1 0.93 %. Untuk melihat hasil confusion matrix dapat dilihat pada gambar 4.32

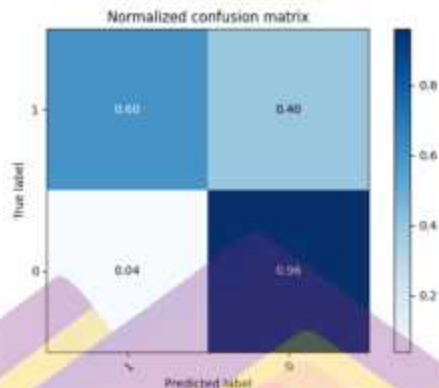
	precision	recall	f1-score	support
0	0.81	0.68	0.69	513
1	0.89	0.96	0.93	1799
accuracy			0.88	2312
macro avg	0.85	0.78	0.81	2312
weighted avg	0.88	0.88	0.87	2312

Gambar 4.32 Pengujian Kelima TrainTest 50 : 50

Pada *Confusion Matrix* tanpa normalisasi dan setelah dilakukan normalisasi untuk model *Naïve Bayes* dengan data *training* dan data *testing* 50 : 50, Gambar tersebut menjelaskan bahwa Nilai *True Positif* adalah 310, *False Positif* adalah 203, untuk Nilai *False Negatif*nya adalah 71 dan untuk nilai *True Negative* nya adalah 1728. Nilai-nilai tersebut dapat digunakan untuk mengevaluasi. Gambar *Confusion matrix* dapat dilihat pada gambar gambar 4.33 dan gambar 4.34



Gambar 4.33 Confusion matrix tanpa normalisasi traintest 50:50



Gambar 4. 34 Confusion matrix normalisasi traintest 50:50

Keterangan Performance : Performance terbaik yang di hasilkan oleh algoritma naïve bayes didapatkan ketika melakukan traintest 80:20 yang kemudian di visualiasiskan dengan confusion matrix pada gambar 4, 24 proses untuk mendapatkan nilai akurasi dihitung dengan persamaan berikut :

1. Mencari Nilai Akurasi :

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

$$Acc = \frac{121 + 698}{77 + 29 + 121 + 698} \times 100 \%$$

$$Acc = \frac{819}{925} \times 100 \%$$

$$Acc = 0,8854 \times 100$$

$$Acc = 89 \%$$

2. Mencari Nilai Presisi :

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

$$Pr = \frac{121}{77 + 121} \times 100 \%$$

$$Pr = \frac{121}{198} \times 100 \%$$

$$Pr = 0,61 \times 100 \%$$

$$Pr = 61 \%$$

3. Mencari Nilai Recall

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

$$Recall = \frac{121}{29 + 121} \times 100 \%$$

$$Recall = \frac{121}{150} \times 100 \%$$

$$Recall = 0,806 \times 100 \%$$

$$Recall = 80 \%$$

4. Mencari Nilai F1 Score

$$F1\ Score = \frac{2 \times recall \cdot presisi}{recall + presisi}$$

$$F1\ Score = \frac{2 \times 0,80 \cdot 0,61}{0,80 + 0,61}$$

$$F1\ Score = \frac{0,976}{1,41}$$

$$F1\ Score = 0,69 \cdot 100$$

$$F1\ Score = 69 \%$$

Dalam penelitian ini korelasi antara nilai Presisi (Precision), Recall, dan F1-Score adalah ukuran evaluasi kinerja dari suatu model klasifikasi, sedangkan Akurasi adalah ukuran yang mengukur seberapa baik model klasifikasi memprediksi kelas yang benar dari seluruh prediksi yang dilakukan. Korelasi antara

Presisi, Recall, dan F1-Score dengan Akurasi dapat bervariasi tergantung pada seberapa seimbang distribusi kelas dari data yang diamati, Korelasi positif dalam penelitian ini ketika Presisi, Recall, dan F1-Score tinggi, maka cenderung Akurasi juga tinggi. Ini terjadi ketika model mampu memprediksi dengan benar baik pada kelas positif maupun negatif, dan tidak terlalu banyak kesalahan dalam prediksinya, dan untuk korelasi negatif dalam kasus tidak sebangunnya distribusi kelas, korelasi antara Presisi, Recall, dan F1-Score dengan Akurasi bisa menjadi negatif. Ini terjadi ketika model cenderung memprediksi mayoritas kelas yang dominan, sehingga akurasi meningkat, tetapi pada saat yang sama Presisi, Recall, dan F1-Score untuk kelas minoritas bisa rendah. Presisi, Recall, dan F1-Score memberikan informasi tambahan tentang kinerja model klasifikasi, Akurasi memberikan gambaran keseluruhan seberapa baik model tersebut dalam memprediksi kelas-kelas yang ada. Korelasi antara Presisi, Recall, dan F1-Score dengan Akurasi tidak selalu linier dan tergantung pada distribusi kelas dalam data.

4.8. Algoritma Random Forest

Percobaan selanjutnya dilakukan dengan algoritma Random Forest. Pada algoritma Random Forest, ada beberapa langkah yang sama yang dilakukan dalam melakukan percobaan dengan data set yang sama dengan algoritma Naïve Bayes dan Random Forest. Karena langkah awal dalam melakukan klasifikasi dengan Naïve Bayes dengan Random forest sama, yaitu import package, import dataset, Variabel independen, Variabel dependen, melakukan split terhadap data uji dan

data training, import packe. Selanjutnya dilakukan Membuat model random forest terhadap Training set seperti pada Gambar 4. 35 di bawah ini.

```

from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import confusion_matrix, accuracy_score

# Inisialisasi model Random Forest
rf_model = RandomForestClassifier(n_estimators=30, random_state=42)

# Latih model
rf_model.fit(X_train, y_train)

```

Gambar 4. 35 Kode Membuat Model Random Forest

Setelah membuat model random forest terhadap data training maka dilakukan proses prediksi terhadap data uji yang sudah ditentukan. Kode yang digunakan untuk melakukan prediksi pada random forest bisa dilihat pada gambar 4. 36

```

# Lakukan prediksi pada data uji
y_pred = rf_model.predict(X_test)

# Hitung confusion matrix
cm = confusion_matrix(y_test, y_pred)

```

Gambar 4. 36 Kode Memprediksi Hasil Tes

Untuk mengevaluasi hasil dari prediksi yang dilakukan sebanyak 5 kali dengan data train yang berbeda dengan jumlah data uji yang sama, maka dilakukan dengan cara membuat confusion matrix adapun kode yang digunakan dapat dilihat pada gambar 4.37

```

# Tampilkan confusion matrix
print("Confusion Matrix:")
print(cm)

# Hitung akurasi
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
# Hitung presisi
precision = precision_score(y_test, y_pred)
print("Precision:", precision)

```

Gambar 4. 37 Kode Membuat Confusion Matrix

4.8.1 Random forest dengan data traintest 80:20

Pengujian algoritma Random Forest pertama kali dilakukan dengan menggunakan data traintest 80:20. Pengujian pertama kali ini dilakukan seperti pada algoritma lainnya pada pengujian dengan traintest, yaitu dilakukan sebanyak satu kali pengujian. Pengujian pertama dengan jumlah *random state* 42. Adapun hasil pengujian pada pengujian pertama random forest memperoleh nilai akurasi sebesar 0.99 atau 99%, Nilai presisi 0.98 atau 98 % untuk kelas 0 dan 0.99 atau 99% untuk kelas 1, Nilai Recall 0.98 atau 98 % untuk kelas 0 dan 1 atau 100 % untuk kelas 1. Nilai F1 Score di peroleh 0.98 atau 98 % untuk kelas 0 dan 1 atau 100 % untuk kelas 1.

Hasil pengujian pertama dapat dilihat pada gambar 4.38 di bawah ini

	precision	recall	f1-score	support
0	0.98	0.98	0.98	196
1	0.99	1.00	1.00	727
accuracy			0.99	925
macro avg	0.99	0.99	0.99	925
weighted avg	0.99	0.99	0.99	925

Gambar 4. 38 Hasil Random Forest Untuk TrainTest 80:20

Dari dua metode klasifikasi yang di gunakan dengan keunggulan masing-masing, dalam konteks penggunaan data Sustainable Development Goals (SDGs) lingkungan di sepuluh Kecamatan di Kabupaten Lombok Timur, terdapat beberapa alasan mengapa Random Forest mungkin lebih baik daripada Naive Bayes diantaranya

1. Kemampuan untuk menangani data yang kompleks

Random Forest cenderung lebih baik dalam menangani data yang kompleks dan beragam. Dalam konteks lingkungan, data dapat bervariasi dari berbagai sumber dan memiliki banyak atribut yang saling terkait. Random Forest mampu menangani korelasi antar atribut dengan baik, sementara Naive Bayes mengasumsikan independensi yang kuat antar atribut, yang mungkin tidak selalu terpenuhi dalam data lingkungan.

2. Ketahanan Terhadap Outlier

Random Forest secara alami lebih tahan terhadap outlier daripada Naive Bayes. Dalam lingkungan, keberadaan outlier atau anomali mungkin menjadi masalah serius karena dapat menyesatkan model. Random Forest, dengan cara membagi dataset ke dalam pohon keputusan, lebih mampu menangani outlier daripada Naive Bayes yang bergantung pada asumsi distribusi atribut yang seragam.

3. Performa yang lebih baik dengan kedalaman model yang dalam

Random Forest cenderung mendapatkan manfaat dari kedalaman model yang lebih dalam, sementara Naive Bayes biasanya tidak memperoleh keuntungan yang signifikan dari kompleksitas model yang lebih tinggi. Dalam beberapa kasus, untuk

memperoleh klasifikasi yang akurat dalam lingkungan yang kompleks, model yang lebih dalam mungkin diperlukan, dan Random Forest memiliki keunggulan dalam hal ini.

4. Manajemen overfitting yang lebih baik

Random Forest memiliki kemampuan bawaan untuk mengatasi overfitting, terutama ketika digunakan dengan jumlah pohon yang memadai. Ini karena setiap pohon dihasilkan dari sub-sample acak dari data pelatihan dan fitur, sehingga mencegah model menjadi terlalu khusus terhadap data pelatihan. Di sisi lain, Naive Bayes cenderung rentan terhadap overfitting, terutama jika distribusi atribut tidak sesuai dengan asumsi naive independensi.

5. Fleksibilitas dalam menangani jenis data yang berbeda

Random Forest lebih fleksibel dalam menangani berbagai jenis data, termasuk data yang tidak terstruktur. Dalam konteks lingkungan, data sering kali terdiri dari campuran data terstruktur dan tidak terstruktur, serta atribut kategoris dan numerik. Random Forest dapat dengan mudah menangani variasi ini, sementara Naive Bayes lebih baik digunakan untuk data yang terstruktur dan atribut yang bersifat kategoris.

4.9. Perbandingan Hasil

Pada tahap ini dilakukan adalah membandingkan hasil prediksi dari kedua algoritma yang digunakan. Naïve Bayes mendapatkan akurasi dengan nilai tertinggi di traintest 80 : 20, dengan nilai akurasi 89 %, dan algoritma Random Forest dengan akurasi 100 % . Pada percobaan yang dilakukan kedua algoritma menggunakan

random state yang sama yaitu 42. Untuk melihat gambar data perbandingan dapat di lihat pada gambar 4.39.



Gambar 4. 39 Perbandingan Hasil

Setelah dilakukan proses evaluasi maka kita dapat membandingkan hasil dari perhitungan algoritma *Naïve Bayes* dan *Random Forest*. Perbandingan hasil dapat kita lihat pada tabel dibawah ini. Tabel 4.3 Perbandingan hasil antara Algoritma *Naïve Bayes* dan *Random Forest*

Tabel 4. 4 Perbandingan Hasil

No	Keterangan	<i>Naïve Bayes</i>	<i>Random Forest</i>
1	Akurasi	89 %	99 %
2	Presisi	90 %	100 %
3	Recall	96 %	100 %
4	F1-Score	93 %	100 %

Metode Klasifikasi *Naïve Bayes* menggunakan data *training* dan data *testing* 80 :20 untuk menciptakan probabilitas setiap kriteria untuk berbagai kelas yang berbeda, dengan tujuan agar nilai-nilai probabilitas dari kriteria ini dapat ditingkatkan untuk memprediksi status tingkat kesehatan lingkungan berdasarkan proses klasifikasi yang dilakukan dengan metode Klasifikasi *Naïve Bayes* dan random forest, pada tabel perbandingan hasil tersebut di dapatkan nilai akurasi 89 % presisi 90 % recall 96 % dan f1- score 93 %. Dan pada algoritma random forest di dapatkan akurasi sebesar 99 %, presisi 100 %, recall 100 % dan f1-score 100 %.

Dari hasil evaluasi diatas dapat disimpulkan bahwa algoritma *random forest* memiliki nilai akurasi tertinggi dibandingkan dengan *naïve bayes*, itu di buktikan dengan nilai akurasi yang di dapatkan yaitu 99 % Dengan demikian dua algoritma tersebut dapat dikatakan layak digunakan untuk memprediksi status Tingkat kesehatan lingkungan di sepuluh kecamatan di kabupaten Lombok Timur.

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan dari penelitian yang dilakukan maka dapat disimpulkan bahwa Untuk mendeteksi tingkat kesehatan lingkungan masyarakat di sepuluh kecamatan di Kabupaten Lombok Timur. Dengan data set 4624 data angket dan menggunakan 26 variabel yang berpengaruh dari 38 variabel angket yang dibagikan.:

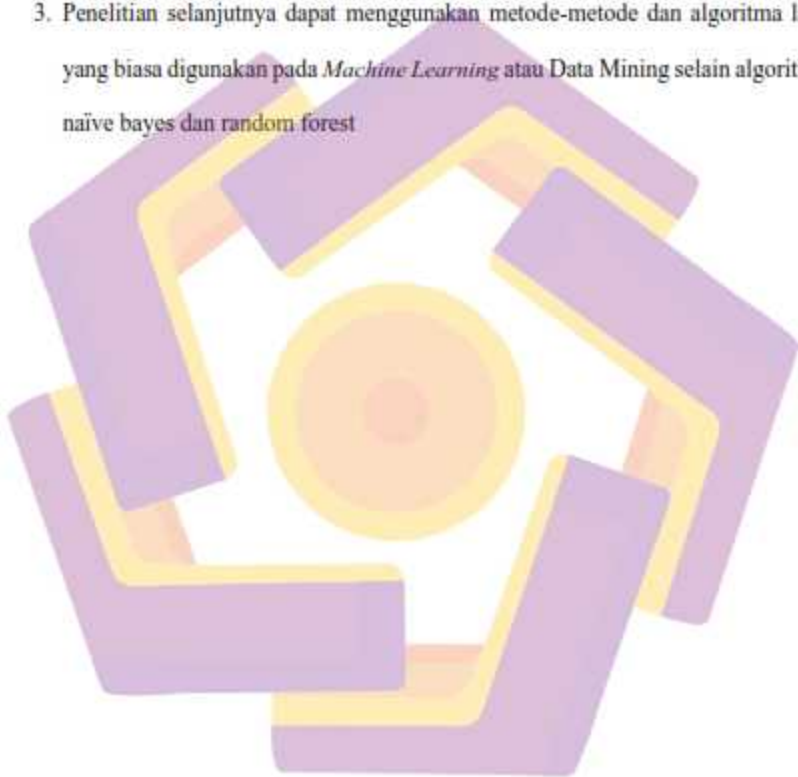
1. Performance Algoritma random forest lebih unggul dari algoritma naïve bayes ketika melakukan pengujian dengan traintest 80:20 dengan randomstate : 42
2. Dalam Pengujian menggunakan algortima naïve bayes dengan beberapa train test nilai akuarasi tertinggi di dapatkan ketika traintest 80:20 dengan nilai akurasi 89 %. Adapun Ketika pengujian algoritma random forest dengan traintest yang sama memperoleh nilai akuarasi 99 %. Dengan demikian algoritma random forest lebih bagus akurasinya ketika membandingkan dataset SDGs Lingkungan di sepuluh kecamatan di kabupaten Lombok Timur.
3. Berdasarkan hasil penelitian yang dilakukan faktor yang mempengaruhi Algoritma naïve bayes dan random forest adalah ketika melakukan split data dan ,melakukan traintest yang berbeda walaupun dengan randomstate yang sama.

5.2. Saran

Adapun beberapa saran untuk pengembangan penelitian ini adalah

1. Pada penelitian ini masih belum dilakukan uji coba berulang kali untuk mendapatkan hasil yang terbaik dalam traintest.

2. Proses pemilihan variabel pada penelitian ini masih mengacu kepada data angket yang disebar kepada Masyarakat sehingga masih belum maksimal, pada penelitian selanjutnya bisa menggunakan algoritma yang lain untuk menentukan variabel mana yang paling berpengaruh.
3. Penelitian selanjutnya dapat menggunakan metode-metode dan algoritma lain yang biasa digunakan pada *Machine Learning* atau Data Mining selain algoritma naïve bayes dan random forest

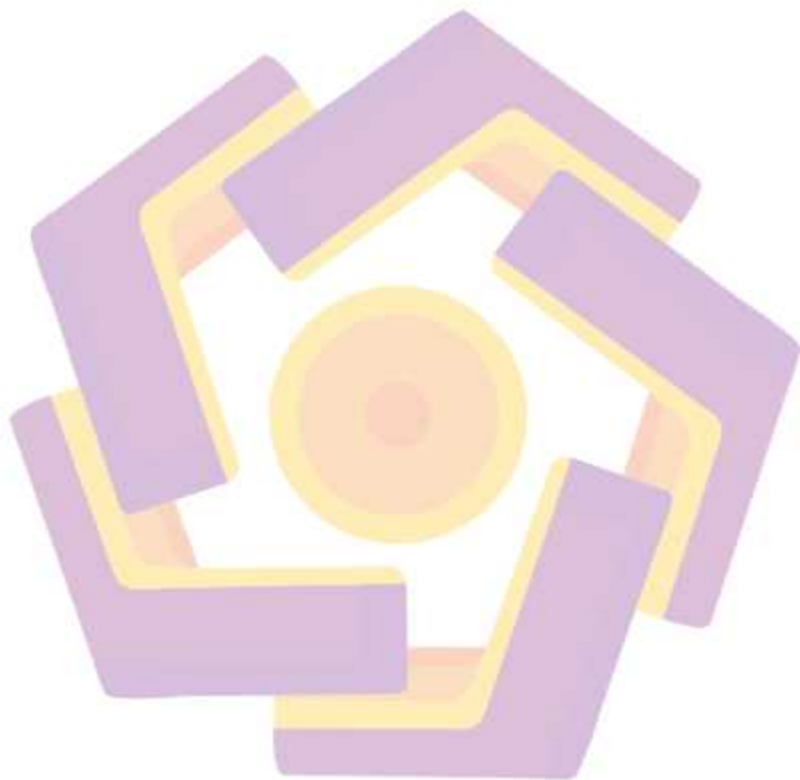


DAFTAR PUSTAKA

PUSTAKA MAJALAH, JURNAL ILMIAH ATAU PROSIDING

- Agustyaningrum, C. I., Gata, W., Nurfalah, R., Radiyah, U., & Maulidah, M. (2020). KOMPARASI ALGORITMA NAIVE BAYES, RANDOM FOREST DAN SVM UNTUK MEMPREDIKSI NIAT PEMBELANJA ONLINE. In *Jurnal Informatika* (Vol. 20, Issue 2).
- Khairina, E., Purnomo, E. P., & Malawnai, A. D. (2020). Sustainable Development Goals: Kebijakan Berwawasan Lingkungan Guna Menjaga Ketahanan Lingkungan Di Kabupaten Bantul Daerah Istimewa Yogyakarta. *Jurnal Ketahanan Nasional*, 26(2), 155. <https://doi.org/10.22146/jkn.52969>
- Kuspani Putra, Y. (2021). Analisis Perbandingan Pengaruh Pertumbuhan Ekonomi Terhadap Tingkat Kesejahteraan Masyarakat Pada Desa Suralaga Dengan Menggunakan Algoritma Naive Bayes Dan Support Vector Machine (Svm). *Jurnal Informatika Dan Teknologi*, 4(1). <https://doi.org/10.29408/jit.v4i1.2961>
- Nugroho, N. A., & Wijayanto, A. W. (n.d.). *Perbandingan Algoritma Machine Learning dalam Pengklasifikasian Tingkat Kemiskinan di Indonesia Tahun 2021 I*.
- Prancevičius, T., & Marcinkevičius, V. (2017). Comparison of Naive Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification. *Baltic Journal of Modern Computing*, 5(2). <https://doi.org/10.22364/bjmc.2017.5.2.05>
- Purwanto, A. D., Ibrahim, A., Ulfa, A., Parwati, E., & Supriyono, A. (n.d.). *PENGEMBANGAN MODEL IDENTIFIKASI HABITAT BENTIK MENGGUNAKAN PENDEKATAN SEGMENTASI OBJECT-BASED IMAGE ANALYSIS (OBIA) DAN ALGORITMA MACHINE LEARNING (STUDI KASUS: PULAU PARI, KEPULAUAN SERIBU)*.
- Samosir, A., Hasibuan, M., Justino, W. E., & Hariyono, T. (2019). Komparasi Algoritma Random Forest, Naïve Bayes dan K-Nearest Neighbor Dalam klasifikasi Data Penyakit Jantung. *Cardiovascular Diseases*, 15(2), 189–194.
- van der Heide, E. M. M., Veerkamp, R. F., van Pelt, M. L., Kamphuis, C., Athanasiadis, I., & Ducro, B. J. (2019). Comparing regression, naive Bayes, and random forest methods in the prediction of individual survival to second lactation in Holstein cattle. *Journal of Dairy Science*, 102(10), 9409–9421. <https://doi.org/10.3168/jds.2019-16295>

Yahya, Y., & Bahtiar, H. (2021). Pengaruh Pertumbuhan Ekonomi Terhadap Tingkat Kesejahteraan Masyarakat Kecamatan Selong Kabupaten Lombok Timur – Nusa Tenggara Barat Menggunakan Algoritma Naive Bayes. *Infotek : Jurnal Informatika Dan Teknologi*, 4(1), 20-28. <https://doi.org/10.29408/jit.v4i1.2981>



LAMPIRAN

Instrumen SDGs untuk Kepala Keluarga

A. Bidang Pendidikan

1. Lengkapi data dasar berikut

Nama Kepala Keluarga	Tanggal Lahir				Umur	Jenis Kelamin	Pendidikan Terakhir	Tempat Sekolah		Putus Sekolah	
								S	N	Ya	Tdk

2. **(DITANYAKAN JIKA ADA ANAK RESPONDEN YANG PUTUS SEKOLAH)** Apa faktor yang menyebabkan anak B/I/S putus sekolah?
- Menikah
 - Tidak ada biaya
 - Malas belajar atau kenakalan remaja
3. Apakah B/I/S membutuhkan program penyetaraan pendidikan seperti program paket A, B, dan paket C?
- saya sangat butuh
 - cukup butuh
 - tidak butuh

DITANYAKAN JIKA ADA ANAK RESPONDEN YANG PENDIDIKANNYA SAMPAI SMA

- Apakah anak B/I/S melanjutkan ke jenjang Perguruan Tinggi?
 - Ya (*lanjut ke pertanyaan no 6*)
 - Tidak (*lanjut ke pertanyaan no 5*)
- Apakah faktor yang menyebabkan sehingga anak B/I/S tidak melanjutkan ke Perguruan Tinggi?
 - Tidak ada biaya kuliah
 - Menikah
 - Kenakalan remaja atau malas belajar
- Ke Perguruan Tinggi mana anak B/I/S melanjutkan studi?

- a. Di wilayah Lombok Timur (*lanjut ke pertanyaan no 7*)
 - b. Kota Mataram
 - c. Luar Daerah
7. Perguruan tinggi mana tempat anak B/I/S melanjutkan studi?
- a. Universitas Hamzanwadi
 - b. Universtias Gunung Rinjani
 - c. Perguruan Tinggi di Anjani
 - d. Lainnya...
8. Dari mana B/I/S mendapatkan informasi Perguruan Tinggi tersebut?
- a. Media sosial
 - b. Selebaran atau spanduk
 - c. Kerabat keluarga
 - d. Cabang atau ranting Organisasi
9. Apakah B/I/S memiliki anggota keluarga yang tidak bisa membaca dan menulis?
- a. Ya, ada
 - b. Tidak ada
10. Berapa umur anggota keluarga B/I/S yang belum bisa membaca dan menulis
- a. 12-15 tahun
 - b. 16-56 tahun
 - c. 60 tahun keatas

B. Bidang Lingkungan


1. Dimana tempat B/I/S sering membuang sampah?
 - a. Sangai
 - b. Sembarang pekarangan kosong
 - c. Parit dekat rumah
 - d. Tempat pembuangan sampah (TPS) yang sudah disediakan
2. Apakah B/I/S tahu cara pengelolaan sampah rumah tangga?
 - a. Ya, saya tahu
 - b. Tidak tahu
3. Apakah B/I/S pernah mengelola sampah rumah tangga menjadi sesuatu yang bermanfaat?
 - a. Ya, saya pernah
 - b. Tidak pernah
4. Apakah B/I/s tahu cara pemanfaatan sampah rumah tangga?
 - a. Ya, saya tahu
 - b. Tidak tahu
5. Apakah B/I/S tahu dampak sampah rumah tangga
 - a. Ya, saya tahu
 - b. Tidak tahu

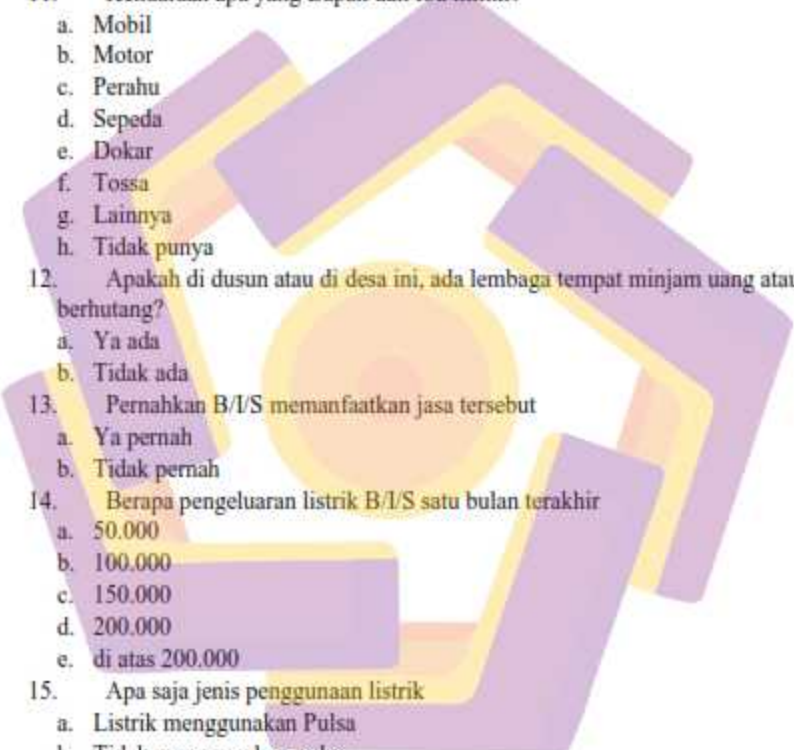
6. Dari dampak sampah rumah tangga berikut, manakah yang paling B/I/S rasakan?
 - a. Banyaknya jumlah lalat disekitar kita khususnya ketika makan
 - b. Merusak pemandangan
 - c. Menimbulkan penyakit
7. Apakah B/I/S mempunyai jamban sendiri?
 - a. Ya, saya punya
 - b. Tidak punya
8. Bagaimana kondisi jamban yang B/I/S miliki?
 - a. Berfungsi dengan baik
 - b. Berapa orang yang menggunakan jamban tersebut?
9. Apakah ada MCK umum di desa B/I/S?
 - a. Ada
 - b. Tidak ada
10. Bagaimana kondisi MCK umum tersebut?
 - a. Berfungsi dengan baik
 - b. tidak berfungsi dengan baik karena rusak
11. kemana B/I/S membuang/mengalirkan air rumah tangga?
 - a. Kali
 - b. Parit dekat rumah
 - c. Pembuangan khusus yang dibuat sendiri
12. Apa sumber mata air B/I/S?
 - a. Sumur
 - b. Air ledeng
 - c. Sungai
13. Bagaimana kondisi sumber mata air B/I/S?
 - a. Jernih, tidak berbau dan rasanya enak
 - b. Tidak jernih tapi tidak berbau dan rasanya enak
 - c. Tidak jernih, berbau dan rasanya tidak enak
14. Apakah B/I/S pernah menebang pohon secara ilegal?
 - a. Pernah
 - b. Tidak pernah
15. Apakah B/I/S memiliki pekarangan rumah?
 - a. Ya, saya punya
 - b. Tidak punya
16. Bagaimana pemanfaatan pekarangan rumah B/I/S?
 - a. Digunakan untuk menanam sayur-sayuran
 - b. Menanam bunga
 - c. Tidak gunakan atau tidak dimanfaatkan
17. Apakah B/I/S pernah menam pohon sebagai bentuk penghijauan?
 - a. Ya, pernah
 - b. Tidak pernah
18. Apakah B/I/S memiliki hewan ternak?

- a. Ya
 - b. Tidak
19. Apa jenis hewan ternak yang B/I/S miliki?
- a. Sapi
 - b. Kambing
 - c. Kuda
 - d. Lainnya....
20. Apakah sebagian besar hewan ternak tersebut miliki B/I/S?
- a. Ya, itu milik saya
 - b. Bukan, itu miliki orang lain
21. Apakah B/I/S menaruh hewan ternak di kandang kolektif?
- a. Kandang kolektif
 - b. Kandang milik sendiri
22. Bagaimana kondisi kandang kolektif B/I/S?
- a. Berfungsi dengan baik
 - b. Tidak berfungsi dengan baik, butuh perbaikan
23. Kenapa B/I/S tidak menaruh hewan ternak di kandang kolektif?
- a. Karena kami tidak memiliki kandang kolektif
 - b. Karena dikandang milik sendiri lebih nyaman dan aman
24. Apa yang menyebabkan B/I/S tidak memiliki kandang kolektif
- a. Tidak diprogramkan oleh desa
 - b. Kurangnya kekompakan para peternak
25. Apakah di desa ini adalah penambangan bahan galian golongan C?
- a. Ya ada
 - b. Tidak ada
26. Apakah B/I/S mengetahui bahwa penambangan bahan galian golongan C dapat merusak lingkungan?
- a. Ya, saya tahu
 - b. Saya tidak tahu

C. Bidang kemiskinan

1. Apa pekerjaan utama Bapak ?
- a. Petani
 - b. Buruh tani
 - c. Pedagang
 - d. Nelayan
 - e. lainnya
2. Berapa penghasilan Bapak setiap bulan?
- a. 200.000 – 500.000
 - b. 500.000 – 750.000
 - c. 750.000 – 1.000.000

- 
- d. 1.000.000 – 1.500.000
e. 1.500.000 – 2.000.000
f. 2.000.000 – 3.000.000
g. Diatas 3 juta
3. Apakah Bapak memiliki pekerjaan sampingan?
a. Ya ada
b. Tidak ada
4. Berapa penghasilan dari pekerjaan sampingan tersebut yang Bapak terima?
a. 200.000 – 500.000
b. 500.000 – 750.000
c. 750.000 – 1.000.000
d. 1.000.000 – 1.500.000
5. Apa pekerjaan utama Ibu?
a. Petani
b. Buruh tani
c. Pedagang
d. Nelayan
e. lainnya
6. Berapa penghasilan Ibu setiap bulan?
a. 200.000 – 500.000
b. 500.000 – 750.000
c. 750.000 – 1.000.000
d. 1.000.000 – 1.500.000
e. 1.500.000 – 2.000.000
f. 2.000.000 – 3.000.000
g. Diatas 3 juta
7. Apakah Ibu memiliki pekerjaan sampingan?
a. Ya ada
b. Tidak ada
8. Berapa penghasilan pekerjaan tambahan yang Ibu terima?
a. 200.000 – 500.000
b. 500.000 – 750.000
c. 750.000 – 1.000.000
d. 1.000.000 – 1.500.000
9. Apakah Bapak dan Ibu memiliki aktivitas ekonomi dalam rumah tangga seperti produksi barang dan jasa?
a. Ya, sebutkan (.....)
b. Tidak ada
10. Hewan ternak apa yang B dan Ibu miliki ?

- 
- a. Sapi
 - b. Kambing
 - c. Ayam
 - d. Ikan
 - e. Bebek
 - f. Lainnya
 - g. Tidak punya
11. Kendaraan apa yang Bapak dan Ibu miliki?
- a. Mobil
 - b. Motor
 - c. Perahu
 - d. Sepeda
 - e. Dokar
 - f. Tossa
 - g. Lainnya
 - h. Tidak punya
12. Apakah di dusun atau di desa ini, ada lembaga tempat minjam uang atau berhutang?
- a. Ya ada
 - b. Tidak ada
13. Pernahkan B/I/S memanfaatkan jasa tersebut
- a. Ya pernah
 - b. Tidak pernah
14. Berapa pengeluaran listrik B/I/S satu bulan terakhir
- a. 50.000
 - b. 100.000
 - c. 150.000
 - d. 200.000
 - e. di atas 200.000
15. Apa saja jenis penggunaan listrik
- a. Listrik menggunakan Pulsa
 - b. Tidak menggunakan pulsa
16. Kondisi rumah responden (**dilisi sendiri oleh suveyor berdasarkan pengamatan**)
- a. Permanen
 - b. semi permanen
 - c. tidak permanen
17. kondisi Fasilitas MCK (**dilisi sendiri oleh suveyor berdasarkan pengamatan**)
- a. Memadai

- b. Cukup memadai
 - c. Tidak memadai
18. Kondisi Sanitasi (**dilisi sendiri oleh suveyor berdasarkan pengamatan**)
- a. Baik
 - b. Cukup baik
 - c. Tidak baik
19. Bahan bakar apa yang B/I/S gunakan untuk memasak?
- a. kayu Bakar
 - b. Gas elpiji
 - c. Minyak Tanah
20. Fasilitas apa saja yang B/I/S gunakan dalam komunikasi di rumah?
- a. telpon, hp, dan internet
 - b. hp, dan internet
 - c. internet
21. Apakah tanah tempat B/I/S membangun rumah adalah milik sendiri?
- a. Ya milik sendiri
 - b. Tidak, milik orang lain

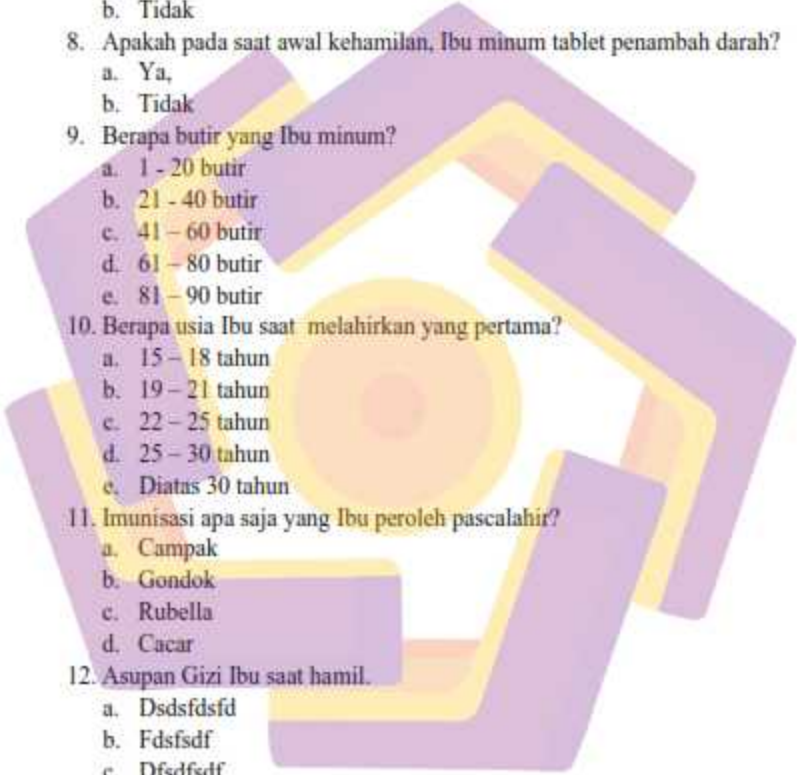
D. Pekerjaan Layak dan Pertumbuhan Ekonomi.

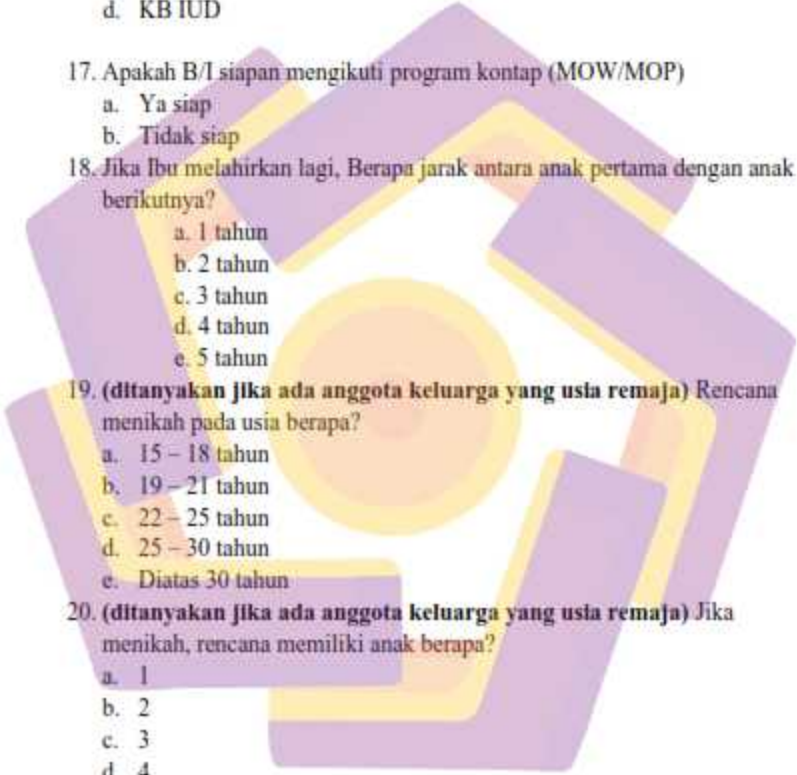
1. Berapa jumlah anggota keluarga B/I/S yang menjadi buruh tani?
.....
2. Apakah ada keluarga yang sedang bekerja sebagai TKI/TKW di luar negeri?
 - a. Ya ada, berapa (....)
 - b. Tidak ada
3. Negara mana tujuan keluarga B/I/S menjadi TKI/TKW?
 - a. Malasiya
 - b. Arab saudi
 - c. Korea
 - d. Lainnya
4. Apa Sumber penghasilan utama sebagian besar penduduk :
 - a. Pertanian
 - b. Pertambangan dan penggalian
 - c. Industri pengolahan (pabrik, kerajinan, dll.)
 - d. Perdagangan besar/eceran dan rumah makan
 - e. Angkutan, pergudangan, komunikasi
 - f. Jasa
 - g. Lainnya
5. Jika sumber penghasilan utama sebagian besar penduduk adalah sektor pertanian. Jenis komoditi apa yang B/I/S geluti :

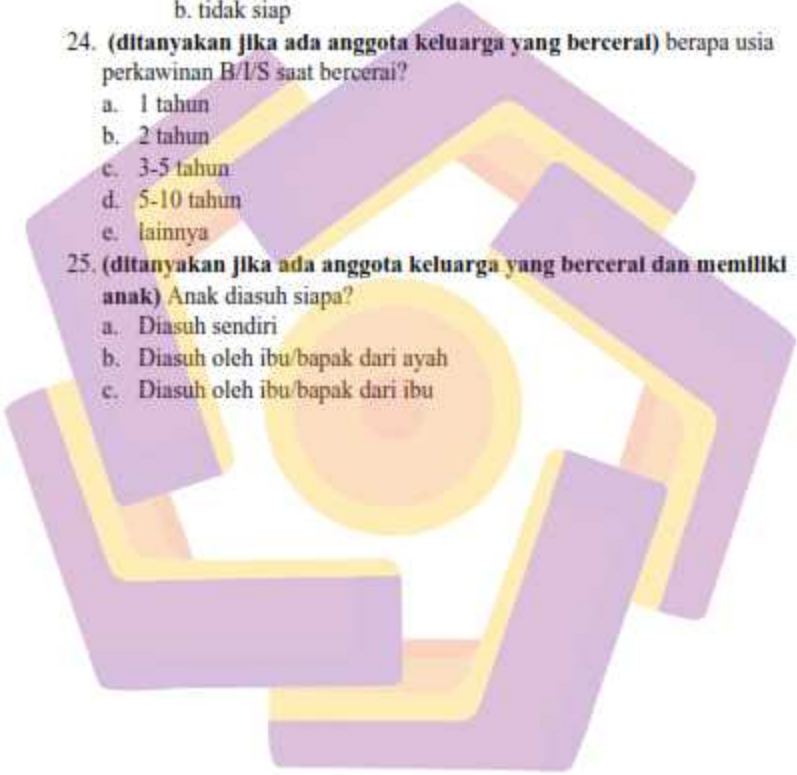
- a. Padi
- b. Palawija (jagung, kacang-kacangan, ubi-ubian)
- c. Hortikultura (buah-buahan, sayur-sayuran,
- d. Perkebunan (cengkeh, kakao, sawit, dll.)
- e. Peternakan (sapi, domba, ayam, dll.)
- f. Perikanan tangkap (termasuk biota lainnya)
- g. Perikanan budidaya (termasuk biota lainnya)
- h. Kehutanan (cemara, jati, pinus, bambu, damar, rotan, dll.)
- i. Jasa pertanian (pembenihan, sewa traktor, dll.)

E. Kesehatan

1. Apakah B/I/S termasuk pasangan usia subur (PUS) dengan umur 20 – 45 tahun?
 - a. Ya
 - b. tidak
2. Berapa umur Bapak ketika nikah (nikah yang pertama)?
 - a. 15 – 18 tahun
 - b. 19 – 21 tahun
 - c. 22 – 25 tahun
 - d. 25 – 30 tahun
 - e. Diatas 30 tahun
3. Berapa umur Ibu ketika nikah (nikah yang pertama)?
 - a. 15 – 18 tahun
 - b. 19 – 21 tahun
 - c. 22 – 25 tahun
 - d. 25 – 30 tahun
 - e. Diatas 30 tahun
4. Ketika Ibu hamil, berapa kali memeriksakan kehamilan ke posyandu atau bidan desa/kelurahan?
 - a. 1-2 kali
 - b. 3-4 kali
 - c. 5-6 kali
 - d. Di atas 6 kali
5. Berapa usia Ibu saat hamil pertama?
 - a. 15 – 18 tahun
 - b. 19 – 21 tahun
 - c. 22 – 25 tahun
 - d. 25 – 30 tahun
 - e. Diatas 30 tahun
6. Siapa yang membantu Ibu saat persalinan?

- 
- a. Ibu Bidan
 - b. Dukun
 - c. Doter
7. Apakah pada saat awal kehamilan, Ibu mendapatkan suntikan imunisasi TT?
 - a. Ya
 - b. Tidak
 8. Apakah pada saat awal kehamilan, Ibu minum tablet penambah darah?
 - a. Ya,
 - b. Tidak
 9. Berapa butir yang Ibu minum?
 - a. 1 - 20 butir
 - b. 21 - 40 butir
 - c. 41 - 60 butir
 - d. 61 - 80 butir
 - e. 81 - 90 butir
 10. Berapa usia Ibu saat melahirkan yang pertama?
 - a. 15 - 18 tahun
 - b. 19 - 21 tahun
 - c. 22 - 25 tahun
 - d. 25 - 30 tahun
 - e. Diatas 30 tahun
 11. Imunisasi apa saja yang Ibu peroleh pascalahir?
 - a. Campak
 - b. Gondok
 - c. Rubella
 - d. Cacar
 12. Asupan Gizi Ibu saat hamil.
 - a. Dsdsfdsfd
 - b. Fdsfsdf
 - c. Dfsdfsdf
 13. Apakah Ibu memberikan asi eksklusif?
 - a. Ya
 - b. Tidak
 14. Berapa umur anak Ibu ketika berhenti minum asi?
 - a. 0,5 bulan
 - b. 1 tahun
 - c. 1,5 tahun
 - d. 2 tahun

- 
15. Setelah melahirkan, apakah Ibu siap ber-KB?
- Ya
 - Tidak siap
16. Jenis alat kontrasepsi yang digunakan?
- Pil KB
 - Suntik KB
 - Kondom
 - KB IUD
17. Apakah B/I siapan mengikuti program kontak (MOW/MOP)
- Ya siap
 - Tidak siap
18. Jika Ibu melahirkan lagi, Berapa jarak antara anak pertama dengan anak berikutnya?
- 1 tahun
 - 2 tahun
 - 3 tahun
 - 4 tahun
 - 5 tahun
19. **(ditanyakan jika ada anggota keluarga yang usia remaja)** Rencana menikah pada usia berapa?
- 15 – 18 tahun
 - 19 – 21 tahun
 - 22 – 25 tahun
 - 25 – 30 tahun
 - Diatas 30 tahun
20. **(ditanyakan jika ada anggota keluarga yang usia remaja)** Jika menikah, rencana memiliki anak berapa?
- 1
 - 2
 - 3
 - 4
 - 5
21. **(ditanyakan jika ada anggota keluarga yang usia remaja)** Pekerjaan yang ditekuni saat ini?
- Tani
 - Buruh tani
 - Berdagang
 - Nelayan
 - lainnya

- 
22. **(ditanyakan jika ada anggota keluarga yang usia remaja)** Apakah anda tahu resiko melakukan hubungan seksual jika usia masih muda?
- ya tahu
 - tidak tahu
23. **(ditanyakan jika ada anggota keluarga yang usia remaja)** Jika sudah menikah dini, apakah siap menunda kehamilan
- ya siap
 - tidak siap
24. **(ditanyakan jika ada anggota keluarga yang bercerai)** berapa usia perkawinan B/I/S saat bercerai?
- 1 tahun
 - 2 tahun
 - 3-5 tahun
 - 5-10 tahun
 - lainnya
25. **(ditanyakan jika ada anggota keluarga yang bercerai dan memiliki anak)** Anak diasuh siapa?
- Diasuh sendiri
 - Diasuh oleh ibu/bapak dari ayah
 - Diasuh oleh ibu/bapak dari ibu