

PERBANDINGAN PREDIKSI PENYAKIT HIPERTENSI MENGGUNAKAN METODE RANDOM FOREST DAN NAÏVE BAYES

Kharits Abdul Khalim, Umi Hayati, Agus Bahtiar

Program Studi Teknik Informatika S1, STMIK IKMI Cirebon
Jalan. Perjuangan Nomor.10B, Karyamulya, Kec. Kesambi Kota Cirebon, Jawa Barat
kharits4190086@gmail.com

ABSTRAK

Hipertensi adalah pengertian medis dari orang yang mengalami tekanan darah tinggi, yaitu sebuah keadaan yang mana tekanan darah seseorang melampaui tekanan darah normal, dimana kondisi darah berada pada angka 140 mmHg atau kondisi darah diastolic melebihi 90 mmHg. Penyakit hipertensi bisa dipengaruhi dari pola hidup yang tidak sehat, selain dari pola hidup penyakit ini juga dapat dipengaruhi dari kondisi medis tertentu seperti stroke, penyakit ginjal dan penyakit jantung, bisa juga dipengaruhi oleh jenis kelamin dan usia. Kebanyakan orang tidak mengetahui bahwa dirinya terkena penyakit hipertensi. Penelitian ini dilakukan untuk mengetahui kinerja dari algoritma random forest dan naïve bayes dengan software orange terhadap penyakit hipertensi dengan menggunakan komponen age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, oldpeak, slope, ca, dan thal. Dari hasil penelitian yang telah dilakukan dengan cara menggunakan model algoritma naïve bayes dan random forest dengan software orange menunjukkan algoritma Naïve Bayes mendapatkan hasil yang lebih rendah dari Random Forest.

Kata kunci: Algoritma; Random Forest; Naïve Bayes; Hipertensi

1. PENDAHULUAN

Hipertensi merupakan salah satu penyebab seorang mengalami kematian tertinggi di dunia. Penyakit ini sering disebut dengan *the silent killer* karena penyakit ini merupakan pembunuh tersembunyi. Angka kematian masyarakat Indonesia sendiri yang disebabkan hipertensi sebanyak 427.218 kematian. Hipertensi sering terjadi pada masyarakat pada umur 31 sampai 44 tahun (31,6 %), untuk umur 45 sampai 54 tahun (45,3%), dan umur 55 sampai 64 tahun (55,2%). Dari prevalensi hipertensi sebanyak 34,1% diketahui bahwa sebesar 8,8% terdiagnosa penyakit hipertensi dan 13,3% orang yang terdiagnosis hipertensi tidak minum obat dan 32,3% tidak rutin minum obat. Hal tersebut menunjukkan sebagian besar penduduk yang menderita hipertensi tidak mengetahui bahwa dirinya hipertensi sehingga tidak mendapatkan pengobatan. Kasus hipertensi menurut Data WHO 2015 menunjukkan sekitar 1,13 miliar orang di dunia menderita hipertensi, yang dimana setiap 1 dari 3 orang di dunia terdiagnosa menderita hipertensi. Jumlah orang yang menderita hipertensi di dunia yang semakin meningkat disetiap tahunnya. Diperkirakan pada tahun 2025 akan ada 1,5 miliar orang yang mengalami hipertensi dan setiap tahun terdapat 9,4 juta orang yang meninggal akibat hipertensi. Selain secara mendunia, penyakit hipertensi juga menjadi salah satu penyakit tidak menular dan paling banyak diderita masyarakat Indonesia (57,6%). Hal tersebut dibuktikan dengan jumlah kunjungan penderita hipertensi di Fasilitas Kesehatan Tingkat Pertama yang semakin meningkat setiap tahunnya. Menurut Riskesdas tahun 2018 juga menunjukkan bahwa terjadinya peningkatan pada Prevalensi tekanan darah tinggi pada penduduk yang berusia 18 tahun keatas

menaik dari 25,8% menjadi 34,1% hal tersebut tercantum dalam RPJMN 2015-2019. (Asri et al., 2022)

Dengan komplikasi beberapa penyakit yang muncul secara bersamaan, Dengan banyaknya jumlah pasien penderita hipertensi, tidak menutup kemungkinan dapat mengakibatkan bertambahnya jumlah angka kematian yang dipengaruhi hipertensi. Hal tersebut disebabkan hipertensi dapat menjadi faktor akibat penyakit –penyakit lainnya seperti, gagal ginjal, gagal jantung, diabetes, stroke dan hipoglikemia. Dengan banyaknya jumlah pasien setiap tahun nya yang menyebabkan data pasien semakin bertambah. Saat ini data –data tersebut yang disimpan pada aplikasi SIMRS (Sistem Informasi Manajemen Rumah Sakit) dan belum dimanfaatkan secara maksimal. Kurang perhatiannya terhadap data pasien ini dapat menyebabkan banyaknya kendala atau masalah yang seharusnya dapat di antisipasi jauh sebelum terjadi permasalahan tersebut muncul. Semakin lama data pasien yang disimpan tersebut jika dibiarkan tentunya akan tambah sulit memprediksikan sebuah solusi untuk memecahkan permasalahan apabila masalah tersebut muncul. Selain itu, data tersebut akan menjadi sebuah tumpukan arsip atau file yang disimpan dan dijadikan formalitas kerja manajemen rumah sakit saja. Oleh karena itu peneliti melakukan penelitian dengan memanfaatkan kumpulan data tersebut agar dapat dimanfaatkan secara maksimal. (Azwanti & Elisa, 2019)

2. TINJAUAN PUSTAKA

Berdasarkan penelitian terdahulu yang di lakukan oleh Salsabilla C Nurzanah, Syariful Alam, Teguh I Hermanto. Berdasarkan penelitian yang telah

dilakukan, disimpulkan bahwa peneliti berhasil menganalisis data rekam medis pasien yang memiliki riwayat hipertensi dari rentang waktu bulan Januari 2020 sampai dengan bulan Maret 2022 dengan jumlah kasus yang berjumlah 2.967 data. Dari hasil pengolahan dan pengujian yang telah dilakukan, algoritma *apriori* mampu digunakan untuk menentukan gejala hipertensi yang lebih banyak dirasakan oleh penderita hipertensi yang melihat dari kecenderungan gejala yang dialami penderita hipertensi. (Nurzanah et al., 2022)

Penelitian terdahulu juga dilakukan oleh Mayanda Mega Santoni, Nurul Chamidah, dan Nurhafifah Matondang Penelitian ini, telah diimplementasikan menggunakan tiga algoritma machine learning yaitu *decision tree*, *naïve bayes* dan *artificial neural network* bertujuan memprediksi penyakit hipertensi. Data dengan beberapa tahapan yaitu melakukan imputasi missing value, oversampling yang akan dipraproses normalisasi data. Pembagian data dilakukan dengan cara *k-fold cross validation* menggunakan nilai $k=5$. Eksperimen dilakukan dengan menggunakan alat analisis data yaitu software KNIME. (Mega Santoni et al., 2020)

Berdasarkan penelitian terdahulu terkait hipertensi Menggunakan Metode *naïve bayes* oleh Sarif Surejo, Yuni Pasmalisya Chaeriko, Pinky Septiana Ananda menjelaskan bahwa Penelitian ini bertujuan menerapkan metode *naïve bayes* dalam upaya mendiagnosa penyakit hipertensi. Pada penelitiannya analisis data hipertensi dilakukan dengan melalui metode algoritma *naïve bayes* yang menggunakan 100 data latih. Metode *naïve bayes* akan digunakan untuk memproses diagnosa hipertensi dengan mealui perhitungan probabilitas pada setiap kelas pada data latih, kemudian dihitung probabilitas sebelumnya, dan hitung nilai probabilitas akhir untuk mendapatkan hasil diagnosa hipertensi. Dari data latih dimana telah diketahui sebelumnya, selanjutnya akan melakukan pengujian terhadap data yang berjumlah 20 data pengujian yang telah dipilih untuk mengetahui hasil akurasi dari metode *naïve bayes*. Dari hasil pengujian, akurasi algoritma *naïve bayes* mendapatkan 95,00%, sehingga bisa disimpulkan bahwa metode *naïve bayes* untuk mendiagnosis hipertensi memberikan proses yang cepat dan metodenya yang mudah untuk diterapkan serta tingkat akurasi sudah tidak dapat disangkal. (Surejo et al., n.d.)

Penelitian terdahulu juga dilakukan oleh Purwono Purwono, Pramesti Dewi, Sony Kartika Wibisono, dan Bala Putra Dewa, Dari hasil penelitian, ditemukan bahwa model klasifikasi neural network bisa dipakai sebagai alat klasifikasi untuk prediksi sebuah penyakit hipertensi. Pengolahan data diperlukan dalam penelitian dikarenakan model neural network membutuhkan transformasi dataset dalam bentuk format numerik. Dataset yang telah diproses selanjutnya dibagi terhadap komposisi data latih sebanyak 70% dan data uji sebanyak 30%. Implementasi pada neural network menggunakan

parameter `solver= 'lbfgs'`, `alpha= 1e-5`, `hidden_layer_sizes= (5, 2)`, `random_state=1`. Parameter tersebut akan diterapkan kembali pada tahap klasifikasi jenis penyakit hipertensi. Hasilnya adalah akurasi prediksi jenis hipertensi yang mendapatkan nilai sebanyak 85%. (Purwono et al., 2022)

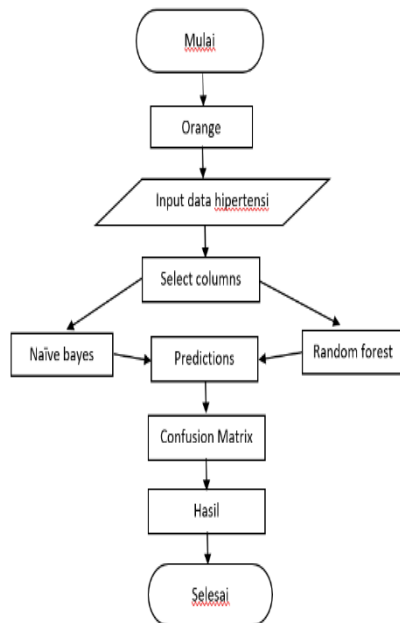
Penelitian terdahulu juga dilakukan oleh Indah Martika Pasaribu, Favorisen Rosyking Lumbanraja, Dewi Asiah Shofiana, dan Aristoteles Penelitian yang memperoleh hasil klasifikasi dengan memakai metode support vector machine (SVM) dengan cara penentuan parameter terbaik menggunakan 10-fold cross validation. Proses pembagian data dilakukan menggunakan aturan 90% data training dan 10% data testing. Terdapat 919 data pada penelitian ini dengan pembagian data training berjumlah 827 data dan data testing berjumlah 92 data. Keseluruhan percobaan hasil klasifikasi menggunakan tiga kernel yang mendapatkan hasil accuracy terbesar pada kernel gaussian yaitu sebesar 99.78%, recall tidak mendapatkan nilai terbesar pada ketiga kernel disebabkan memiliki nilai yang sama yaitu sebesar 99.83%, dan precision terbesar pada kernel gaussian adalah sebesar 99.83%. (Martika Pasaribu et al., 2021)

Penelitian terdahulu juga dilakukan oleh Misna, Rais, Iut Tri Utami, Mendapatkan hasil penelitian yang menunjukkan bahwa terdapat faktor-faktor yang mempengaruhi penderita hipertensi adalah lama seorang merokok (X1), jenis rokok apa yang dihisap (X3) dan cara menghisap rokoknya (X4). Persamaan regresi logistik biner dengan fungsi logit yang mendapatkan hasil yaitu $g(x) = 0,981 + 2,478X1 - 2,364X3 + 1,549X4$. Nilai ketepatan klasifikasi penderita hipertensi dengan menggunakan analisis regresi logistik biner mendapatkan nilai sebesar 77,6%. (Tri Utami, 2018)

Penelitian terdahulu juga dilakukan oleh Insani & Ramadhani, Hasil penelitian ini menunjukkan bahwasannya konsumsi garam, junk food, makanan yang diasinkan, dan makanan yang dibakar merupakan faktor risiko yang dapat berhubungan secara signifikan dengan kejadian hipertensi ($p < 0.05$). Faktor risiko yang signifikan secara multivariat kemudian diubah kedalam bentuk skor. Hasil AUC dari total skor mendapatkan nilai sebesar 0,725 (p -value: 0,000, 95% CI: 0,648-0,803). Sensitifitas dan spesifisitas dari total skor yang didapat sebesar 0,478 dan 0,858 dengan titik potong sebesar 1,5 dengan prevalensi dari hipertensi semakin banyak dengan semakin tingginya skor. Saran bagi tenaga kesehatan untuk senantiasa melaksanakan kegiatan deteksi dini dan promosi kesehatan mengenai pola konsumsi dan hipertensi yang bertujuan meningkatkan kesadaran masyarakat dalam menjaga pola konsumsi sehari-hari. Hal ini dengan tujuan untuk mengurangi angka prevalensi kejadian hipertensi pada Kecamatan Aek Kuasan. (Insani & Ramadhani, 2022) hasil yang dapat ditarik pada penelitian ini adalah mendapatkan hasil hubungan yang signifikan antara tingkat stress dengan kejadian hipertensi pada

anggota Prolanis di Wilayah Kerja Puskesmas Parongpong. Semakin tinggi tingkat stress maka akan memicu peningkatan tekanan darah baik systolic maupun diastolic. Rekomendasi yang dapat diberikan kepada anggota Prolanis untuk memperbaiki pola hidup yang lebih baik dan tetap menjaga pola hidup sehat telah dilakukan dengan cara mengontrol pikiran agar terhindar dari tingkat stress yang dapat mempengaruhi kesehatannya terlebih peningkatan tekanan darah. (Situmorang et al., 2020)

3. METODE PENELITIAN



Gambar 1. Metode Penelitian

3.1. Dataset

Dataset merupakan kumpulan dari objek dan sifat atau karakteristik dari suatu objek itu sendiri (atribut). Dataset terdiri dari sekumpulan file atau dokumen. Pada statistik, dataset biasanya berasal dari hasil pengamatan aktual yang diperoleh dengan cara mengambil sampel populasi. Dataset dapat menghasilkan algoritma yang bisa dipakai dalam pengujian pada perangkat lunak. (Prasojo & Haryatmi, 2021)

3.2. Software Orange

Orange merupakan software *open source* untuk pengolahan sebuah data *analytics* atau data mining. Dibandingkan dengan software data mining yang lainnya, orange lebih unggul dalam hal visualisasi atau yang biasa disebut *visual programming*. Orange sudah menyediakan banyak widget yang telah diletakkan pada canvas atau drawing board yang kemudian dapat dihubungkan dengan widget lainnya. Melalui media canvas ini maka dapat memudahkan pemakai untuk bermain dengan data dan melakukan proses data analisis secara intuitif. Setiap widget yang ada memiliki fungsi masing-masing dan bisa menerima

input dan mengeluarkan output. (Istiqomah & Maesaroh, n.d.)

3.3. Random Forest

Random Forest merupakan algoritma dalam *machine learning* yang digunakan untuk pengklasifikasian dataset dalam jumlah besar. Karena fungsinya bisa dipakai dengan banyak dimensi dan dengan berbagai skala dan performa yang tinggi. *Random Forest* diimplementasikan untuk berbagai sektor seperti perbankan, kesehatan, analisis keuangan, *e-commerce* dan sebagainya. Cara kerja algoritma ini adalah dengan membuat beberapa pohon keputusan dan menyartukannya untuk memperoleh prediksi yang lebih stabil serta akurat. Hutan (*Forest*) yang dibangun oleh algoritma tersebut merupakan kumpulan pohon keputusan dimana biasanya dilatih dengan metode *bagging*, yang diharapkan menghasilkan model *machine learning* yang lebih baik. (Sudrajat et al., 2022)

3.4. Naïve Bayes

Naive Bayes adalah sebuah pengklasifikasian probabilistic sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma yang memakai teorema Bayes dan mengasumsikan semua atribut independen atau "tidak" saling bergantung yang diberikan oleh nilai terhadap variabel kelas. Definisi lain juga mengatakan bahwa *Naive Bayes* adalah pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang yang akan datang berdasarkan hasil pengalaman di masa sebelumnya. *Naive Bayes* didasarkan pada asumsi penyederhanaan yaitu nilai atribut secara kondisional saling bebas jika diberikan nilai *output*. Dengan kata lain, yang diberikan nilai *output*, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. Keuntungan penggunaan *Naive Bayes* yaitu tata cara ini hanya memerlukan jumlah informasi pelatihan (*Training Informasi*) yang kecil untuk memastikan parameter yang dibutuhkan dalam proses pengklasifikasian. (Herdiana et al., n.d.)

3.5. Confusion Matrix

Tabel 1. Confusin Matrix

Kelas	klasifikasi Positive	klasifikasi Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	True Negative (TN)	False Negative (FN)

Tabel 1 merupakan klasifikasi dari *confusion matrix*. *True positive* (TP) merupakan model klasifikasi yang menyatakan benar dengan memberi label jumlah tupel positif. *True Negative* (TN)

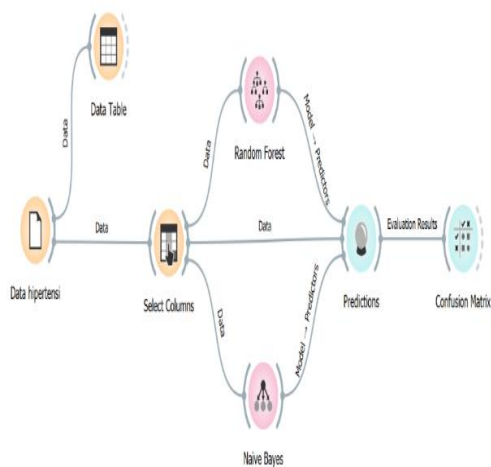
menyatakan bahwa model klasifikasi dinyatakan benar dengan memberi label jumlah tupel negatif. *False positive* (FP) menyatakan bahwa model klasifikasi yang memberi label yang salah untuk jumlah tupel negatif. *False Negative* (FN) menyatakan bahwa model klasifikasi yang memberi label yang salah untuk jumlah tupel positif. (Jenis et al., n.d.)

Berikut adalah Rumus *confusion matrix* untuk melakukan perhitungang *accuracy*, *precision*, dan *recall*.

1. $Accuracy = \frac{TP+TN}{Total}$
2. $Precision = \frac{TP}{TP+FP}$
3. $Recall = \frac{TP}{TP+FN}$

(Normawati & Prayogi, 2021)

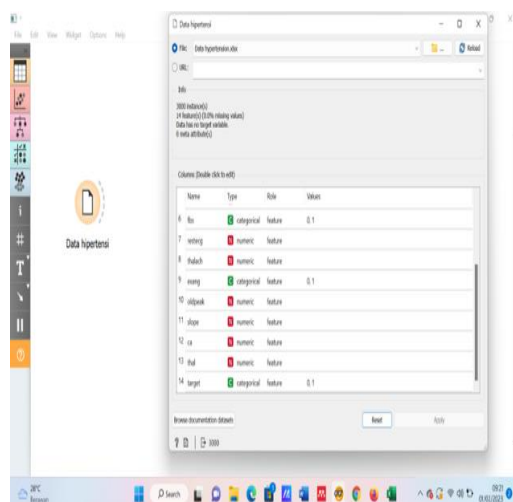
4. HASIL DAN PEMBAHASAN



Gambar 2. Model Penelitian

Hasil penelitian dituliskan berdasarkan tahapan penelitian dari langkah awal penelitian hingga akhir seperti pada gambar 2.

4.1. Widget File



Gambar 3. Wdget File

Pada Gambar 3 melakukan *input* terhadap data hipertensi ke dalam *widget file* yang dimana datanya terdiri dari 3000 *instance* dan 14 *feature*.

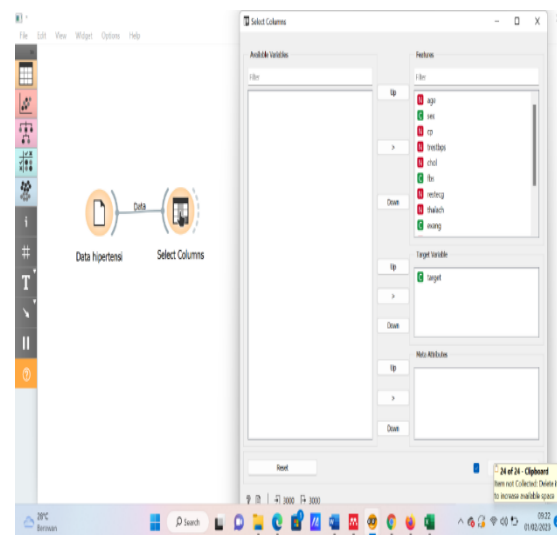
4.2. Widget Data Tabel

	age	sex	tp	breast	chol	fat	redlog	haush	esorg	disgus	snow	ca
2972	59.1	0	140	260.0	0	0	179.0	1.2	1	2		
2973	55.1	0	175	384.0	0	0	152.1	8.0	2	3		
2980	52.9	0	135	282.0	0	0	156.1	8.0	2	0		
2981	59.9	1	160	248.0	0	1	152.1	2.9	1	1		
2982	51.1	0	174	249.0	0	1	143.1	8.0	1	0		
2983	51.1	0	157	274.0	0	1	88.1	1.2	1	1		
2984	51.1	0	132	164.0	0	0	105.1	2.1	1	1		
2985	41.9	0	134	274.0	0	0	164.0	8.5	1	0		
2986	71.9	0	134	486.0	0	0	159.1	1.9	1	2		
2987	26.9	1	160	246.0	0	1	138.1	8.0	1	3		
2988	59.1	1	192	280.0	0	0	195.0	8.0	2	1		
2989	57.9	1	140	254.0	0	0	146.0	2.9	1	3		
2990	46.9	0	140	260.0	0	1	122.1	4.2	1	3		
2991	59.9	0	137	247.1	0	0	143.1	8.1	1	4		
2992	47.9	0	138	284.1	0	1	106.0	1.9	1	3		
2993	59.9	1	160	273.0	0	0	125.0	8.0	2	0		
2994	57.1	0	135	256.0	0	0	138.1	3.0	1	2		
2995	55.1	0	144	280.0	0	0	126.1	8.9	1	0		
2996	53.1	1	120	227.0	0	1	152.1	3.6	1	0		
2997	57.9	0	135	228.1	0	1	155.1	1.9	1	3		
2998	59.9	0	112	220.0	0	1	168.0	8.0	2	1		
2999	41.1	0	131	287.0	0	1	55.1	2.9	1	2		
3000	72.9	0	160	268.0	0	1	169.1	1.6	1	2		

Gambar 4. Bentuk Data

Widget data tabel tempat untuk melihat data yang sudah di input kedalam widget file, Gambar 4 adalah bentuk data hipertensi yang digunakan dalam penelitian.

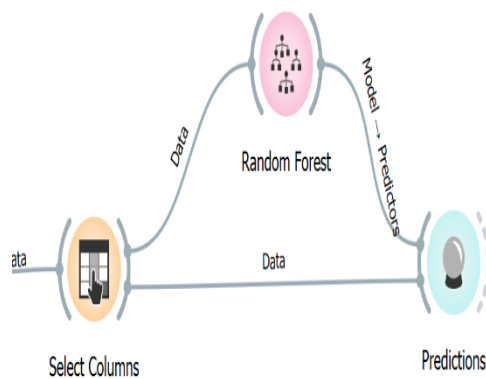
4.3. Widget Selec Columns



Gambar 5 Widget File

Pada gambar 5 setelah data terhubung ke *select columns* akan dilakukan pengambilan satu features yang akan dijadikan sebagai target variable, selanjutnya akan dilakukannya pengujian terhadap data tersebut menggunakan metode algoritma *Random Forest* dan *Naïve Bayes*.

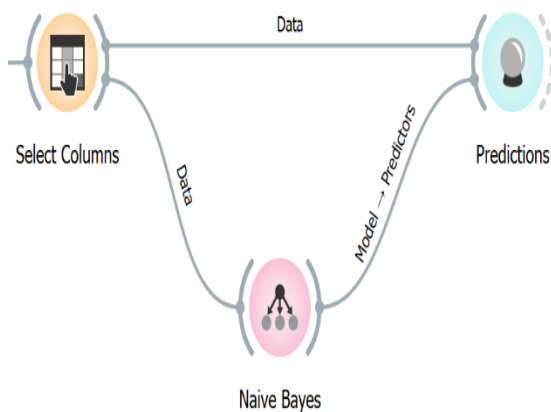
4.4. Widget Random Forest



Gambar 6 Proses Random Forests

Pada gambar 6 dilakukan penghubungan data dari *widget select columns* ke dalam *widget Random Forest*, setelah terhubung selanjutnya akan lanjutkan ke proses *widget prediction*. *Widget prediction* membutuhkan dua data yang sudah dilakukan perhitungan algoritma dan yang belum dilakukan perhitungan algoritma oleh karena itu peneliti menghubungkan data dari *widget select columns* ke dalam *widget prediction* untuk melihat hasilnya.

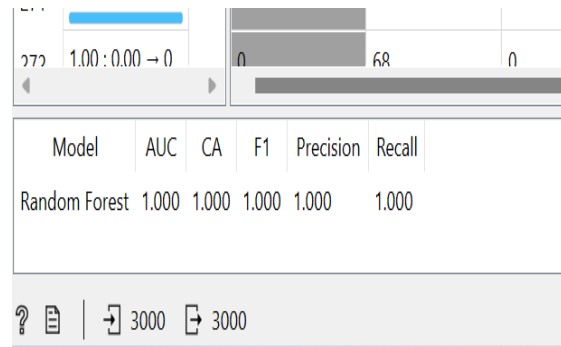
4.5. Widget Naïve Bayes



Gambar 7 Proses Nanive Bayes

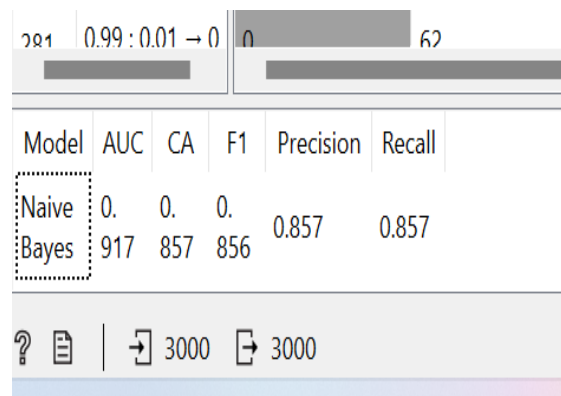
Pada gambar 7 dilakukan penghubungan data dari *widget select columns* ke *widget Naïve Bayes*. Setelah terhubung dengan *Naïve Bayes* selanjutnya akan dihubungkan ke dalam *widget predictions*. Dalam *widget predictions* membutuhkan dua data yang sudah dilakukan perhitungan algoritma dan data yang belum dilakukan perhitungan algoritma oleh karena itu peneliti menghubungkan data dari *select columns* ke *widget predictions* untuk melihat hasilnya.

4.6. Widget Predictions



Gambar 8 Predictions Random Forest

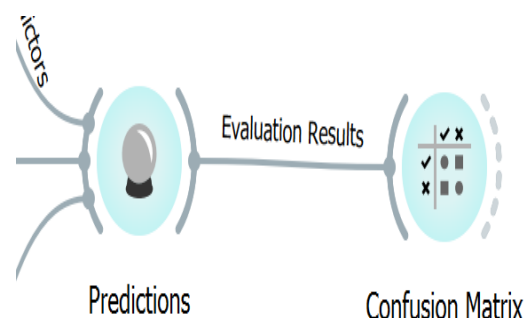
Pada gambar 8 didapatkan hasil prediksi menggunakan algoritma *Random Forest* adalah 1.000 disetiap model *AUC*, *CA*, *F1*, *Precision*, dan *Recall*.



Gambar 9 Predictions Naive Bayes

Pada gambar 9 mendapatkan nilai dari prediksi menggunakan metode algoritma *naïve bayes* yaitu *AUC* (0,917), *CA* (0,857), *F1* (0,856), *Precision* (0,857), dan *Racall* (0,857).

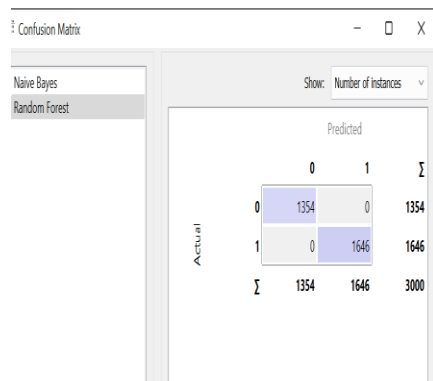
4.7 Widget Confusion Matrix



Gambar 10. Widget Confusion Matrix

Setelah melakukan predictions, pada gambar 10 peneliti melakukan *evaluation result* menggunakan *confusion matrix* dengan cara menghubungkan *widget prediction* yang didalamnya sudah terdapat data hipertensi yang telah dilakukan *predictions*

menggunakan *Random Forest* dan *Naïve Bayes* ke dalam *widget confusion matrix* yang terdapat pada gambar 11 dan 12.



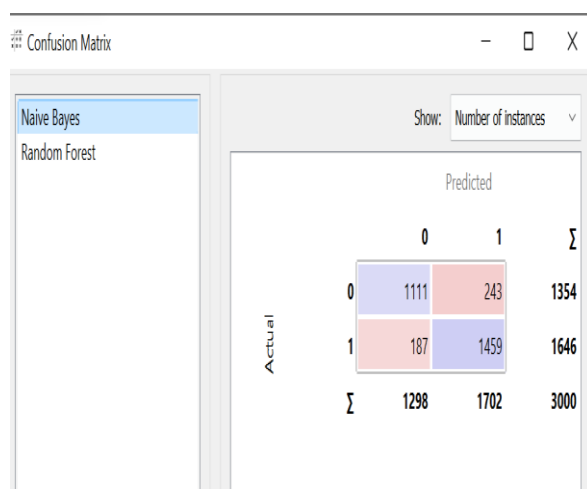
Gambar 11. Hasil dari Confusions Matrix Random Forest

Pada gambar 11 yang di lambankan angka 0 melambangkan No sedangkan angka 1 melambangkan Yes. Dalam baris actual 0 terdapat 1.354 data hipertensi yang dinyatakan benar tidak terkena penyakit hipertensi adalah 1.354 data sedangkan yang dinyatakan terkena penyakit hipertensi adalah 0 data, dan dari baris actual 1 terdapat 1.646 data hipertensi yang tidak terkena penyakit hipertensi adalah 0 data dan yang benar terkena adalah 1.646 data. Maka didapatkan nilai benar positif = 1.354, benar negatif = 1.646, salah positif = 0, dan salah negatif = 0, presisi, dan penarikan metode *Random Forest* adalah sebagai berikut:

$$Accuracy = \frac{1354+1646}{1354+0+1646+0} \times 100\% = 100\%$$

$$Precision = \frac{1354}{1354+0} \times 100\% = 100\%$$

$$Recall = \frac{1354}{1354+0} \times 100\% = 100\%$$



Gambar 12. Hasil Confusion Matrix Naive Bayes

Pada gambar 12 yang di lambankan angka 0 melambangkan No sedangkan angka 1 melambangkan Yes. Dalam baris actual 0 terdapat 1.354 data

hipertensi yang dinyatakan benar tidak terkena penyakit hipertensi adalah 1.111 data dan yang terkena penyakit hipertensi adalah 243 data, dan dalam baris actual 1 terdapat 1.646 data hipertensi yang dinyatakan tidak terkena penyakit hipertensi adalah 187 data dan yang dinyatakan benar terkena adalah 1459 data. Maka didapatkan nilai benar positif = 1.111, benar negatif = 1.459, salah positif = 187, dan salah negatif = 243, presisi, berikut ini adalah penarikan algoritma naïve bayes adalah:

$$Accuracy = \frac{1111+1459}{1111+187+243+1459} \times 100\% = 85,9\%$$

$$Precision = \frac{1111}{1111+187} \times 100\% = 85,5\%$$

$$Recall = \frac{1111}{1111+243} \times 100\% = 82, \%$$

Tabel 2. Hasila Evaluasi

Metode	Accuracy	Precision	Recall
Naïve Bayes	85,90%	85,50%	82,60%
Random Forest	100%	100%	100%

Berdasarkan evaluasi yang dilaksanakan pada widget confusion matrix yang mendapatkan hasil perbandingan *accuracy*, *precision* dan *recall* dari ketiga metode tersebut seperti pada tabel 2.

Klasifikasi menggunakan algoritma *Random Forest*. Dari keseluruhan tahapan yang dilakukan pada algoritma *Random Forest* mendapatkan hasil prediksi untuk *AUC*, *CA*, *F1*, *Precision*, dan *Recall* masing masing mendapatkan nilai 1.000 yang kemudian di validasi menggunakan *confusions matrix* yang dimana mendapatkan hasil benar positif = 1.354, benar negatif = 1.646, salah positif = 0, dan salah negatif = 0. Hasil kemudian dihitung menggunakan rumus *Accuracy*, *Precision* dan *Recall* mendapatkan hasil dari ketiga rumus tersebut mendapatkan nilai 100%. Sedangkan pada algoritma *Naïve Bayes* mendapatkan hasil yaitu auc (0,917), ca (0,857), f1(0,856), presision (0,857), dan recall (0,857) yang didapatkan hasil benar positif = 1.111, benar negatif = 1.459, salah positif = 187, dan salah negative (FN) = 243 kemudian dihitung menggunakan rumus *Accuracy*, *Precision* dan *Recall* mendapatkan hasil untuk *Accuracy* = 85,9%, *Precision* =85,5%, *Recall* =82,6%.

5. KESIMPULAN DAN SARAN

Dari hasil penelitian yang sudah dilakukan memakai data hipertensi yang bertujuan untuk membandingkan hasil prediksi menggunakan dua metode algoritma *Random Forest* dan *Naïve Bayes* menggunakan *software orange*. Pada data yang dipakai didalam penelitian yang dijalankan menunjukkan algoritma *Naïve Bayes* mendapatkan hasil prediksi yang lebih rendah dari *Random Forest*. Penelitian tersebut telah di evaluasi menggunakan *confusion matrix* yang menggunakan tiga rumus yaitu *Accuracy*, *Precision* dan *Recall* yang mendapatkan nilai masing masing, *Naive Bayes* mendapatkan nilai *Accuracy* 85,9%, *Precision* 85,5% dan *Recall* 82,6%. Sedangkan

Random Forest mendapatkan nilai 100% dari ketiga rumus yaitu *Accuracy*, *Precision* dan *Recall*. Dalam penelitian ini membuktikan bahwa *Random Forest* lebih unggul dari *Naïve Bayes* dalam menentukan prediksi hipertensi yang mendapatkan nilai keakuratan 100%.

DAFTAR PUSTAKA

- [1] Asri, I. P., Pitriani Salamah, N., Maryanah Putri, A., Putri Nabila E, S., Khairunnisa, A., Afifah, F., & Kusumastuti, I. (2022). Analisis Faktor Risiko Kejadian Hipertensi di Wilayah Kota Depok. *Journal of Public Health Education*, 1(3), 170–184. <https://doi.org/10.53801/jphe.v1i3.51>
- [2] Azwanti, N., & Elisa, E. (2019). Analisis Pola Penyakit Hipertensi Menggunakan Algoritma C4.5. *InfoTekJar (Jurnal Nasional Informatika Dan Teknologi Jaringan)*, 3(2), 116–123. <https://doi.org/10.30743/infotekjar.v3i2.944>
- [3] Herdiana, R., Hmonangan, R., Hayati, U., & Suprapti, T. (n.d.). *PREDIKSI FINANCIAL DISTRESS PERUSAHAAN FOOD AND BEVERAGE MENGGUNAKAN METODE NAIVE BEYES*.
- [4] Insani, A., & Ramadhani, H. A. (2022). DETERMINAN KEJADIAN HIPERTENSI BERDASARKAN POLA KONSUMSI: MODEL PREDIKSI DENGAN SISTEM SKORING. *Quality : Jurnal Kesehatan*, 16(1), 9–20. <https://doi.org/10.36082/qjk.v16i1.399>
- [5] Istiqomah, S., & Maesaroh, D. S. (n.d.). *Sistem Prediksi Penjualan Hijab Menggunakan Algoritma Prediksi Di Aplikasi Orange (Studi Kasus : Kota Tasikmalaya)*.
- [6] Jenis, K., Berdasarkan, H., Pribadi, F., Dewi, P., Purwono, P., & Kurniawan, S. D. (n.d.). *Pemanfaatan Teknologi Machine Learning pada*.
- [7] Martika Pasaribu, I., Rosyking Lumbanraja, F., Asiah Shofiana, D., Studi Ilmu Komputer, P., Lampung Jalan Soemantri Brojonegoro No, U., Meneng, G., & Lampung, B. (2021). *Klasifikasi Kejadian Hipertensi dengan Metode Support Vector Machine (SVM) Menggunakan Data Puskesmas di Kota Bandar Lampung* (Vol. 2, Issue 2).
- [8] Mega Santoni, M., Chamidah, N., & Matondang, N. (2020). *Prediction of Hypertension using Decision Tree, Naïve Bayes and Artificial Neural Networks in KNIME Analytics Platform* (Vol. 19, Issue 4).
- [9] Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. In *Jurnal Sains Komputer & Informatika (J-SAKTI)* (Vol. 5, Issue 2).
- [10] Nurzanah, S. C., Alam, S., & Hermanto, T. I. (2022). ANALISIS ASSOCIATION RULE UNTUK IDENTIFIKASI POLA GEJALA PENYAKIT HIPERTENSI MENGGUNAKAN ALGORITMA APRIORI (STUDI KASUS: KLINIK RAFINA MEDICAL CENTER). *Jurnal Informatika Dan Komputer) Akreditasi KEMENRISTEKDIKTI*, 5(2). <https://doi.org/10.33387/jiko>
- [11] Prasojo, B., & Haryatmi, E. (2021). Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 7(2), 79–89. <https://doi.org/10.25077/teknosi.v7i2.2021.79-89>
- [12] Purwono, P., Dewi, P., Wibisono, S. K., Putra Dewa, B., Informatika, P., Harapan Bangsa, U., Keperawatan, P., & Bangsa, H. (2022). *Model Prediksi Otomatis Jenis Penyakit Hipertensi dengan Pemanfaatan Algoritma Machine Learning Artificial Neural Network*. 7(2), p.
- [13] Situmorang, F. D., Sri, I., & Wulandari, M. (2020). *HUBUNGAN TINGKAT STRES DENGAN KEJADIAN HIPERTENSI PADA ANGGOTA PROLANIS DI WILAYAH KERJA PUSKESMAS PARONGPONG*. 2(1). <http://ejournal.unklab.ac.id/index.php/kjn>
- [14] Sudrajat, D., Purnamasari, A. I., Dikananda, A. R., Kurnia, D. A., & Bahtiar, A. (2022). Klasifikasi Mutu Pembelajaran Hybrid berdasarkan Algoritma C.45, Random Forest dan Naïve Bayes dengan Optimasi Bootstrap Areggating (Bagging) pada masa COVID-19. *JURIKOM (Jurnal Riset Komputer)*, 9(6), 2227. <https://doi.org/10.30865/jurikom.v9i6.5179>
- [15] Surorejo, S., Chaeriko, Y. P., Ananda, S., Studi, P., Informasi, S., Ymi Tegal, S., & Kunci, K. (n.d.). *PENERAPAN METODE NAÏVE BAYES PADA SISTEM PAKAR UNTUK DIAGNOSA PENYAKIT HIPERTENSI*.
- [16] Tri Utami, I. (2018). Analisis Regresi Logistik Biner Untuk Mengklasifikasi Penderita Hipertensi Berdasarkan Kebiasaan Merokok Di RSUD Mokopido Toli-Toli Binary Logistic Regression Analysis For Classification Of Hypertension Patients Based On Smoking Habits In Mokopido Toli-Toli Hospital. *Natural Science: Journal of Science and Technology ISSN*, 7(3), 341–348.