

Data Mining

Implementasi Algoritma Naive bayes dalam Klasifikasi Status Kesejahteraan Masyarakat Desa Gunungsari

Nur Syifa Fauzia, Raditya Danar Dana

Program Studi Manajemen Informatika, STMIK IKMI Cirebon, Cirebon, Indonesia

INFORMASI ARTIKEL

Diterima Redaksi: 19 Februari 2023

Revisi Akhir: 21 Maret 2023

Diterbitkan Online: 24 Maret 2023

KATA KUNCI

Data Mining; Klasifikasi; Naive Bayes

KORESPONDENSI

Phone: 083123506881

E-mail: fauziahsyifa241@gmail.com

ABSTRAK

Kesejahteraan merupakan suatu pencapaian tatanan dalam kehidupan sosial, material, maupun spiritual yang diliputi oleh rasa keselamatan serta ketentraman lahir batin yang memungkinkan setiap individu masyarakat untuk melakukan usaha-usaha dalam memenuhi kebutuhan hidupnya guna mendapatkan kehidupan yang layak serta mampu mengembangkan dirinya dalam melaksanakan fungsi sosialnya. Pemerintah mempunyai peranan penting dalam mensejahterakan warganya, berbagai upaya pembangunan telah dilakukan oleh pemerintah dengan tujuan utama untuk mencapai kesejahteraan masyarakat melalui berbagai program pembangunan. Namun dalam pelaksanaannya dirasa kurang berjalan efektif, sering kali ditemukan kesalahan salah satunya dalam menentukan kelayakan penerima bantuan. Dalam Implementasi pada Desa Gunungsari belum terdapatnya metode penilaian yang tepat dalam memastikan status kesejahteraan masyarakat. Konsep Data mining akan memudahkan dalam menangani permasalahan yang belum optimal salah satunya metode klasifikasi naive bayes yang mampu menciptakan model yang membedakan dalam memastikan status kesejahteraan masyarakat dengan memprediksi peluang di masa depan bergantung kepada kumpulan jenis data yang ada. Naive Bayes menghitung probabilitas bersumber pada kriteria-kriteria yang sudah ditetapkan. Metode ini menerapkan teknik supervised objek dengan menetapkan target kelas berdasarkan kategori sudah sejahtera dan belum sejahtera. Dari keseluruhan data yang berjumlah 1109 akan digunakan sebanyak 70% Sebagai data training, dan akan diuji sebanyak 30% menggunakan data testing. Variabel yang digunakan sebanyak 21 variabel. Hasil pengujian yang didapatkan menghasilkan akurasi sebesar 93,69 serta akurasi error sebesar 6,31%. Dengan adanya hasil prediksi kondisi penduduk setiap keluarga tersebut, diharapkan penelitian ini dapat memberikan informasi mengenai status kesejahteraan masyarakat dan dapat digunakan dalam membuat keputusan tepat dan akurat dan menjadi bahan pertimbangan yang bertujuan untuk kesejahteraan masyarakat.

PENDAHULUAN

Kesejahteraan merupakan suatu pencapaian tatanan dalam kehidupan sosial, material, maupun spiritual suatu masyarakat yang diliputi oleh rasa keselamatan, kesehatan dan ketentraman lahir batin yang memungkinkan setiap individu untuk melakukan usaha-usaha dalam rangka memenuhi kebutuhan hidupnya guna mendapatkan kehidupan yang layak serta mampu mengembangkan dirinya dalam melaksanakan fungsi sosialnya. Namun pada kenyataannya permasalahan yang berkaitan dengan kesejahteraan cenderung meningkat, baik dari segi kualitas maupun kuantitas, saat ini masih terdapat masyarakat yang belum dapat memenuhi kebutuhan dasarnya, hal ini disebabkan oleh terhambatnya ekonomi sosial, sehingga mengalami kesulitan dalam menikmati kehidupan yang layak.[1]. Pemerintah mempunyai peranan penting dalam mensejahterakan warganya dari kemiskinan, berbagai upaya pembangunan terus dilakukan oleh pemerintah dengan tujuan utama untuk mencapai kesejahteraan masyarakat melalui berbagai program pembangunan. Demikian, upaya mengonversikan kesejahteraan secara terus menjadi kajian di kalangan para ekonomi, namun dalam pelaksanaannya,

sering kali ditemukan kesalahan serta kekeliruan dalam menentukan status kesejahteraan masyarakat, salah satunya dalam kelayakan penerima bantuan dari pemerintah [2].

Pada dasarnya, pencapaian kesejahteraan masyarakat merupakan tujuan utama dari setiap pembangunan ekonomi. Pencapaian kesejahteraan di Indonesia merupakan salah satu tujuan negara kesatuan republik Indonesia (NKRI) sebagaimana yang dijelaskan dalam pembukaan UUD 1945 alinea ke-empat yaitu “melindungi segenap bangsa Indonesia dan seluruh tumpah darah Indonesia dan untuk memajukan kesejahteraan umum, mencerdaskan kehidupan bangsa”. Pemahaman dan teknik dalam mengukur kesejahteraan dalam membangun perekonomian terus mengalami perkembangan, kesejahteraan masyarakat menengah ke bawah dapat direpresentasikan dari tingkat hidup masyarakat. Tingkat hidup masyarakat ditandai oleh rentannya terhadap kemiskinan, tingkat pendapatan, perolehan tingkat Pendidikan serta tingkat produktivitas masyarakat. Konsep dasar dalam kesejahteraan yaitu sebagai padanan dari makna konsep martabat manusia yang dapat dilihat dari empat indikator yaitu terciptanya rasa aman, kehidupan yang sejahtera, memiliki kebebasan dalam mengekspresikan diri, serta mengenal jati dirinya [3].

Berdasarkan penelitian sebelumnya yang telah dilakukan oleh Arianto Tarigan dkk.2019 dalam sebuah artikel dengan judul “Klasifikasi Status Kesejahteraan Rumah Tangga di Kabupaten Siak Menggunakan Algoritma Naive Bayes Classifier” penelitian ini melakukan model Algoritma Naive Bayes dengan menggunakan 19.612 Data Kesejahteraan Rumah Tangga tahun 2017 yang diperoleh dari Dinas Sosial Kabupaten Siak, penelitian ini melakukan pembagian data latih dan data uji dengan menggunakan k-fold cross validation dengan melakukan 4 cross dan 10 kali percobaan menggunakan tools weka dan data kesejahteraan tahun 2017. Hasil dari prediksi menunjukkan bahwa akurasi tertinggi pada k4 dengan akurasi rata-rata 53%. Hasil dari perhitungan tersebut diimplementasikan kedalam sebuah sistem informasi yang telah dibangun. [2]. Penelitian sebelumnya juga dilakukan oleh Deny Novianty tahun 2019 dengan judul “Implementasi Algoritma Naive Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner” penelitian ini menerapkan algoritma naive bayes untuk memprediksi hepatitis menggunakan rapidminer, hasil penelitian menunjukkan bahwa metode naive bayes dengan rapidminer mampu memprediksi hasil akurasi sebesar 76,77% [4]. penelitian sebelumnya dilakukan oleh Made Ayu Dusea Widayada dkk.2019 dengan judul “Implementasi Metode Naive Bayes dalam penentuan tingkat kesejahteraan keluarga” penelitian ini menerapkan algoritma naive bayes dalam proses pendataan kesejahteraan masyarakat dan meminimalisir terjadinya kesalahan dalam pendataan yang dapat digunakan sebagai pendukung program dengan pengujian pada sistem. [5]

Dimensi kesejahteraan penduduk yang sangat luas dan kompleks ini mengakibatkan rentan terjadinya kesenjangan permasalahan sosial dalam bermasyarakat. Tepatnya pada wilayah Desa Gunungsari kecamatan Sadananya yang akan menjadi studi kasus dalam penelitian ini. Berdasarkan hasil observasi yang dilakukan, permasalahan ditemukan yaitu ketidakakuratan data dalam menentukan dan mengukur status kesejahteraan masyarakat, pemerintah menentukan tingkat kesejahteraan berdasarkan musyawarah dengan menentukan status kesejahteraan berdasarkan aspek-aspek yang telah ditentukan, namun dalam pelaksanaannya kurang berjalan secara efektif, seringkali ditemukan kesalahan salah satunya dalam memastikan kelayakan penerima bantuan, sehingga masih terdapat keluarga Penerima manfaat (KPM) dengan kategori sejahtera yang mendapatkan bantuan sedangkan keluarga yang belum sejahtera belum mendapat bantuan. Masalah lain yang ditemukan yaitu data yang masih abstrak dan banyaknya atribut yang digunakan dalam menentukan suatu kebijakan, hal ini akan memakan waktu yang lama jika masih menggunakan pengukuran secara manual.

Oleh karena itu metode klasifikasi naive bayes mampu menciptakan model serta mampu memprediksi peluang di masa depan dengan melakukan prediksi dalam memastikan suatu informasi bergantung kepada kumpulan jenis data yang ada. Naive Bayes menghitung probabilitas yang bersumber pada kriteria-kriteria yang sudah ditetapkan. Metode ini menerapkan teknik supervised klasifikasi objek dengan menetapkan label kelas berdasarkan kategori. Pengujian dalam perhitungan klasifikasi ini mengimplementasikan Aplikasi Rapidminer dengan menggunakan data training dan data testing dimana data training atau data latih digunakan untuk melatih algoritma serta mengembangkan dan mencari model dari kumpulan data yang sesuai. Sedangkan data testing digunakan untuk menguji dan mengetahui performa model yang didapatkan.

Berdasarkan uraian di atas dan didukung oleh penelitian sebelumnya yang menjadi acuan dasar maka dalam penelitian ini akan menentukan klasifikasi status kesejahteraan masyarakat pada Desa Gunungsari Kecamatan Sadananya menggunakan Algoritma Naive Bayes Classifier dengan mengimplementasikannya pada Aplikasi Rapid Miner, sistem

aplikasi ini dapat memberikan solusi dalam melakukan analisis terhadap data mining, text mining dan analisis prediksi dalam membuat keputusan terbaik, yang diharapkan dapat digunakan sebagai pengetahuan yang bermanfaat dan dapat membantu dalam menentukan status kesejahteraan masyarakat secara cepat dan akurat.

TINJAUAN PUSTAKA

Data Mining

Data mining merupakan proses dalam pengelolaan data yang jumlahnya besar untuk memperoleh informasi yang dapat dimanfaatkan untuk pengambilan keputusan. teknik data mining memanfaatkan pengalaman bahkan kesalahan data di masa lalu untuk meningkatkan kualitas model maupun hasil analisisnya dengan kemampuan yang dimiliki teknik data mining salah satunya klasifikasi. [6]

Klasifikasi

Klasifikasi merupakan proses untuk menemukan model dan fungsi yang menjelaskan atau membedakan konsep dengan tujuan untuk mendapatkan perkiraan kelas dari suatu objek yang labelnya tidak diketahui. Dalam prosesnya klasifikasi membentuk sebuah model yang mampu membedakan data ke dalam kelas yang berbeda-beda berdasarkan pada aturan serta fungsi tertentu.terdapat beberapa algoritma dalam klasifikasi meliputi, Logistic Regression, K-Nearesr Neighbors, Naive Bayes, Decission Tree, Random Forest, Neural Network. [7]

Naive Bayes

Naive Bayes merupakan pengklasifikasian probablisti sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari *dataset*. Teorema bayes mengasumsikan semua atribut independen serta tidak saling keterantungan terhadap nilai yang diberikan kepada kelaas variabel. Metode ini hanya membutuhkan jumlah data training yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikian. [8]

Persamaan Teorema Naive Bayes sebagai berikut:

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

Keterangan :

$P(X|Y)$ = Posterior|probability yaitu nilai probabilitas X berdasarkan kondisi Y

$P(Y|X)$ = probabilitas Y yang ditentukan X adalah benar

$P(X)$ = Peluang evidence penyakit X

$P(Y)$ = Probabilitas dari nilai Y

Confusion Matrix

Confusion Matrix merupakan suatu metode yang digunakan untuk mengukur akurasi pada data mining, dengan memiliki dua kelas yaitu kelas bersifat positif dan kelas bersifat negatif. Terdiri dari empat sel yaitu True Positif(TP), False Positif (FP), True Negatif (TN) dan False Negatif(FN). Perhitungan akurasi dilakukan dengan cara membagi jumlah data yang diklasifikasi secara benar dengan total sample data testing yang diuji. Dalam Confusion matterdapat precision, Recall, dan Kurva ROC. [9].

Tabel 1. Confusion matrix untuk 2 model kelas

CLASSIFICATION	PREDIKTED CLASS	
	Class=YES	Class=NO
	Class=YES	Class=NO
OBSERVED CLASS	a (true positive-TP)	b (false negative-FN)
	c (false positive-FP)	d (true negative-TN)

METODOLOGI

1. Sumber Data

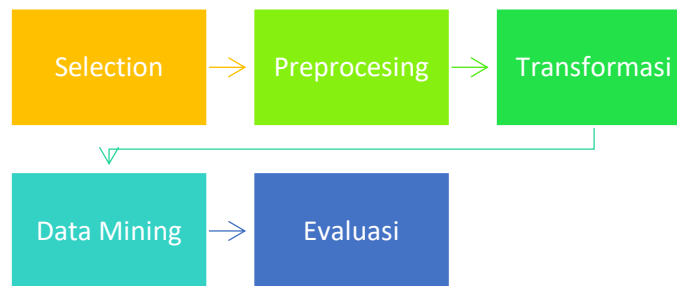
Dalam mengumpulkan data-data yang dapat menunjang penelitian ini menggunakan Sumber data Sekunder yaitu pengumpulan data yang diperoleh langsung dari sumber objek penelitian, pengumpulan data dilakukan dengan mempertimbangkan penggunaan data berdasarkan jenis dan data sumbernya. Data yang digunakan dalam penelitian ini yaitu data sensus penduduk desa gunungsari tahun 2022.

2. Teknik pengumpulan Data

Studi pustaka dalam Teknik pengumpulan data ini bersumber dari data sekunder yang telah terkumpul dalam sebuah file spreadsheet berekstensi .xls, dengan jumlah 1109 record data, dengan jumlah atribut 20 kolom dan 1 Label. Namun demikian data tersebut tidak dapat langsung digunakan dalam proses pemodelan menggunakan algoritma Naïve Bayes karena masih terdapat format data yang belum standar. Sehingga perlu dilakukan preprocessing data terlebih dahulu.

3. Tahapan perancangan

Tahapan Perancangan model yang digunakan dalam penelitian ini menggunakan Knowledge Discovery in Database (KDD) adalah proses yang bertujuan untuk menggali dan menganalisis data yang sangat besar menjadi informasi yang berguna untuk pengetahuan



Gambar 1. Tahapan perancangan KDD

HASIL DAN PEMBAHASAN

Dalam hasil penelitian ini akan membahas tahapan-tahapan penerapan data mining dengan metode algoritma Naive Bayes dan akan dilakukan pengujian metode Naive Bayes dengan menggunakan Rapidminer yang akan digunakan dalam klasifikasi status kesejahteraan masyarakat.

Data selection

Tahap seleksi data dilakukan dengan menganalisis data-data yang tidak relevan, *inkonsisten*, noise ataupun *redundant*. Dataset yang digunakan memiliki 1 target variabel sebagai kelas label yaitu status penduduk sudah sejahtera dan status penduduk belum sejahtera dan 20 atribut sebagai variabel predictor.

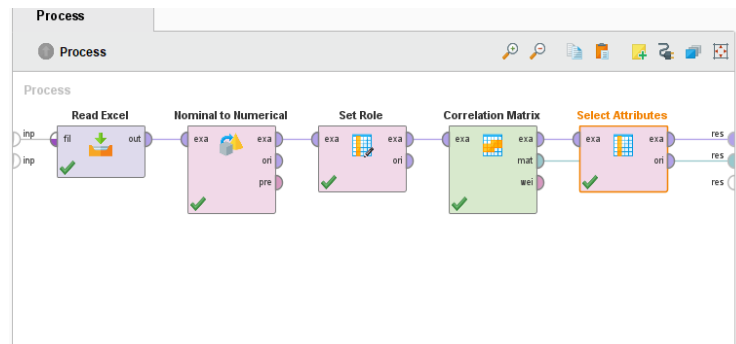
Berdasarkan tabel 4.1 dibawah ini Menerangkan atribut serta keterangan yang akan digunakan dalam proses data mining. Pada tahap ini dilakukan seleksi data dengan variabel-variabel prediktor dan satu target variable. Variabel-variabel prediktor yaitu nama keluarga, tempat tinggal, status lahan, jenis lantai, jenis dinding, atap, penerangan rumah, energi untuk memasak, fasilitas MCK, sumber air mandi, sumber air minum, rumah berada di bantaran sungai, rumah di lereng bukit, kondisi rumah, BLT, PKH, BST, BANPRES, Bantuan UMKM, bantuan pendidikan anak, sedangkan variabel yang akan dijadikan sebagai target label yaitu status.

Berikut table dibawah ini adalah data atribut yang digunakan sesuai dengan dataset awal yang diperoleh :

Tabel 2. Daftar Atribut

No	Nama Atribut	Sub-atribut
1	Nama keluarga	Nama kepala keluarga yang menjadi tanggungjawab
2	Tempat tinggal	- Bebas sewa - Milik sendiri - Sewa
3	Status lahan	- milik sendiri - milik orang lain
4	Jenis lantai	- Granit - Kayu - Keramik - Semen - Ubin
5	Jenis dinding	- Bilik - GRC - Tembok
6	Atap	- Genteng
7	Penerangan rumah	- Listrik PLN - Listrik Non PLN
8	Energi untuk memasak	- Gas LPG - Kayu bakar
9	Fasilitas MCK	- Sendiri - MCK umum - Berkelompok
10	Sumber air mandi	- Mata air - Perpipaan berbayar - Sungai
11	Sumber air minum	- Air isi ulang - Mata air
12	Rumah dibantaran sungai	- Ya - Tidak
13	Rumah dilembar bukit	- Ya - Tidak
14	Kondisi rumah	- Tidak kumuh
15	BLT	- Ya - Tidak
16	PKH	- Ya - Tidak
17	BST	- Ya - Tidak
18	Banpres	- Ya - Tidak
19	Bantuan UMKM	- Ya - Tidak
20	Bantuan pendidikan anak	- Ya - Tidak
21	Status	- Sudah sejahtera - Belum sejahtera

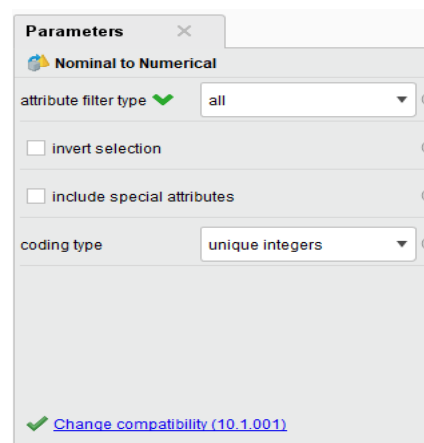
Proses select attribute dengan mengimplementasikan tools pada Rapid miner yang akan ditunjukkan pada gambar proses dibawah ini:



Gambar 2. Proses seleksi data

Adapun penjelasan dari setiap operator adalah sebagai berikut:

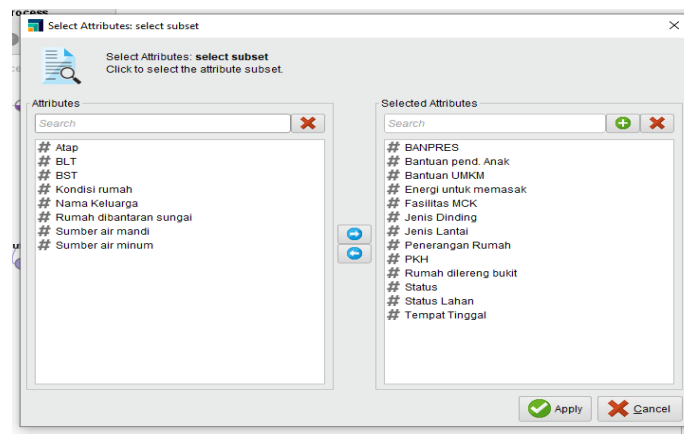
1. Read Excel, operator ini berfungsi untuk membaca data yang telah diimport. Data sensus penduduk diambil dari tempat observasi dan sudah dalam format excel.
2. Nominal To Numerical, operator ini digunakan untuk mengubah tipe data nominal menjadi tipe data numerik. Pada operator ini terdapat parameters “attribute filter type” akan dipilih all, karena dalam proses correlation matrix harus bernilai numerikal. Hasilnya semua atribut diubah menjadi angka 0 dan 1. Seperti pada gambar 3.



Gambar 3. Parameters Nominal to Numerical

3. Set role, operator ini digunakan untuk menentukan atribut kelas atau label dari data. Pada operator ini terdapat parameters “edit list” kemudian pilih “attribute name” digunakan untuk menentukan atribut yang akan dijadikan kelas atau label, pada data ini status yang akan menjadi kelas atau label, lalu pada parameters “target role” akan dipilih label yang digunakan untuk menentukan bahwa atribut tersebut adalah kelas label.
4. Correlation matrix, berfungsi untuk menentukan korelasi antar semua atribut yang dapat dihasilkan vektor bobot berdasarkan korelasi terkait. Korelasi ini merupakan teknik statistik yang dapat menunjukkan seberapa kuat korelasi antar atribut terkait. Korelasi merupakan angka antara -1 dan +1. Dalam hal ini nilai positif menyiratkan korelasi asosiasi positif sedangkan nilai negatif mengimplikasikan korelasi asosiasi negatif atau terbalik. Atribut yang terkorelasi harus dihapus karena hal ini akan berpengaruh saat perhitungan prediksi dan dapat menghambat proses berjalannya data mining. Berdasarkan hasil dari proses correlation matrix variabel prediktor terhadap variabel target label status terdapat korelasi yang bernilai negatif diantaranya nama keluarga, sumber air mandi, sumber air minum, rumah dibantaran sungai, BLT, dan BST, sedangkan pada atribut atap dan kondisi rumah memiliki nilai *redundant* artinya dalam dataset yang lebih dari satu *record* berisi nilai yang sama. Oleh karena itu atribut-atribut tersebut harus dihilangkan dan akan di proses dalam tahap select attribute.
5. Select attribute, operator ini digunakan untuk memilih atribut yang akan digunakan dalam pemrosesan data. Dalam penelitian ini data dari sensus penduduk akan dipilih dari 21 atribut menjadi 13 atribut, atribut yang digunakan meliputi : Tempat tinggal, Status lahan, Jenis lantai, Jenis dinding, Penerangan rumah, Energi untuk memasak, Fasilitas MCK, Rumah dilemang bukit, PKH, Banpres, Bantuan UMKM, Bantuan pendidikan anak

dan status. Emilihan atribut ini didasarkan pada hasil dari correlation matrix. Yang dapat dilihat pada gambar dibawah ini !

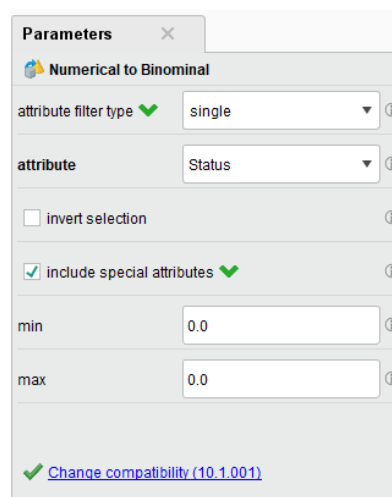


Gambar 4. Select attribute

Berdasarkan hasil dari proses correlation matrix variabel prediktor terhadap variabel target label status terdapat korelasi yang bernilai negatif diantaranya nama keluarga, sumber air mandi, sumber air minum, rumah dibantaran sungai, BLT, dan BST, sedangkan pada atribut atap dan kondisi rumah memiliki nilai *reducent* artinya dalam dataset yang lebih dari satu *record* berisi nilai yang sama. Oleh karena itu atribut-atribut tersebut harus dihilangkan dan akan di proses dalam tahap select attribute.

Data Transformation dan preprocessing

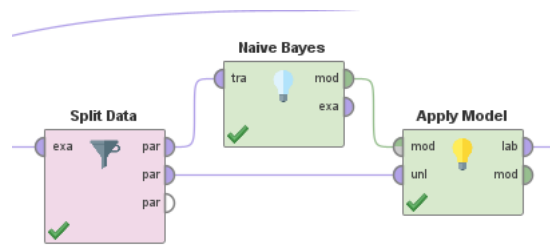
Pada tahap transformation ini hasil dari data selection akan ditransformasikan menggunakan operator numerikal to binominal, operator ini berfungsi mengubah atribut numerik yang dipilih menjadi tipe binominal atau biner, operator ini akan memetakan nilai target atribut yaitu status yang dijadikan sebagai label ke nilai binominal yang memiliki dua kemungkinan nilai “true” dan “false”



Gambar 5. Parameter numerical to Binominal

Data Mining

Proses Data Mining merupakan tahapan dalam menemukan pola atau informasi dari sekumpulan data dengan menggunakan metode klasifikasi dan algoritma naive bayes untuk menemukan informasi mengenai data sensus penduduk. Proses data mining akan diimplementasikan secara aplikatif menggunakan rapidminer studio. Operator yang digunakan dalam pemodelan mining ini dapat dilihat pada gambar dibawah ini:



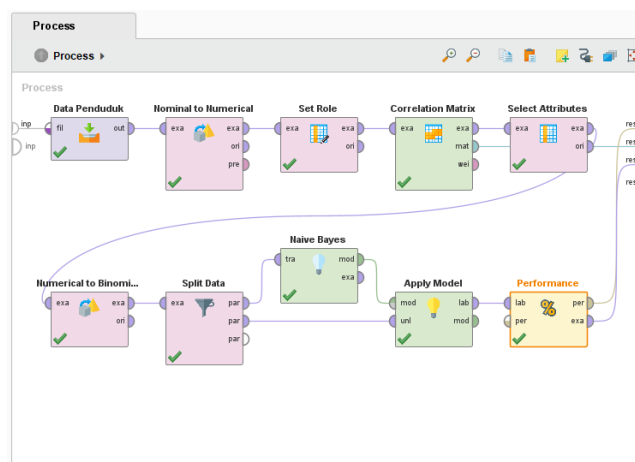
Gambar 6. Proses pemodelan Algoritma

Split Data operator ini berfungsi untuk memisahkan dataset dengan membagi menjadi dua bagian, dalam mendapatkan model yang sebaik mungkin maka rasio dalam data tarining harus lebih besar dari rasio data testing, maka pada praktik ini akan membagi 70% dari dataset sebagai data training dan 30% dari dataset

Data Evaluation

Proses klasifikasi dengan rapid miner menggunakan algoritma naive bayes pada data sensus penduduk dengan membandingkan data testing dengan data training yang sudah diketahui dari rule-rulanya sebelumnya, berikut langkah-langkahnya :

Pada tampilan process masukan operator read excel untuk membaca data dari file excel, kemudian ubah nilai data nominal menjadi numerik dengan menggunakan operator nominal to numerical selanjutnya pilih atribut yang akan dijadikan sebagai target label, untuk menentukan korelasi antar atribut digunakan operator correlation matrix yang akan menjadi penentu dalam proses select attribute, setelah melakukan select attribute selanjutnya operator mengubah atribut numerik yang dipilih menjadi tipe binominal atau biner, operator ini akan memetakan nilai target atribut yaitu status yang dijadikan sebagai label ke nilai binominal yang memiliki dua kemungkinan nilai “true” dan “false”, kemudian melakukan pemisahan dataset dengan membagi menjadi data training dan data testing yang selanjutnya akan di masukan kedalam proses pemodelan algoritma naive bayes dan apply model, hasil akhir akan dilakukan pengukuran akurasi dengan menggunakan operator performance.



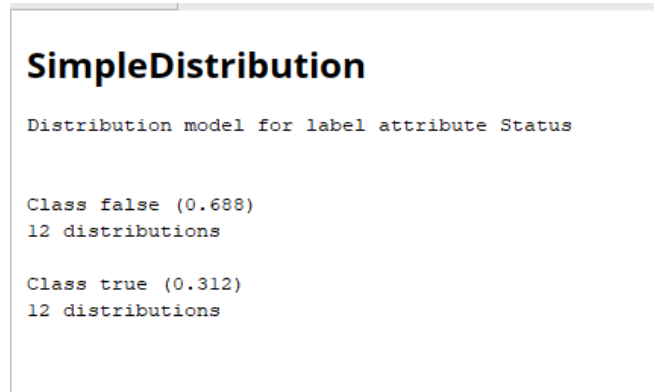
Gambar 7. proses accuracy prediksi

accuracy: 93.69%

	true false	true true	class precision
pred. false	214	6	97.27%
pred. true	15	98	86.73%
class recall	93.45%	94.23%	

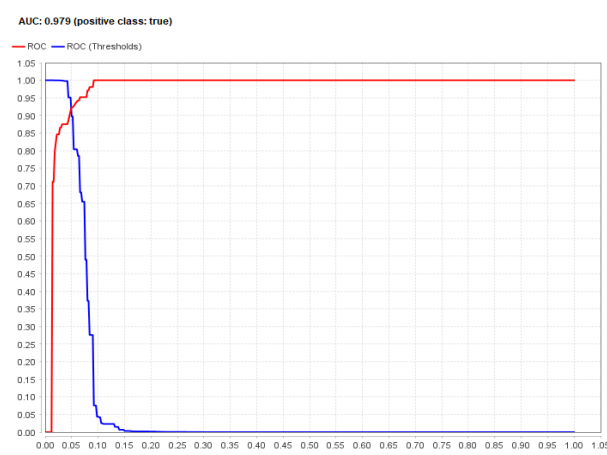
Gambar 8. Hasil Akurasi performance

Hasil pengukuran accuracy yang diperoleh dari data training mencapai 93,69%. Jumlah prediksi false(sudah sejahtera) yang diklasifikasikan sudah sejahtera oleh classifier yaitu 214 data dan jumlah prediksi sudah sejahtera yang diklasifikasikan true (belum sejahtera) sebagai belum sejahtera oleh classifier yaitu 6 dengan mencapai eclass precission 97,27%. Sedangkan jumlah prediksi belum sejahtera yang diklasifikasikan sebagai sudah sejahtera oleh classifier yaitu 15 data, dan jumlah prediksi belum sejahtera yang diklasifikasikan sebagai sudah sejahtera oleh classifier yaitu 98 data, dengan pencapaian class precision 86,73%. Untuk kelas recall dengan true false (sudah sejahtera) mencapai 93,45%, sedangkan untuk class recall dengan true true (belum sejahtera) mencapai 95,23%.



Gambar 9. Simple distribution

Berdasarkan gambar simple ditribution menunjukkan bahwa jika nilai class mendekati angka 1 maka hasilnya termasuk dalam kelas sudah sejahtera



Gambar 10. Kurva ROC

Kurva ROC menunjukkan akurasi dan perbandingan secara visual dan ROC akan mengekspresikan *confusion matrix*. Tingkat keakurasian AUC diklasifikasikan menjadi lima kelompok yaitu :

1. 0.90-1.00 = unggul
2. 0.80-0.90 = Baik
3. 0.70-0.80=Cukup
4. 0.60-0.70=Kurang
5. 0.50-0.60=Gagal

Nilai AUC yang dihasilkan dalam proses pengujian berdasarkan kurva ROC menggunakan metode naive bayes sebesar 0,979% berdasarkan kriteria dalam tingkat keakurasian AUC menunjukkan klasifikasi yang dihasilkan termasuk kedalam kelompok unggul.

PerformanceVector

```

PerformanceVector:
accuracy: 93.69%
ConfusionMatrix:
True:  false  true
false:  214    6
true:   15    98
classification_error: 6.31%
ConfusionMatrix:
True:  false  true
false:  214    6
true:   15    98
AUC (optimistic): 0.989 (positive class: true)
AUC: 0.979 (positive class: true)
AUC (pessimistic): 0.979 (positive class: true)
precision: 86.73% (positive class: true)
ConfusionMatrix:
True:  false  true
false:  214    6
true:   15    98
recall: 94.23% (positive class: true)
ConfusionMatrix:
True:  false  true
false:  214    6
true:   15    98

```

Gambar 11. Deskripsi performance vektor

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil pengujian dalam penilaian akurasi status kesejahteraan masyarakat menggunakan algoritma naive bayes, variable data yang digunakan yaitu Tempat Tinggal, Status lahan, Jenis lantai, Jenis dinding, Penerangan Rumah, Energi Untuk Memasak, Tempat Pembuangan sampah, Fasilitas MCK, Sumber air mandi, sumber air minum, rumah berada di bantaran sungai, rumah di lereng bukit, BLT, PKH, BST, BANPRES, Bantuan UMKM, Bantuan Pendidikan Anak. Metode Algoritma Naive bayes dapat diimplementasikan dan mampu mengidentifikasi *class* kesejahteraan masyarakat dengan akurasi hasil prediksi sebesar 93,69%. Sehingga sangat cocok diterapkan dalam memprediksi peluang dimasa yangnng akan datang berdasarkan kepada pengalaman dimasa sebelumnya sehingga dapat memudahkan pemerintah desa dalam menentukan kesejahteraan masyarakat.

Saran

Untuk penelitian selanjutnya diharapkan dapat dilakukan pengukuran performa classification dengan teknik yang berbeda. Untuk penelitian selanjutnya diharapkan dapat menerapkan algoritma lain seperti algoritma decision tree, algoritma neural network, algoritma support vektor machine agar dapat mengembangkan penelitian-penelitian dimasa yang akan datang. Penelitian ini dapat dikembangkan dengan membandingkan algoritma data mining yang lain atau dapat juga mengoptimalkan dengan parameter lainnya.

DAFTAR PUSTAKA

- [1] A. Khairi, A. Fais Ghozali, dan A. Darul Nur Hidayah, "IMPLEMENTASI K-NEAREST NEIGHBOR (KNN) UNTUK KLASIFIKASI MASYARAKAT PRA SEJAHTERA DESA SAPIKEREK KECAMATAN SUKAPURA," vol. 2, no. 3, hlm. 319–323, 2021.
- [2] M. Ayu, D. Widyadara, dan R. H. Irawan, "Implementasi Metode Naïve Bayes Dalam Penentuan Tingkat Kesejahteraan Keluarga," 2019.
- [3] W. F. T. F. Gorahe Vivi Lusya *dkk.*, "Dampak Pandemi Covid 19 Terhadap Kesejahteraan Masyarakat di Desa," 2021. [Daring]. Tersedia pada: <https://www.kemkes.go.id/>,
- [4] D. Novianti, S. Nusa, M. Jakarta, dan C. Sitasi, "Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner," vol. 21, no. 1, 2019, doi: 10.31294/p.v20i2.
- [5] M. Ayu, D. Widyadara, dan R. H. Irawan, "Implementasi Metode Naïve Bayes Dalam Penentuan Tingkat Kesejahteraan Keluarga," 2019.

- [6] R. L. Hasanah, M. Hasan, W. E. Pangesti, F. F. Wati, dan W. Gata, “Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode Knn (K-Nearest Neighbor),” *Jurnal Techno Nusa Mandiri*, vol. 16, no. 1, hlm. 1–6, 2019, doi: 10.33480/techno.v16i1.25.
- [7] B. S. Prakoso, D. Rosiyadi, D. Aridarma, H. S. Utama, dan F. Fauzi, “OPTIMALISASI KLASIFIKASI BERITA MENGGUNAKAN FEATURE INFORMATION GAIN UNTUK ALGORITMA NAIVE BAYES TERHUBUNG,” vol. 15, no. 2, hlm. 211–218, 2019, doi: 10.33480/pilar.v15i2.684.
- [8] C. A. Sugianto, F. R. Maulana, dan D. Mining, “Algoritma Naïve Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai (Studi Kasus Kelurahan Utama),” vol. 18, no. 4, hlm. 321–331, 2019.
- [9] B. Utami dan P. Aliandu, “KLASIFIKASI PENENTUAN TIM UTAMA OLAHRAGA HOCKEY MENGGUNAKAN ALGORITMA C4.pdf,” *The Proceedings of International Conferences of Information, Communication, Technology, and Systems*, vol. 5, no. 4, hlm. 1–5, 2013.