Printer : Jurnal Pengembangan Rekayasa Informatika dan Komputer
Vol. 1 No. 2 Nopember 2023
Hal. 88 – 102
e-ISSN 3052-9142
DOI : 10.29408/jprinter.v1i2.22007 **URL : https://doi.org/ 10.29408/jprinter.v1i2.22007**

Prediksi Tingkat Kesehatan Masyakarat Kecamatan Suralaga Berdasarkan Penggunaan Alat Kontrasepsi Menggunakan Algoritma Random Forest

Rabiatul Adawiyah[1*], Yahya[2], Muhammad Saiful[3]

[1,2,3]Program Studi Sistem Informasi, Universitas Hamzanwadi

*rabiatulc27gmail.com

Abstrak

Kesehatan masyarakat adalah ilmu dan seni untuk mencegah penyakit, memperpanjang masa hidup, dan meningkatkan kualitas sumber daya manusia. Dalam penelitian ini permasalahan yang akan dikaji adalah penggunaan alat kontrasepsi yang digunakan Keluarga berancana karena berada di tingkat yang masih rendah. Berdasarkan hal tersebut, diperlukan adanya suatu tindakan dalam menentukan keputusan yang diambil untuk mengetahui alat kontrasepsi yang digunakan salah satunya dengan cara memanfaatkan data yang ada. Pengolahan data dilakukan menggunakan algoritma Random Forest untuk mendapatkan hasil atau performa yang terbaik dalam memprediksi. dari dataset yang ada kita menggunakan 10 atribut yaitu nama lengkap, umur, pendidikan terakhir, pasangan usia subur, usia hamil pertama,usia pertama melahirkan, siap kb, alat kontrasepsi, mengikuti program kontap, dan jarak melahirkan. Hasil accuracy yang didapatkan sebanyak 71,99% dan hasil AUC yang didapatkan sebanyak 0,680% dengan output pohon keputusan bahwa alat kontrasepsi yang digunakan yaitu KB Iud, Pil KB dan Suntik KB.

Kata kunci : *Alat Kontrasepsi, Kesehatan Masyarakat, Random Forest.*

Abstract

Public health is the science and art of preventing disease, extending life span, and improving the quality of human resources. In this research, the problem that will be studied is the use of contraception used by planning families because it is still at a low level. Based on this, it is necessary to take action to determine the decision taken to find out which contraceptive device to use, one of which is by utilizing existing data. Data processing is carried out using the Random Forest algorithm to get the best results or performance in predicting. From the existing dataset we use 10 attributes, namely full name, age, highest level of education, partner of childbearing age, age at first pregnancy, age at first birth, ready for family planning, contraception, following a contact program, and birth spacing. The accuracy results obtained were 71.99% and the AUC results obtained were 0.680% with the decision tree output that the contraceptives used were IUD, birth control pills and birth control injections.

Keywords : *Contraception, Public Health, Random Forest.*

Printer : Jurnal Pengembangan Rekayasa Informatika dan Komputer
Vol. 1 No. 2 Nopember 2023
Hal. 88 – 102
e-ISSN 3052-9142
DOI : 10.29408/jprinter.v1i2.22007 **URL : https://doi.org/ 10.29408/jprinter.v1i2.22007**

1. Pendahuluan

Kecamatan Suralaga merupakan kecamatan yang berada di Kabupaten Lombok Timur dan mempunyai 15 desa. Kecamatan Suralaga terdapat 2 pusat kesehatan masyarakat atau puskesmas yang berada di desa kerongkong dan desa suralaga, dengan jumlah penduduk sebanyak 69.394 jiwa dan jumlah Kepala Keluarga sebanyak 26.311. Kesehatan masyarakat adalah ilmu dan seni untuk mencegah penyakit, memperpanjang masa hidup, dan meningkatkan kualitas sumber daya manusia. Di kecamatan Suralaga sendiri memiliki masalah kesehatan terhadap penggunaan alat kontrasepsi yang masih rendah dilihat dari data pada tahun 2019 jumlah kepala keluarga yang menggunakan alat kontrasepsi sebanyak 790 sedangkan jumlah pertumbuhan penduduk setiap tahunnya terus mengalami peningkat sebanyak 71.803, tahun 2021 72.280 dan tahun 2022 69.394.

Berdasarkan jumlah data yang ada bahwa jumlah pertumbuhan terus mengalami peningkatan setiap tahunnya maka pemerintah melakukan tindakan untuk meminimalisir jumlah pertumbuhan penduduk dengan membuat program Keluarga Berencana menggunakan Alat Kontrasepsi. Pemerintah mulai menekan terhadap jumlah pertumbuhan penduduk dengan membatasi jumlah anak melalui program *Keluarga Berencana* dengan menggunakan *alat kontrasepsi.*Pengguna alat kontrasepsi memiliki keluhan yang dirasakan para pengguna KB yang membuat pengguna KB merasakan keraguan dalam menggunakan alat kontrasepi. Sebenarnya efek samping yang ditimbulkan tidak berbahaya, akan tetapi sering membuat penggunanya merasa tidak nyaman. Efek samping yang sering ditimbulkan pada pengguna alat kontrasepsi yaitu gangguan pada siklus menstruasi, perubahan pada berat badan, mual/muntah, pusing/sakit kepala, timbulnya jerawat dan flek hitam diwajah.

Dalam penelitian ini permasalahan yang dikaji adalah penggunaan alat kontrasepsi dikalangan ibu-ibu yang menggunakan KB alat kontrasepsi karena berada di tingkat yang masih rendah, dimana jumlah pengguna alat kontrasepsi masih sangat sedikit dimana dapat meningkat jumlah pertumbuhan setiap tahunnya. Jenis-jenis alat kontrasepsi yang digunakan adalah KB IUD, SUNTIK KB, KONDOM dan PIL KB. Berdasarkan hal tersebut, diperlukan adanya suatu tindakan dalam menentukan keputusan yang diambil untuk mengetahui jenis-jenis alat kontrasepsi yang lebih banyak digunakan

Printer : Jurnal Pengembangan Rekayasa Informatika dan Komputer
Vol. 1 No. 2 Nopember 2023
Hal. 88 – 102
e-ISSN 3052-9142
DOI : 10.29408/jprinter.v1i2.22007 URL : https://doi.org/ 10.29408/jprinter.v1i2.22007

salah satu dengan cara memanfaatkan data yang ada. Dengan seperti itu, Pemerintah Kecamatan Suralaga dapat menentukan dan memilih strategi penanggulangan jumlah penggunaan alat kontrasepsi berdasarkan jenis-jenis alat kontrasepsi.

Dalam hal ini penulis melakukan pengolahan data mining menggunakan algoritma Random Forest yang dapat melakukan prediksi dan klasifikasi menggunakan penggabungan beberapa pohon keputusan dimana setiap nodenya mempresentasikan nilai dari atribut, pohon keputusan yang dihasilkan dengan menguji data yang ada.

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Adapun penelitian sebelumnya yang menggunakan algoritma Random Forest sebagai metode dalam pengolahan data sebagai pendukung penelitian kali ini, antara lain:

1. Penelitian yang dilakukan oleh Yuri Yuliani dalam jurnal yang berjudul "Algoritma Random Forest Untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung Menggunakan Seleksi Fitur Bestfirst" dapat disimpulkan bahwa dari hasil penelitian yang menggunakan aplikasi weka dengan melakukan seleksi fitur dengan metode bestfirst serta metode class balancer untuk menangani class yang tidak balance dan perbandingan terhadap 3 algoritma yang menunjukkan performa terbaik yaitu algoritma random forest dengan metode percentage split 80% yang menghasilkan accurasi 91,45%, mean absolute error 0.1874, incorrectly classified instances 8.55%, precision 0.915, recall 0.914, AUC 0.953[1].
2. Penelitian yang dilakukan oleh Yahya dan Hariman Bahtiar dalam jurnal yang berjudul "Pengaruh Pertumbuhan Ekonomi Terhadap Tingkat Kesejahteraan Masyarakat Kecamatan Selong Kabupaten Lombok Timur - Nusa Tenggara Barat Menggunakan Algoritma Naïve Bayes" berdasarkan hasil penelitian yang dilakukan menggunakan data ekonomi masyarakat Kecamatan Selong menggunakan Algoritma Naive Bayes, dapat disimpulkan bahwa masyarakat Kecamatan Selong merupakan masyarakat yang tergolong masyarakat yang sudah sejahtera apabila dilihat dari faktor ekonomi, dengan tingkat akurasi kedekatan dengan keadaan yang sebenarnya sebesar 93,45%. Dari data yang diperoleh 93,45% atau 0,9345 x jumlah data (kk) = 0,9345 x 1130 kk = 1056 kk yang menunjukkan

DOI : 10.29408/jprinter.v1i2.22007 URL : https://doi.org/ 10.29408/jprinter.v1i2.22007

masyarakatnya sejahtera dan 6,55% x 1130 = 74 kk yang menyatakan masyarakat yang tidak sejahtera dan dapat dijadikan sebagai acuan dalam pengentasan kemiskinan melalui program yang sudah dicanangkan oleh pemerintah[2].

3. Penelitian yang dilakukan oleh Mahpuz Yahya dan Muhammad Wasil dalam jurnal yang berjudul "Implementasi Algoritma Decision Tree Untuk Mengetahui Faktor Kredit Macet Dan Lancar di Koperasi Serba Usaha Daruzzakah Rensing Lombok Timur" dapat disimpulkan berdasarkan hasil pengujian yang telah dilakukan sebanyak 9 kali yaitu dari validation 2 sampai dengan 10 makadidapatkan dua hasil akurasi terbaik dari K-Fold Validation 9 dan K-Fold Validation 10 yang masing-masing memiliki nilai accuracy sebesar 96,43% pada K-Fold Validation 9 dan 96,45% dari K-Fold Validation 10. Dan selisih akurasi keduanya adalah 0.02% yang menunjukan bahwa terjadi peningkatan 0.02% pada rentang k-fold validation 9 dan 10. Peningkatan pada setiap uji coba tidaklah selalu sama ini dipengaruhi oleh seberapa banyak pembagian data yang dilakukanMaka didapatkan nilai accuracy terbaik pada KFold Validation 10 dengan akurasi sebesar 96,45% dimana data dibagi menjadi 10 bagian untuk ditraining dan ditesting. serta didapatkan juga nilai AUC (Area Under Curve) sebesar 0.942 yang dimana AUC merupakan parameter yang digunakan dalam analisis klasifikasi dalam menentukan model terbaik untuk prediksi suatu kelas atau atribut, AUC sendiri mempunyai rentang nilai dari 0 sampai 1, yang artinya semakin nilai AUC mendekati angka 1 maka prediksi atau diagnosa atributnya semakin bagus. Ini menjunjukan nilai AUC sebesar 0.942 yang didapatkan setelah melakukan pengujian dengan K-Fold Validation 10 sangat baik karena hampir mendekati angka 1[3].

4. Penelitian yang dilakukan oleh Muhammad Saiful, Syamsuddin dan Moh. Farid Wajdi dalam jurnal yang berjudul "Implementasi Algoritma Naive Bayes Untuk Memprediksi Predikat Ketuntasan Belajar Siswa Pasca Pandemi Covid 19" berdasarkan implementasi dan pembahasan dalam penelitian dilakukan untuk memprediksi ketuntsan belajar Siswa SMA Negeri 3 Selong, selama pandemic pasca covid 19, maka dapat disimpulkan bahwa dalam menentukan

Printer : Jurnal Pengembangan Rekayasa Informatika dan Komputer
Vol. 1 No. 2 Nopember 2023
Hal. 88 – 102
e-ISSN 3052-9142
DOI : 10.29408/jprinter.v1i2.22007 URL : https://doi.org/ 10.29408/jprinter.v1i2.22007

data tuntas dan tidak tuntas dapat diprediksi dan dievaluasi dengan memanfaatkan teknik data mining menggunakan algoritma Naïve Bayes. Algoritma ini mampu menganalisa pola data tuntas dan tidak tuntas, dalam mengambil kebijakan untuk melengkapi data siswa yang di katagorikan tuntas dan tidak tuntas dengan memanfaatkan teknik data mining,dimana dilakukan dengan cara melakukan menampilkan output berupa nilai akurasi terbaik dan akurat. Setelah dilakukannya pengujian sebanyak 9 kali yaitu dari validation 2 sampai dengan 10 maka didapatkan dua hasil akurasi terbaik dari K-Fold Validation 4 dan K-Fold Validation 5 yang masing-masing memiliki nilai accuracy sebesar 83.89% pada K-Fold Validation 4 dan 82.74% dari K-Fold Validation 5. Dan selisih akurasi keduanya adalah 0.01% yang menunjukan bahwa terjadi peningkatan 0.80 % pada rentang k-fold validation 4 dan 8 . Peningkatan pada setiap uji coba tidak selalu sama ini dipengaruhi oleh seberapa banyak pembagian data yang dilakukan[4].

5. Penelitian yang dilakukan oleh Evy Tri Susanti dan Haniva Lukma Sari dalam jurnal yang berjudul "Pendidikan Kesehatan Tentang Jenis-Jenis Alat Kontrasepsi Terhadap Pemilihan Alat Kontrasepsi" berdasarkan hasil penelitian bahwa sebelum dilakukan pendidikan kesehatan tentang jenis-jenis kontrasepsi, pengetahuan Ny. L masih kurang, dapat menjawab 6 dari 15 pertanyaan. Setelah diberikan pendidikan kesehatan pengetahuan baik, dapat menjawab 12 dari 15 pertanyaan. Ny. L dapat memilih alat kontrasepsi IUD yang akan digunakan dengan alasan bukan obat, tidak menimbulkan kegemukan, dapat digunakan jangka lama, tidak mengganggu menstruasi dan masa subur cepat kembali[5].

2.2. Random Forest

Random Forest adalah algoritma klasifikasi dan regresi yang menjadi bagian dari kelompok ensemble learning. Metode random forest merupakan pengembangan dari decision tree dimana setiap decision tree telah dilakukan proses pelatihan dengan menggunakan sampel individu. Random forest yang dihasilkan memiliki banyak tree dan setiap tree ditanam dengan cara yang sama. Seiring dengan bertambahnya dataset, maka tree juga ikut berkembang[6]. Dalam random forest, Pemilihan atribut pada setiap kali sebuah node akan dipecah akan diambil secara acak. Setiap tree diberi sampel data

Printer : Jurnal Pengembangan Rekayasa Informatika dan Komputer
Vol. 1 No. 2 Nopember 2023
Hal. 88 – 102
e-ISSN 3052-9142
DOI : 10.29408/jprinter.v1i2.22007 URL : https://doi.org/ 10.29408/jprinter.v1i2.22007

pelatihan dengan menggunakan metode bagging dan tiap tree dibangun menggunakan metode yang sama untuk membangun CART (classification and regression tree).

Random forest yang dihasilkan memiliki banyak tree dan setiap tree akan tumbuh dengan cara yang sama. Tree dengan variabel x akan ditempatkan pada jarak yang jauh dengan tree dengan variabel y. Sejalan dengan bertambahnya dataset maka tree pun ikut berkembang. Penempatan tree yang saling berjauhan akan memudahkan dalam deteksi jenis tree. Tree yang berada disekitar tree x maka tree tersebut merupakan perkembangan dari tree x sedangkan tree yang berada disekitar tree y maka tree tersebut merupakan perkembangan dari tree y. Pembangunan tree akan berhenti ketika data sudah homogen atau jika batas jumlah data minimum sudah terlewati.

2.3. Prediksi

Dalam arti non-statistik, istilah "prediksi" sering digunakan untuk merujuk pada tebakan atau opini yang diinformasikan. Prediksi semacam ini mungkin diinformasikan oleh penalaran abduktif, penalaran induktif, penalaran deduktif, dan pengalaman seseorang yang memprediksi ; dan mungkin berguna jika orang yang memprediksi adalah orang yang berpengetahuan luas di bidangnya. The metode adalah teknik untuk memunculkan prediksi berbasis ahli-penilaian seperti dengan cara yang terkontrol. Jenis prediksi ini mungkin dianggap konsisten dengan teknik statistik dalam arti bahwa, minimal, "data" yang digunakan adalah pengalaman kognitif pakar prediksi yang membentuk "kurva probabilitas" intuitif.

Dalam statistika, Prediksi merupakan bagian dari inferensi statistik. Salah satu pendekatan khusus untuk inferensi semacam itu dikenal sebagai inferensi prediktif, tetapi prediksi dapat dilakukan dalam salah satu dari beberapa pendekatan untuk inferensi statistik. Memang satu gambaran statistik yang mungkin adalah bahwa statistik menyediakan sarana untuk mentransfer pengetahuan tentang sampel populasi ke seluruh populasi, dan populasi terkait lainnya, yang tidak selalu sama dengan prediksi dari tahun ke tahun. Ketika informasi ditransfer lintas waktu, seringkali ke titik waktu tertentu, prosesnya dikenal sebagai peramalan. Peramalan biasanya membutuhkan metode time series, sedangkan prediksi sering dilakukan pada data cross-sectional.

Teknik statistik yang digunakan untuk prediksi termasuk analisis regresi dan berbagai sub-kategorinya seperti regresi linier, model linier umum (regresi logistik ,

Printer : Jurnal Pengembangan Rekayasa Informatika dan Komputer
Vol. 1 No. 2 Nopember 2023
Hal. 88 – 102
e-ISSN 3052-9142
DOI : 10.29408/jprinter.v1i2.22007 URL : https://doi.org/ 10.29408/jprinter.v1i2.22007

regresi Poisson , regresi Probit), dll. Dalam hal peramalan, model rata-rata bergerak autoregresif dan model regresi vektor dapat dimanfaatkan. Saat ini dan/atau terkait, kumpulan regresi atau metode pembelajaran mesin yang diterapkan dalam penggunaan komersial, bidang ini dikenal sebagai analitik prediktif[7].

2.4. Confusion Matrix

Confusion Matrix juga dikenal sebagai error matrix merupakan tata letak tabel khusus yang memungkinkan kinerja suatu algoritma, biasanya pembelajaran yang diawasi (dalam pembelajaran tanpa pengawasan biasanya disebut matriks yang cocok). Setiap baris matriks mewakili instance di kelas aktual sementara setiap kolom mewakili instance di kelas yang diprediksi, atau sebaliknya – kedua varian ditemukan dalam literatur. Nama berasal dari fakta bahwa memudahkan untuk melihat apakah sistem ini membingungkan dua kelas (yaitu biasanya salah memberi label satu sama lain). Dalam penelitian ini menggunakan perhitungan akurasi dimana jumlah data yang telah diklasifikasikan secara benar dan salah dibagi dengan total sampel data testing[7].

Tabel 1. Confusion matrix

	Actual Values	
Predicted values	TP	FP
	FN	TN

- TP (True Positive), yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.
- TN (True Negative), yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
- FN (False Negative), yaitu jumlah data negative namun terklasifikasi salah oleh sistem.
- FP (False Positive), yaitu jumlah data positif namun terklasifikasi salah oleh sistem.

Dari tabel confusion matrix tersebut kita bisa mencari nilai akurasi, precision dan recallnya.

2.5. Area Under The Curve (AUC)

Area Under The Curve adalah daerah yang terdapat di bawah Receiver Operating Characteristic (ROC). Dimana ROC adalah sebuah kurva hasil dari proses antara sensitivitas dan spesifikasi dari berbagai titik potong. Secara teori nilai AUC adalah antara 0 dan 1. Nilai AUC adalah nilai yang memberikan gambaran pengukuran terhadap kesesuaian yang dihasilkan dari algoritma yang digunakan, semakin besar nilai area under curve maka nilai variable

Printer : Jurnal Pengembangan Rekayasa Informatika dan Komputer
Vol. 1 No. 2 Nopember 2023
Hal. 88 – 102
e-ISSN 3052-9142
DOI : 10.29408/jprinter.v1i2.22007 URL : https://doi.org/ 10.29408/jprinter.v1i2.22007

yang diteliti untuk memprediksi semakin baik.

2.6. Rapid Miner

RapidMiner adalah salah satu software yang digunakan dalam proses data mining. Keunggulan rapidminer ini adalah mampu menerapkan dengan berbagai algoritma dan penyatuan fitur-fitur visualisasi data. Rapidminer sangat mudah dan efisien untuk proses komputasi dengan waktu yang relatif cepat dibanding dengan software lainnya. Rapidminer dapat melakukan text mining, Yaitu menganalisis teks, mengekstrak pola serta melakukan kombinasi dalam metode Statistika, Database, dan Kecerdasan buatan. Selain itu rapidminer juga dimanfaatkan untuk memahami Deep learning, Machine learning, dan Analisis prediktif. Rapidminer dapat diaplikasikan dalam Bisnis, Komersial, Pelatihan, Pendidikan, Riset, dan Sebagainya. Proses data mining yang dilakukan oleh rapidminer adalah ETL (Extraction, Transformation, Loading), Data Preprocessing, Visualization, Modelling, dan Evaluation. Proses ini digambarkan dengan XML, dibuat dengan GUI (Graphic user interface), dan menggunakan bahasa pemrograman java.

2.7. K-Fold Cross Validation

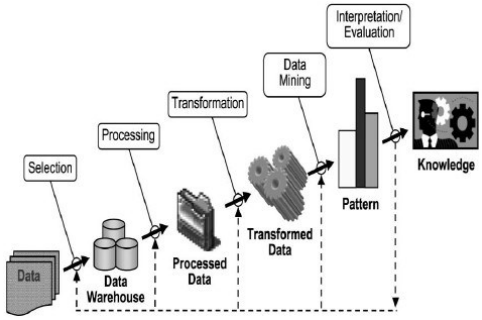
K-fold cross validation digunakan untuk mengestimasi kesalahan prediksi dalam mengevaluasi kinerja model. Data dibagi menjadi himpunan bagian k berjumlah hampir sama. Model dalam klasifikasi dilatih dan diuji sebanyak k. Disetiap pengulangan, salah satu himpunan bagian akan digunakan sebagai data training dan data testing[8]. Langkah-langkah dari k-fold cross validation yaitu :

1. Total data dibagi menjadi k bagian.
2. Fold ke-1 adalah ketika bagian ke-1 menjadi data uji (testing data) dan sisanya menjadi data latih(training data). kemudian, hitung akurasi atau kesamaan atau kedekatan suatu hasil pengukuran dengan angka atau data yang sebenarnya berdasarkan porsi data tersebut.
3. Fold ke-2 adalah ketika bagian ke-2 menjadi data uji (testing data) dan sisanya menjadi data latih (training data). kemudian hitung akurasi berdasarkan porsi data tersebut.
4. Demikian seterusnya hingga mencapai fold ke-k. Hitung rata-rata akurasi dari k buah akurasi diatas. Rata-rata akurasi ini menjadi akurasi final.

Printer : Jurnal Pengembangan Rekayasa Informatika dan Komputer
Vol. 1 No. 2 Nopember 2023
Hal. 88 – 102
e-ISSN 3052-9142
DOI : 10.29408/jprinter.v1i2.22007 **URL : https://doi.org/ 10.29408/jprinter.v1i2.22007**

2.8. Tahapan Data Mining

Data Mining merupakan salah satu dari rangkaian *knowledge discovery in database (KDD)*. Adapun serangkaian proses tahapan dalam Data Mining tersebut sebagai berikut:



Gambar 1. Tahapan data mining

a) Pembersihan Data (*Data Cleaning*)
Merupakan proses untuk membuang data yang tidak konsisten dan menghilangkan *noise.* Data yang diperoleh baik dari database suatu perusahaan maupun hasil eksperimen memiliki isian yang tidak sempurna seperti data yang hilang, data tidak valid atau bahkan sekedar salah ketik. Data-data yang tidak relevan lebih baik dibuang, pembersihan data sangat penting dilakukan karena akan mempengaruhi performasi dari teknik data mining.

b) Integrasi Data (*Data Integration*)
Merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database atau file teks. Integrasi dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan, dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya.

c) Seleksi Data (*Data Selection*)
Data yang ada pada database seringkali tidak semuanya dipakai. Oleh karena itu, hanya data yang sesuai untuk dianalisis yang akan diambil dari database.

d) Transformasi Data (*Data Transformasi*)
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa di aplikasikan.

Printer : Jurnal Pengembangan Rekayasa Informatika dan Komputer
Vol. 1 No. 2 Nopember 2023
Hal. 88 – 102
e-ISSN 3052-9142
DOI : 10.29408/jprinter.v1i2.22007 **URL : https://doi.org/ 10.29408/jprinter.v1i2.22007**

e) Proses Mining

Proses mining merupakan proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

f) Evaluasi Pola (*Pattern Evaluation*)

Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi di evaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternative yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining yang lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil diluar dugaan yang mungkin bermanfaat.

3. Metode Penelitian

Dalam tahap pengumpulan data, Penelitian ini menggunakan metode pengumpulan data dua tipe yaitu pengumpulan data primer dan data sekunder. Data primer adalah data yang dikumpulkan pertama kali dan untuk melihat apa yang sesungguhnya terjadi. Data Sekunder adalah data yang sebelumnya pernah dibuat oleh seseorang baik diterbitkan atau tidak. dalam pengumpulan data primer, peneliti menggunakan metode observasi dan interview, dengan menggunakan data yang berhubungan dengan penggunaan data Kesehatan Masyarakat secara langsung kepada pihak yang terlibat secara langsung. sedangkan dalam pengumpulan data sekunder menggunakan jurnal,buku, publikasi, dan lain-lain.

Sumber data yang dikumpulkan langsung oleh pembuat laporan disebut dengan sumber primer, sedangkan apabila melalui tangan kedua disebut sumber sekunder. data yang diperoleh adalah data sekunder karena diperoleh melalui proses pendataan pada pelaksanaan KKN Tematik Desa Gemilang Universitas Hamzanwadi pada tahun 2019. Data yang dikumpulkan adalah data Kesehatan Kepala keluarga pada tahun 2019 data yang terkumpul sebanyak 790 data.

4. Hasil dan Pembahasan

Dataset yang digunakan yaitu dataset Kessehatan masyarakat kecamatan suralaga berdasarkan penggunaan alat kontrasepsi dengan Jumlah data yang terkumpul sebanyak 790 data dengan 10 atribut yaitu : nama lengkap, umur, pendidikan terakhir, pasangan usia subur, usia hamil pertama,usia pertama melahirkan, siap kb,

Printer : Jurnal Pengembangan Rekayasa Informatika dan Komputer
Vol. 1 No. 2 Nopember 2023
Hal. 88 – 102
e-ISSN 3052-9142
DOI : 10.29408/jprinter.v1i2.22007 **URL : https://doi.org/ 10.29408/jprinter.v1i2.22007**

alat kontrasepsi, mengikuti program kontap, dan jarak melahirkan. Dari 10 atribut tersebut dapat digunakan sebagai komponen penentu keputusan akhir untuk memperkirakan penggunaan alat kontrasepsi dikecamatan suralaga.

Tabel 2. Sampel dataset

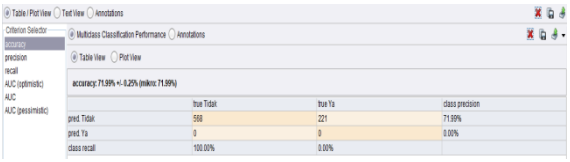
NAMA LENGKAP	UMUR	PENDIDIKAN TERAKHIR	PASANGAN USIA SUBUR	USIA HAMIL PERTAMA	USIA PERTAMA MELAHIRKAN	SIAP KB	ALAT KONTRASEPSI	MENGIKUTI PROGRAM KONTAP
humaidi	40	SD	Ya	19-21 tahun	-	Ya	Suntik KB	Tidak
rafiqah	40	SMA	Ya	15-18 tahun	15-18 tahun	Ya	Suntik KB	Tidak
Mahsan	46	SD	Ya	19-21 tahun	19-21 tahun	Ya	Suntik KB	Tidak
abdul qadir jaela	19	SMA	Ya	15-18 tahun	15-18 tahun	Tidak	TIDAK KB	Tidak
Hardi	32	SD	Ya	22-25 tahun	22-25 tahun	Ya	Suntik KB	Tidak
SALEHUDIN	37	SMA	Ya	19-21 tahun	19-21 tahun	Ya	KB IUD	Tidak
Nasri	35	SMP	Ya	19-21 tahun	19-21 tahun	Ya	Pil KB	Tidak
andiyono	32	SMA	Ya	15-18 tahun	15-18 tahun	Ya	KB IUD	Tidak
toki wantono	35	SMA	Ya	19-21 tahun	19-21 tahun	Ya	Suntik KB	Tidak

Berikut adalah hasil dari pengolahan data menggunakan rapidminer dengan metode klasifikasi dan algoritma random forest. Pengujian dilakukan dengan K-Fold 2-10 cross validation.

Tabel 3. Hasil Pengolahan Data Algoritma Random Forest

NO	Cross Validation	Accuracy (%)	Precision (%)	Recall (%)	AUC
1	k-2	71.99 %	71,99 %	100 %	0.660
2	k-3	71.99 %	71,99 %	100 %	0.661
3	k-4	71.99 %	71,99 %	100 %	0.680
4	k-5	71.99 %	71,99 %	100 %	0.647
5	k-6	71.99 %	71,99 %	100 %	0.673
6	k-7	71.99 %	71,99 %	100 %	0.676
7	k-8	71.99 %	71,99 %	100 %	0.678
8	k-9	71.99 %	71,99 %	100 %	0.640
9	k-10	71.99 %	71,99 %	100 %	0.671

Dari hasil pengujian menggunakan k-fold cross validation hasil akurasi tertinggi berada di k-4 yaitu 71,99% untuk nilai accurasi dan 0,680 untuk nilai AUC.



Gambar 2. Hasil Accuracy K-fold Cross Validation

Dari tabel pada gambar 2, maka kita dapat melihat True tidak mengikuti program kontap sebesar568 , hasil tersebut adalah hasil klasifikasi yang berarti true positif dimana bahwa faktanya masyarakat itu tidak mengikuti program kontap dan hasil prediksi juga benar. Dan untuk nilai true mengikuti program kontap sebesar 221 dan hasil prediksinya adalah salah, sedangkan false negatif sebesar 0 yang artinya antara nilai factual dan hasil klasifikasi adalah salah, bahwa masyarakat itu tidakmengikuti program dan true negatifnya juga sebesar 0. Adapun bentuk confusion matrix.

Tabel 4. Confusion Matrix

Printer : Jurnal Pengembangan Rekayasa Informatika dan Komputer
Vol. 1 No. 2 Nopember 2023
Hal. 88 – 102
e-ISSN 3052-9142
DOI : 10.29408/jprinter.v1i2.22007 **URL : https://doi.org/ 10.29408/jprinter.v1i2.22007**

	Ya	Tidak
Ya	568	221
Tidak	0	0

Adapun perhitungan confusion matrix sebagai berikut :

- Kita bisa menghitung tingkat error atau kesalahan dengan menggunakan symbol E dan persamaan berikut :

$$E = \frac{\text{FP} + \text{FN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}} X100\%$$

$$E = \frac{221 + 0}{221 + 0 + 568 + 0} X100\%$$

$$E = \frac{221}{789} X100\%$$

$$E = 0{,}2801 = 2{,}82\%$$

- Perhitungan untuk mencari accuracy

$$Acc = \frac{\text{TP} + \text{TN}}{FP + FN + TP + TN} X100\%$$

$$Acc = \frac{568 + 0}{221 + 0 + 568 + 0} X100\%$$

$$Acc = \frac{568}{789} X100\%$$

$$Acc = 0.71989 = 71{,}99\%$$

- perhitungan untuk mencari precision

$$Pr = \frac{\text{TP}}{FP + TP} X100\%$$

$$Pr = \frac{568}{221 + 568} X100\%$$

$$Pr = \frac{568}{789} X100\%$$

$$Pr = 0.71989 = 71{,}99\%$$

- Perhitungan untuk mencari Recall

$$Recall = \frac{\text{TP}}{\text{FN} + \text{TP}} X100\%$$

$$Recall = \frac{568}{0 + 568} x100\%$$

$$Recall = \frac{568}{568} x100\%$$

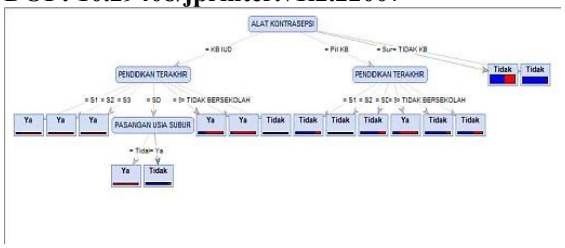
$$Recall = 1\ X\ 100\% = 100\%$$

Dari perhitungan tabel confusion matrix tersebut maka dapat kita lihat hasil untuk nilai accurasi sebesar 71,99%, berarti algoritma tersebut telah berhasil mengklasifikasikan tingkat penggunaan alat kontrasepsi dikecamatan suralaga, untuk selanjtnya nilai Precission sebesar 71,99% ini berarti sistem telah berhasil memberikan informasi yang tepat, sedangkan yang terakhir adalah nilai dari recall sebesar 100% yang artinya sistem telah berhasil memberikan informasi yang benar.

Gambar 3. ROC dan Nilai AUC

Grafik ROC menjelaskan bahwa nilai untuk AUC adalah sebesar 0,679%.

Printer : Jurnal Pengembangan Rekayasa Informatika dan Komputer
Vol. 1 No. 2 Nopember 2023
Hal. 88 – 102
e-ISSN 3052-9142
DOI : 10.29408/jprinter.v1i2.22007 **URL : https://doi.org/ 10.29408/jprinter.v1i2.22007**



Gambar 4. Model Random Forest

5. Kesimpulan

Dari penelitian dan pengolahan data yang sudah dilakukan maka dapat disimpulkan bahwa dalam menentukan tingkat kesehatan kepala keluarga kecamatan suralaga berdasarkan penggunaan alat kontrasepsi dapat diprediksikan menggunakan algoritma random forest. Setelah dilakukan pengujian sebanyak 9 kali pengujian dari k-fold cross validation 2 sampai dengan 10 menggunakan tools *RapidMiner* maka didapatkan hasil akurasi sebesar 71,99% dengan Nilai AUC diperoleh sebanyak 0,680%. Metode ini mampu menganalisa penggunaan alat kontrasepsi, dimana dilakukan dengan cara menampilkan output berupa pohon keputusan yang mampu memberikan analisa bahwa alat kontrasepsi yang digunakan yaitu KB Iud, Pil KB dan Suntik KB.

Daftar Pustaka

[1] Y. Yuliani, *Algoritma Random Forest Untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung Menggunakan Seleksi Fitur Bestfirst*, vol. 5, no. 2. 2022. doi: 10.29408/jit.v5i2.5896.

[2] Y. Yahya and H. Bahtiar, "Pengaruh Pertumbuhan Ekonomi Terhadap Tingkat Kesejahteraan Masyarakat Kecamatan Selong Kabupaten Lombok Timur – Nusa Tenggara Barat Menggunakan Algoritma Naive Bayes," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 20–28, 2021, doi: 10.29408/jit.v4i1.2981.

[3] S. T. Informatika *et al.*, "e-ISSN 2614-8773 92," vol. 3, no. 2, pp. 92–103, 2020.

[4] A. Lalu Kertawijaya, Rina septiana, "Infotek : Jurnal Informatika dan Teknologi," *Sist. Inf. Berbas. Web Penyewaan Wedd. Organ. Pada Doni Organ.*, vol. 4, no. 1, pp. 96–104, 2021.

[5] E. T. Susanti and H. L. Sari, "Pendidikan Kesehatan Tentang Jenis-Jenis Alat Kontrasepsi Terhadap Pemilihan Alat Kontrasepsi," *J. Kesehat.*, vol. 9, no. 1, p. 53, 2020, doi: 10.46815/jkanwvol8.v9i1.95.

[6] D. Irawan, E. B. Perkasa, Y. Yurindra, D. Wahyuningsih, and E. Helmud, "Perbandingan Klassifikasi SMS Berbasis Support Vector

DOI : 10.29408/jprinter.v1i2.22007 **URL : https://doi.org/ 10.29408/jprinter.v1i2.22007**

Machine, Naive Bayes Classifier, Random Forest dan Bagging Classifier," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 10, no. 3, pp. 432–437, 2021, doi: 10.32736/sisfokom.v10i3.1302.

[7] Baskoro, Sriyanto, and L. S. Rini, "Prediksi Penerima Beasiswa dengan Menggunakan Teknik Data Mining di Universitas Muhammadiyah Pringsewu," *Semin. Nas. Has. Penelit. dan Pengabdi. Masy.*, pp. 87–94, 2021.

[8] L. Mardiana, D. Kusnandar, and N. Satyahadewi, "Analisis Diskriminan Dengan K Fold Cross Validation Untuk Klasifikasi Kualitas Air Di Kota Pontianak," *Bul. Ilm. Mat. Stat. dan Ter.*, vol. 11, no. 1, pp. 97–102, 2022.