

## Prediksi Tingkat Kesehatan Masyarakat Berdasarkan Penggunaan Alat Kontrasepsi Menggunakan Algoritma Random Forest

Yahya<sup>1\*</sup>, Nurhidayati<sup>2</sup>, Andri Suherman<sup>3</sup>

<sup>1,2,3</sup>Program Studi Sistem Informasi, Universitas Hamzanwadi

\*ayhay7078@gmail.com

### Abstrak

Kesehatan masyarakat merupakan bagian dari pola hidup yang digunakan untuk mencegah penyakit, memperpanjang masa hidup, dan meningkatkan kualitas sumber daya manusia. Oleh sebab itu, kesehatan masyarakat ditinjau dari berbagai aspek sangat perlu untuk dikaji, agar tujuan dan sasaran yang hendak dicapai dapat terlaksana. Salah satu kajian yang dikemukakan adalah penggunaan alat kontrasepsi bagi masyarakat pedesaan, khususnya masyarakat pedesaan yang berada di Kecamatan Suralaga Kabupaten Lombok Timur – Nusa Tenggara Barat. Penggunaan alat kontrasepsi yang merupakan salah satu cara yang digunakan dalam penerapan Keluarga Berencana masih berada pada tingkat yang tergolong masih rendah. Berdasarkan data yang dimiliki yang berkaitan dengan hal tersebut, diperlukan adanya suatu analisis dan pengolahan data dalam menentukan keputusan yang akan diambil untuk mengetahui pengaruh alat kontrasepsi dalam peningkatan kesehatan masyarakat. Pengolahan dan analisis data dilakukan menggunakan algoritma Random Forest, yang disebabkan oleh karakteristik datasets yang digunakan sesuai. Untuk mendapatkan hasil atau performa yang terbaik dalam mengetahui tingkat akurasi, dari datasets yang digunakan memiliki 10 atribut yaitu : nama lengkap, umur, pendidikan terakhir, pasangan usia subur, usia hamil pertama, usia pertama melahirkan, siap kb, alat kontrasepsi, mengikuti program kontak, dan jarak melahirkan. Akurasi yang diperoleh sebanyak 71,99% dengan output pohon keputusan bahwa alat kontrasepsi yang paling banyak dan efisien untuk digunakan di wilayah Kecamatan Suralaga adalah Pil KB dan Suntik KB.

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

### Abstract

*Public health is part of a lifestyle that is used to prevent disease, extend life span, and improve the quality of human resources. Therefore, it is very necessary to study public health from various aspects, so that the goals and targets to be achieved can be realized. One of the studies presented was the use of contraceptives for rural communities, especially rural communities in Suralaga District, East Lombok Regency - West Nusa Tenggara. The use of contraceptives, which is one of the methods used in implementing family planning, is still at a relatively low level. Based on the data held relating to this matter, analysis and data processing are needed in determining the decisions to be taken to determine the effect of contraceptives in improving public health. Data processing and analysis is carried out using the Random Forest algorithm, which is due to the appropriate characteristics of the datasets used. To get the best results or performance in determining the level of accuracy, the datasets used have 10 attributes, namely: full name, age, last education, couple of childbearing age, age at first pregnancy, age at first birth, ready for birth control, contraceptives, following the contact program, and birth distance. The accuracy obtained was 71.99% with the decision tree output that the most common and efficient contraceptives used in the Suralaga District area were birth control pills and birth control injections.*

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

## 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 peningkatan 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 kontrasepsi. 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 di wajah. 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 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:

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].

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  $0,9345 \times \text{jumlah data (kk)} = 0,9345 \times 1130 \text{ kk} = 1056 \text{ kk}$  yang menunjukkan masyarakatnya sejahtera dan  $6,55\% \times 1130 = 74 \text{ kk}$  yang menyatakan masyarakat yang tidak sejahtera dan dapat dijadikan sebagai acuan dalam pengentasan kemiskinan melalui program yang sudah dicanangkan oleh pemerintah[2].

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 menunjukkan 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 dilakukan. Maka 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 menunjukkan nilai AUC sebesar 0.942 yang didapatkan setelah melakukan pengujian dengan K-Fold Validation 10 sangat baik karena hampir mendekati angka 1[3].

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

ketuntasan belajar Siswa SMA Negeri 3 Selong, selama pandemic pasca covid 19, maka dapat disimpulkan bahwa dalam menentukan 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 menunjukkan 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].

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. Landasan Teori

### 1. 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 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. 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[7].

### 3. Klasifikasi

Untuk membantu mengklasifikasikan tingkat kesejahteraan masyarakat Suralaga di wilayah Lombok Timur, maka data yang dimiliki akan diklasifikasikan lalu diolah dengan algoritma Random Forest. 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[8]

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, 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[9].

#### 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.

Tabel 1. Confusion matrix

Item	Actual Values	
	TP	FP
Predicted values	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.

### 5. K-Fold Cross Validation

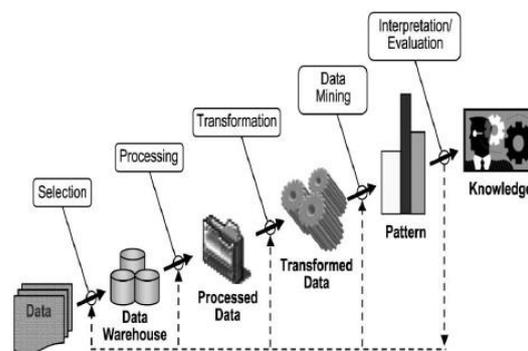
K-fold cross validation digunakan untuk mengestimasi kesalahan prediksi dalam mengevaluasi kinerja model. Data dibagi menjadi himpunan bagian k berjumlah hampir sama. Model dalam klasifikasi dilatih dan diuji sebanyak k. Disetiap pengulangan, salah satu himpunan bagian akan digunakan sebagai data training dan data testing[10]. 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.

### 2.3. Tahapan Penelitian

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

#### 1. 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 performansi dari teknik data mining.

#### 2. 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.

### 3. 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.

### 4. 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.

### 5. Proses Mining

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

### 6. 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

Penelitian ini dilakukan di Kantor Camat Suralaga, Jl. Jurusan Suralaga, Tumbuh Mulia, Suralaga, Kabupaten Lombok Timur Nusa Tenggara Barat. Metode pengumpulan data yang dilakukan dengan cara : 1. Melakukan penginputan data masyarakat Suralaga yang terdiri dari beberapa desa yang menggunakan alat kontrasepsi. 2. Melakukan pemisahan data tentang jumlah masyarakat yang menggunakan alat kontrasepsi berdasarkan jenisnya seperti Pil KB, Suntik KB, Kondom. 3. Merekap seluruh data masyarakat Suralaga yang menggunakan alat kontrasepsi yang dijadikan sebagai datasets. 4. Melakukan pengolahan data yang berjumlah 790 datasets

### 4. Hasil dan Pembahasan

Dataset yang digunakan yaitu dataset Kesehatan 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, 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

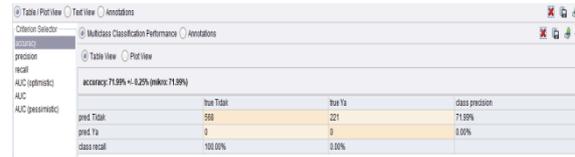
NAMA LENGKAP	UMUR	PENDIDIKAN TERAKHIR	PASANGAN USIA SUBUR	USIA HAMIL PERTAMA	USIA PERTAMA MELAHIRKAN	SIAP KB	ALAT KONTRASEPSI	MENGIKUTI PROGRAM KONTAP
humaidi	40	SD	Ya	19-21 tahun	-	Ya	Suntik KB	Tidak
rafiqah	40	SMA	Ya	15-18 tahun	15-18 tahun	Ya	Suntik KB	Tidak
Mahsan	46	SD	Ya	19-21 tahun	19-21 tahun	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 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 sebesar 568, 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 tidak mengikuti program dan true negatifnya juga sebesar 0. Adapun bentuk confusion matrix.

Tabel 4. Confusion matrix

Item	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 :

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

$$E = \frac{221 + 0}{221 + 0 + 568 + 0} \times 100\%$$

$$E = \frac{221}{789} \times 100\%$$

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

**Perhitungan untuk mencari accuracy**

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

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

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

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

**perhitungan untuk mencari precision**

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

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

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

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

**Perhitungan untuk mencari Recall**

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

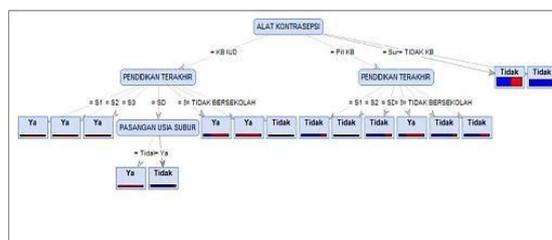
$$Recall = \frac{568}{0 + 568} x100\%$$

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

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

Dari perhitungan tabel confusion matrix tersebut maka dapat kita lihat hasil untuk nilai accuracy 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. Model Random Forest.

**5. Kesimpulan**

Hasil akurasi tertinggi yang diperoleh sebesar 71,99% dengan Nilai AUC diperoleh sebanyak 0,680%. Metode ini masih kurang mampu menganalisa penggunaan alat kontrasepsi secara menyeluruh, tapi hanya mampu menampilkan output berupa pohon keputusan yang memberikan analisa bahwa alat kontrasepsi yang digunakan yaitu KB lud, Pil KB dan Suntik KB yang banyak digunakan. Sedangkan penggunaan kondom sangat kurang digunakan oleh kaum laki-laki di Kecamatan Suralaga.

**6. 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 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 Pengabd. Masy.*, pp. 87–94, 2021.
- [8] Agustiningrum, C. I., Gata, W., Nurfalah, R., & Radiyah, U. (2020). KOMPARASI ALGORITMA NAIVE BAYES , RANDOM FOREST DAN SVM UNTUK MEMREDIKSI NIAT. 20(2).
- [9] 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.
- [10] Imran, B., Subki, A., Yani, A., Alfian, M. R., Engineering, I., Engineering, S., & Mataram, U. T. (2019). DATA MINING USING RANDOM FOREST , NAÏVE BAYES , AND ADABOOST MODELS FOR PREDICTION AND CLASSIFICATION OF BENIGN AND. 37–46
- [11] S. Linawati, S. Nurdiani, K. Handayani, and L. Latifah, “Prediksi Prestasi Akademik Mahasiswa Menggunakan Algoritma Random Forest Dan C4.5,” *J. Khatulistiwa Inform.*, vol. 8, no. 1, pp. 47–52, 2020, doi: 10.31294/jki.v8i1.7827.
- [12] N. Cahyana and A. Aribowo, *Metode Data Mining K-Means Untuk Klasterisasi Data Penanganan Dan Pelayanan Kesehatan Masyarakat*, no. 5. 2018.
- [13] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, “Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5,” *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020.
- [14] M. Mahpuz, A. Muliawan Nur, and L. M. Samsu, “Penerapan Algoritma C4.5 Dalam Mengklasifikasi Status Gizi Balita Pada Posyandu Desa Dames Damai Kabupaten Lombok Timur,” *Infotek J. Inform. dan Teknol.*, vol. 5, no. 1, pp. 72–81, 2022, doi: 10.29408/jit.v5i1.4414.
- [15] Y. Yahya and R. Zuliana, “Prediksi Jumlah Penggunaan BBM Perbulan Menggunakan Algoritma Decition Tree (C4.5) Pada Kantor Dinas Lingkungan Hidup dan Kebersihan Kecamatan Selong Kabupaten Lombok Timur,” *Infotek J. Inform. dan Teknol.*, vol. 1, no. 1, pp. 56–63, 2018, doi: 10.29408/jit.v1i1.895.