 ISSN:2302-8149
e-ISSN:2540-9719

Optimasi Fitur dengan *Forward Selection* pada Estimasi Tingkat Obesitas menggunakan *Random Forest*

Feature Optimization with Forward Selection on Obesity Rate Estimation using Random Forest

[1]Agung Bia Alpiansah*,[2]Yudi Ramdhani
[1,2]Teknik Informatika, Teknologi Informasi, Universitas Adhirajasa Reswara Sanjaya, Bandung Antapani,
Jalan Terusan Sekolah No.1-2, Cicaheum, Kota Bandung, Jawa Barat
*e-mail: agungbiaa04@gmail.com

(***received*:** 14 Juli 2023, ***revised*:** 29 Juli 2023, ***accepted*:** 13 Agustus 2023)

Abstrak

Obesitas remaja di Indonesia sedang meningkat, karena kebiasaan makan yang buruk dan gaya hidup yang kurang gerak. Obesitas meningkatkan risiko masalah kesehatan yang serius seperti penyakit jantung, stroke, diabetes, dan lain-lain yang memerlukan tindakan segera. Obesitas berkembang ketika jumlah kalori yang dikonsumsi melebihi jumlah kalori yang dibakar. Obesitas telah menjadi masalah kesehatan masyarakat yang sangat besar di seluruh dunia. Menurut Organisasi Kesehatan Dunia, sekitar 1,9 miliar orang berusia 18 tahun ke atas mengalami kelebihan berat badan, dengan 600 juta orang mengalami obesitas. Menurut Survei Kesehatan dan Morbiditas Nasional, wanita 29,6% lebih mungkin mengalami obesitas dibandingkan pria, dibandingkan dengan 25% pria. Dataset rekam medis gagal jantung akan ditangani dalam dua tahap percobaan berdasarkan validasi. Empat algoritma klasifikasi yang berbeda, termasuk *Random Forest*, K-Nearest Neighbor, Decision Tree, dan Naive Bayes, akan dicoba pada langkah pertama. Untuk *Testing*, metode *Cross Validation* yang menggunakan *Random Forest* mengungguli empat algoritma lainnya dalam *Testing* algoritma. Setelah *Testing*, metode algoritma *Random Forest* menghasilkan nilai akurasi tertinggi, dan dievaluasi kembali menggunakan *Split Validation* dan rasio split yang bervariasi dengan *Forward Selection* sebagai fitu seleksi. Hanya *Testing* yang menggunakan metode *Forward Selection* mengungguli *Testing* yang menggunakan algoritma *Random Forest*.

Kata kunci: Obesitas, *Forward Selection*, *Random Forest*, *Split Validation*, *Cross Validation*.

Abstract

Adolescent obesity is on the rise in Indonesia, owing to bad eating habits and a sedentary lifestyle. Obesity increases the chance of significant health problems such as heart disease, stroke, diabetes, and others that necessitate prompt treatment. Obesity occurs when the amount of calories ingested exceeds the amount of calories burnt. Obesity has become a major public health issue all over the world. The World Health Organization estimates that approximately 1.9 billion people aged 18 and up are overweight, with 600 million obese. According to the National Health and Morbidity Survey, women are 29.6% more likely than males to be obese, while men are 25% more likely to be obese. Based on validation, the heart failure medical record dataset will be handled in two experimental stages. In the first step, four distinct classification algorithms will be tested: Random Forest, K-Nearest Neighbor, Decision Tree, and Naive Bayes. In algorithm Testing, the cross validation technique that employs Random Forest outperforms the other four algorithms. The Random Forest algorithm approach produces the maximum accuracy value after Testing, and it is re-evaluated using split validation with varied split ratios and Forward Selection as a selection feature. Only tests employing the Forward Selection approach outperform those employing the Random Forest algorithm.

Keywords: *Obesity Disease, Forward Selection, Random Forest, Split Validation, Cross Validation*

 ISSN:2302-8149
e-ISSN:2540-9719

1 Pendahuluan

Obesitas dihasilkan oleh kelebihan kalori yang dikonsumsi, dan ini ditentukan oleh penumpukan lemak tubuh yang cepat. Obesitas terjadi ketika jumlah kalori yang dikonsumsi melebihi jumlah kalori yang dibakar [1]. Di seluruh dunia, obesitas kini menjadi wabah masalah kesehatan yang serius. Orang yang kelebihan berat badan berjumlah lebih dari 1,9 miliar orang berusia 18 tahun ke atas, dengan 600 juta di antaranya mengalami obesitas, menurut WHO *(World Health Organization).* Temuan serupa dari Survei Kesehatan dan Morbiditas Nasional dilaporkan, menunjukkan bahwa perempuan lebih mungkin mengalami obesitas dibandingkan laki-laki, pada tingkat 29,6% dibandingkan dengan 25% untuk laki-laki [2]. Obesitas disebabkan oleh berbagai alasan, termasuk faktor biologis, perkembangan, lingkungan, perilaku, dan keturunan. Komorbiditas dewasa seperti diabetes melitus tipe 2, hipertensi, penyakit hati berlemak non-alkohol, apnea tidur obstruktif, dislipidemia, dan sindrom metabolik sedang meningkat, seperti halnya prevalensi obesitas. selama masa kanak-kanak dan remaja. Selain kelainan menstruasi pada remaja dan pubertas dini pada anak-anak, obesitas Selain itu, anak-anak dan remaja yang obesitas dapat mengalami masalah psikologis seperti melankolis, kecemasan, rendah diri, masalah dengan interaksi teman sebaya, dan gangguan makan. Ketimpangan dalam keseimbangan energi, atau terlalu banyak kalori masuk tanpa cukup kalori keluar, adalah penyebab utama obesitas pada masa kanak-kanak dan remaja [3].

Dalam ilmu komputer, teknik *Data Mining* digunakan untuk menganalisis kumpulan data. Istilah "algorithm *Machine Learning*" mengacu pada kelas teknik terkait yang menggunakan pola berbasis model untuk meramalkan dan mengkalsifikasi data baru [4]. Deteksi penyakit sering menggunaka *Data Mining*. *Data Mining* adalah metode untuk menyusun pengetahuan dengan menggunakan algoritma untuk menemukan pola, tren, dan prinsip mekanis tertentu dalam data. Teknik ini digunakan untuk menentukan hubungan antara data yang sebelumnya tidak terlihat [5].

Klasifikasi salah satu tantangan *Data Mining* yang paling sulit. Inputnya adalah kumpulan data catatan *Training*, masing-masing dengan kumpulan atribut yang unik. Karakteristik kategori memiliki domain non-numerik, sedangkan atribut numerik memiliki domain numerik. Selain itu, ada properti unik yang dikenal sebagai label kelas. Tujuan dari klasifikasi ini adalah untuk membuat model yang dapat memperkirakan label kelas yang akan datang untuk record yang tidak berlabel [6]. Untuk mengantisipasi kelas dari suatu objek yang belum ditentukan labelnya, klasifikasi adalah proses pengembangan fungsi atau model yang menggambarkan kelas pada data atau konsep. Karena klasifikasi termasuk dalam kategori pembelajaran yang diawasi, data *Training* diperlukan untuk membuat model Clasifikasi [7].

Fase penting dalam proses *Data Mining* untuk meningkatkan kualitas dan akurasi klasifikasi. Pada penelitian ini dilakukan optimasi fitur pada dataset gambar Pap Smear dengan menggunakan algoritma Genetic Algorithm (GA) untuk memilih fitur yang optimal. GA adalah metode pencarian dan pengoptimalan yang didasarkan pada konsep genetik dan seleksi alam. Beberapa individu terbaik dipilih berdasarkan kualitasnya dalam seleksi fitur menggunakan GA, atau bisa juga dilakukan dengan menggunakan proportional random sampling. Dimensi ruang dataset dapat diturunkan dengan melakukan optimasi fitur, meningkatkan tingkat akurasi klasifikasi [8].

Pengembangan metode Decision Tree yang menggunakan banyak Decision Tree, yang masing-masing telah dilatih menggunakan sampel terpisah adalah *Random Forest*, dan setiap karakteristik dipecah menjadi pohon yang dipilih secara acak. *Random Forest* menawarkan berbagai keuntungan, termasuk kemampuan untuk meningkatkan akurasi hasil ketika ada data yang hilang, menolak outlier, dan menyimpan data secara efisien. Kemampuan untuk mengatasi kesulitan overfitting, kemampuan untuk menggunakannya untuk klasifikasi dan regresi, dan kemampuan untuk mengekstraksi karakteristik yang paling penting dari dataset *Training* adalah semua keuntungan menggunakan algoritma hutan acak sebagai strategi klasifikasi bagian dalam. Penelitian sebelumnya [9]. Metode Decision Tree berfungsi sebagai dasar untuk algoritma *Random Forest* berbasis *ansambel*, yang terkenal dengan kinerjanya. Algoritma *Machine Learning* yang dikenal sebagai algoritma berbasis *ansambel* menggabungkan metode *Machine Learning* yang berbeda ke dalam model prediksi tunggal. Pendekatan ini dirancang untuk mengurangi bias dan kesalahan serta meningkatkan akurasi klasifikasi [10].

Pemilihan fitur yang digunakan sebelum dilakukan klasifikasi. Telah terbukti efektif dalam memecahkan masalah pemilihan fitur yang berlaku untuk data *Forward Selection*, yang menggunakan

 ISSN:2302-8149
e-ISSN:2540-9719

pemilihan karakteristik yang terkait dengan data yang mempengaruhi hasil klasifikasi. Juga, karena kerja algoritma klasifikasi menjadi lebih efisien dan efektif, data dimensi berkurang [11]. *Forward Selection* dilakukan dengan memilih fitur-fitur yang signifikan terhadap temuan klasifikasi. Juga, ketika peningkatan yang efisien dan efektivitas metode klasifikasi beroperasi, dimensi data berkurang. Metode *Forward Selection* dipilih sebagai metode seleksi fitur. Teknik *Forward Selection* merupakan salah satu metode kategori pembungkus (wrap method) dalam pemilihan fitur, yang memiliki tiga kategori filter, wrapper, dan embedding [12].

Pada penelitian ini, penerapan teknik *Data Mining*, khususnya algoritma *Machine Learning* seperti *Random Forest*, akan meningkatkan kemampuan untuk mengidentifikasi dan mengklasifikasikan obesitas pada anak dan remaja. Berdasarkan pemahaman tentang *Data Mining* dan kemampuan algoritma *Random Forest* untuk menangani kompleksitas data, penelitian ini menyiratkan bahwa algoritma *Random Forest* dapat mengungkap pola dan hubungan tersembunyi dalam kumpulan data obesitas, memungkinkan identifikasi dan klasifikasi tingkat obesitas yang lebih akurat pada anak dan remaja. Diharapkan dengan menggunakan teknik ini, akan ada pengetahuan yang lebih besar tentang penyakit obesitas dan kemungkinan mengadopsi intervensi yang lebih efektif untuk mencegah dan mengobati obesitas pada rentang usia ini. Selain itu, dapat dibuktikan bahwa menggabungkan pendekatan *Data Mining* dengan algoritma *Machine Learning* seperti *Random Forest* akan membawa wawasan baru dan pengetahuan yang berguna dalam mengidentifikasi faktor-faktor yang berkontribusi terhadap obesitas pada masa kanak-kanak dan remaja. Teknik *Random Forest* dapat membantu mengungkap variabel dan faktor yang signifikan dalam perkembangan obesitas pada populasi ini dengan mengevaluasi dan mengekstraksi pola dari kumpulan data obesitas yang terkait dengan pola makan, gaya hidup, faktor genetik, dan faktor lingkungan lainnya. Dengan demikian, penelitian ini mengusulkan bahwa penggunaan teknik *Data Mining* dan algoritma Pembelajaran Mesin dapat memberikan pemahaman yang lebih dalam tentang etiologi obesitas pada anak-anak dan remaja, memungkinkan pengembangan strategi pencegahan dan intervensi obesitas yang lebih efektif.

2 Tinjauan Literatur

Mengklasifikasikan Tingkat Probabilitas Obesitas Pada Mahasiswa Sistem Informasi UIN Suska Riau Menggunakan Teknik Klasifikasi *Naive Bayes*. Obesitas adalah masalah yang terkenal di kalangan mahasiswa, yang dapat menyebabkan berbagai penyakit. Pendekatan studi yang digunakan adalah survei berbasis kuesioner, dengan data diolah menggunakan algoritma Rapid Miner dan *Naive Bayes Classification*. Hasil penelitian menunjukkan bahwa 32 siswa (36,36%) dari 88 sampel siswa yang dianalisis mengalami obesitas. Studi ini dapat membantu siswa dan lembaga pendidikan meningkatkan kesadaran tentang pentingnya pemeliharaan kesehatan dan pencegahan obesitas [13].

Kebutuhan akan suatu sistem yang dapat membantu dokter dalam mengidentifikasi obesitas secara lebih tepat dan efektif telah diidentifikasi menjadi perhatian. Pengumpulan data melalui analisis sistem pakar, desain sistem, dan analisis data menggunakan *Backward Chaining* merupakan bagian dari metodologi penelitian. temuan menunjukkan bahwa sistem pakar yang dibuat dapat membantu dokter mendiagnosis obesitas pada orang dewasa dengan akurasi yang lebih tinggi. Namun penelitian ini masih memiliki kekurangan dalam pengumpulan data gejala yang hanya berdasarkan gejala fisik yang dialami oleh pasien, sehingga diperlukan sistem berdasarkan gejala berdasarkan hasil pemeriksaan laboratorium untuk meningkatkan akurasi diagnosis. Selanjutnya, untuk menjaga kebenaran data pada sistem pakar, basis pengetahuan harus dimutakhirkan secara berkala [14].

Kualitas tema songket Palembang ditentukan menggunakan ekstraksi ciri dengan *Scale-Invariant Feature Transform* (SIFT) dan klasifikasi dengan *Random Forest*. Pada tahap awal penelitian, identifikasi masalah diselesaikan dengan mengumpulkan materi berupa jurnal terbaru dan teori klasifikasi topik songket Palembang, metode ekstraksi SIFT, dan metodologi klasifikasi *Random Forest*. Temuan penelitian menunjukkan bahwa teknik SIFT tidak dapat mengekstraksi karakteristik dari suatu item secara memadai karena warna backdrop hampir sama dengan warna objek itu sendiri. Namun hasil penilaian karakteristik klasifikasi motif songket Palembang menunjukkan tingkat pengenalan yang paling tinggi. Dalam *Testing*, model *Random Forest* digunakan. Pengklasifikasian kualitas songket Palembang dapat dinilai dengan menggunakan akurasi, presisi, dan recall [15].

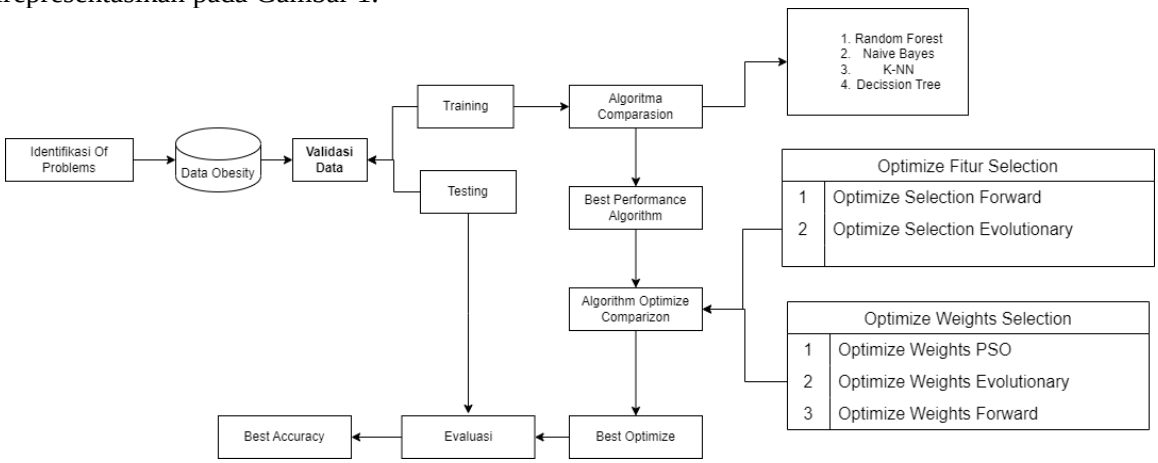
Temuan penelitian terdahulu menekankan perlunya penggunaan metode dan pendekatan untuk memerangi obesitas pada mahasiswa dan meningkatkan kualitas penilaian tema songket Palembang.

 ISSN:2302-8149
e-ISSN:2540-9719

Penerapan pendekatan *Naive Bayes Classification* untuk memprediksi obesitas pada mahasiswa berpotensi meningkatkan kesadaran kesehatan dan menghindari obesitas. Sementara itu, pengembangan sistem pakar menggunakan algoritma *Backward Chaining* dapat membantu klinisi mendeteksi obesitas dengan lebih akurat, meskipun masih perlu meningkatkan pengumpulan data gejala dengan informasi uji laboratorium. Meskipun ditemukan berbagai masalah dengan teknik ekstraksi, penelitian tentang kualitas Tema songket Palembang menggunakan ekstraksi ciri SIFT dan klasifikasi *Random Forest* menghasilkan pengenalan ciri motif yang positif. Untuk meningkatkan hasil penelitian di masa mendatang, tantangan dan keterbatasan tersebut harus diatasi agar sistem dan metode yang digunakan dapat memberikan kontribusi yang lebih efektif untuk mengatasi permasalahan yang ada.

3 Metode Penelitian

Metode *Random Forest (RF)* digunakan dalam penelitian ini untuk mengkategorikan tingkat obesitas pada remaja, sedangkan *Optimize Forward Selection* digunakan untuk memilih karakteristik yang menghasilkan akurasi prediksi terbaik. Penelitian ini dilakukan secara bertahap, seperti yang direpresentasikan pada Gambar 1.



Gambar 1. Desain Penelitian

3.1.1 Identifikasi Masalah

Seperti yang ditunjukkan dalam pendahuluan, untuk pemilihan fitur dalam kategorisasi obesitas, algoritma *Random Forest* akan digunakan bersamaan dengan optimalisasi metode *Forward Selection*. pemilihan fitur untuk akurasi yang lebih baik, tujuan utama penelitian ini adalah untuk meningkatkan akurasi dataset *Obesity* dengan menggunakan algoritma *Random Forest* dan mengoptimalkan algoritma *Forward Selection* sebagai mesin pemilihan fitur untuk mengklasifikasikan seseorang sebagai obesitas.

3.1.2 Dataset

Untuk mengumpulkan data,peneliti memanfaatkan dataset public yang tersedua di uci.repository. Setelah dilakukan pengumpulan, dataset tersebut berisi data estimasi level obesitas seseorang di negara Peru,Mexsico dan Colombia, berdasarkan pola makan, Dataset ini terdiri dari 17 Atribut dan 2.111, Berikut penjelasan atribut pada Tabel 1.

Tabel 1. Penjelesan Atribut

Fitur	**Deskripsi**
Gender	Jenis Kelamin
Age	Umur
Height	Tinggi Badan
Weight	Berat Badan
family_HWO	Riwayat Keluarga Dengan Kelebihan Berat Badan
FAVC	Frekuensi konsumsi makanan berkalori tinggi
FCVC	Frekuensi konsumsi sayuran
NCP	Jumlah makanan utama
CAEC	Konsumsi makanan di antara waktu makan
SMOKE	Merokok
CH2O	Konsumsi Air Harian

 ISSN:2302-8149
e-ISSN:2540-9719

SCC	Atribut yang berhubungan dengan kondisi fisik adalah: Pemantauan konsumsi kalori
FAF	Frekuensi aktivitas fisik
TUE	Waktu menggunakan perangkat teknologi
CALC	Konsumsi alcohol
MTRANS	Transportasi yang digunakan
NObeyesdad	Tingkat Obesitas

3.1.3 Validasi Data

Sebelum memulai analisis, peneliti melakukan *preprocessing* data dengan mencari missing value pada dataset yang dihasilkan. Untuk mengatasi masalah ini, peneliti menggunakan operator filter contoh untuk membuang atau memfilter data yang tidak lengkap atau hilang. Langkah ini sangat penting karena nilai yang hilang dapat memengaruhi kesimpulan analisis dan menimbulkan bias. Dengan menghapus data yang tidak lengkap, peneliti dapat memastikan bahwa kumpulan data yang digunakan dalam penelitian ini berkualitas baik dan dapat diandalkan untuk menghasilkan hasil yang akurat. Dalam hal ini, data penelitian dipisahkan menjadi dua bagian data untuk *Training* dan *Testing*. Peneliti menggunakan prosedur *Cross Validation* dan *Split Validation* untuk membagi data. Untuk menetapkan kinerja terbaik dari model yang akan diuji, digunakan pendekatan *Cross Validation*. Selama proses ini, data *Training* dipisahkan menjadi beberapa subset yang masing-masing digunakan sebagai data *Training* dan data *Testing* secara bergantian. Hasilnya, peneliti dapat menguji model secara menyeluruh dan memperoleh evaluasi kinerja model yang lebih tepat. Sementara itu, *Split Validation* digunakan untuk menguji model tertentu. Dengan menggunakan prosedur *Cross Validation* dan *Split Validation*, data dalam penelitian ini dibagi menjadi data *Training* dan *Testing* melalui langkah validasi data. *Cross Validation* digunakan untuk menguji kinerja terbaik model dengan membagi data berulang kali dan menghitung kinerja rata-rata., sedangkan *Split Validation* digunakan untuk menguji suatu model tertentu dengan membagi data menjadi dua bagian secara acak dengan menggunakan split ratio 0,5 hingga 0,9. Dengan menggunakan kedua metode ini, penelitian ini dapat menghasilkan informasi yang penting mengenai performa dan akurasi model yang digunakan dalam klasifikasi penyakit obesitas, Berikut sebaran jumlah data *Training* dan data *Testing* pada Tabel 2.

Tabel 2. Sebaran Data Training dan Testing

Sebaran Data	Data Training	Data Testing
0,9	1899	212
0,8	1689	422
0,7	1478	633
0,6	1267	844
0,5	1056	1055

3.1.4 Komparasi Algoritma

Penelitian ini menyelidiki empat algoritma yang berbeda pada tahap perbandingan algoritma, yaitu *Random Forest (RF), Decision Tree, K-Nearest Neighbors (K-NN), dan Naive Bayes (NB). Random Forest* adalah model terbaik untuk klasifikasi penyakit Obesitas. Pendekatan *Random Forest* memiliki dua keunggulan utama: mengatasi overfitting dan menghasilkan akurasi klasifikasi yang tinggi. *Random Forest* dapat mengatasi overfitting yang terjadi ketika model terlalu kompleks, memungkinkan untuk memberikan prediksi yang lebih umum dan dapat diterapkan pada data yang tidak diketahui. Selain itu, dengan menggabungkan temuan dari multiple *decision tree*, metode ini mampu menghasilkan akurasi yang tinggi dalam mengidentifikasi data. Selain itu, *Random Forest* efektif dalam mengatasi masalah data yang tidak merata, yang umum terjadi pada klasifikasi obesitas, karena dapat menangani ketidakseimbangan ini melalui strategi pengambilan sampel dan pembobotan yang tepat [16]. Salah satu metode klasifikasi yang digunakan dalam *Data Mining* dan *Machine Learning* adalah *Naive Bayes*. Prosedur ini didasarkan pada teorema Bayes, yang menegaskan bahwa kemungkinan terjadinya suatu peristiwa dapat diestimasi dengan mengalikan probabilitas peristiwa yang terkait dengan peristiwa itu dengan probabilitas peristiwa yang terkait dengan peristiwa itu. *Naive Bayes* digunakan dalam klasifikasi data untuk memprediksi kelas suatu objek berdasarkan atributnya. Algoritma ini bekerja dengan menghitung kemungkinan setiap kelas berdasarkan atribut objek dan memilih kelas dengan probabilitas tertinggi sebagai kelas yang diantisipasi. Terlepas dari

 ISSN:2302-8149
e-ISSN:2540-9719

kesederhanaannya, *Naive Bayes* sangat efektif dalam banyak situasi, terutama saat menangani data dengan banyak atribut dan sampel dalam jumlah besar[17]. *K-Nearest Neighbor (KNN)* adalah algoritma klasifikasi *Machine Learning* yang memprediksi kelas suatu data berdasarkan kelas data yang paling dekat dengan data tersebut. Jarak antara data yang akan diprediksi dengan data yang sudah ada dalam dataset dihitung dengan menggunakan metode ini. Kemudian, dengan menggunakan nilai k yang telah dihitung, KNN akan memilih kelas data yang terdekat dengan data yang akan diramalkan. Nilai k ini menunjukkan berapa banyak tetangga terdekat yang akan diperiksa saat memilih kelas data. Semakin banyak tetangga yang diperiksa, semakin besar nilai k, dan semakin sedikit tetangga yang dipertimbangkan, semakin kecil nilai k. Karena KNN sering digunakan dalam mengkategorikan data dengan beberapa fitur atau dimensi [18]. *Decission Tree* adalah kategorisasi untuk memperkirakan nilai target berdasarkan serangkaian kriteria. Setiap node di *Decision Tree* mewakili atribut, sedangkan cabang mewakili kemungkinan nilai atribut. Metode *Decission Tree* akan memilih atribut yang paling informatif pada setiap tingkat pohon keputusan untuk membagi data menjadi dua atau lebih kelompok yang berbeda. Tujuan *Decission Tree* adalah untuk mengurangi kesalahan klasifikasi sekaligus meningkatkan akurasi prediksi. *Decission Tree* digunakan secara luas karena mudah dipahami dan diinterpretasikan oleh manusia, dan dapat digunakan untuk meramalkan nilai target baik dalam bentuk diskrit maupun kontinu [19].

3.1.5 Komparasi Algoritma Optimasi

Dalam tahap perbandingan algoritma optimasi, penelitian ini menguji lima fitur optimasi selection yang berbeda. Optimalisasi *Forward Selection* adalah proses memilih atribut yang relevan untuk dimasukkan dalam prosedur klasifikasi atau pengelompokan. Ini juga dikenal sebagai pemilihan fitur, pemilihan subset, pemilihan atribut, atau pemilihan variabel. Kompleksitas algoritma klasifikasi, menentukan akurasi algoritma klasifikasi, dan mengidentifikasi atribut yang mempengaruhi tingkat akurasi [20]. Salah satu pendekatan optimasi yang digunakan dalam klasifikasi *Data Mining* adalah *Optimize* Selection *(Evolutionary)*. Strategi ini meningkatkan pemilihan fitur dalam algoritma klasifikasi yang digunakan. *Optimize* Selection *(Evolutionary)* menggunakan teknik algoritma genetika untuk menemukan kombinasi fitur optimal yang dapat meningkatkan nilai akurasi algoritma klasifikasi selama fase optimasi [21]. Selanjutnya *Optimize Weights* PSO *(Particle Swarm Optimization)* adalah salah satu operator yang digunakan untuk mengoptimalkan bobot fitur pada model algoritma PSO (*Particle Swarm Optimization)*. Untuk menghasilkan bobot fitur yang sesuai, operator ini menggunakan pendekatan optimalisasi sekumpulan partikel. Selama prosedur, sekumpulan partikel akan mencari posisi optimal dalam ruang fitur untuk masalah pengoptimalan. Metode optimasi *PSO (Particle Swarm Optimization)* memiliki konsep dasar, mudah diimplementasikan, dan efisien dalam perhitungan jika dibandingkan dengan algoritma matematika dan teknik optimasi heuristik lainnya [22]. Optimasi *Evolutionary* adalah optimasi *Data Mining* yang digunakan untuk meningkatkan kinerja model klasifikasi. meningkatkan pemilihan fitur Serly Agustin dan rekan-rekannya sedang menyelidiki penerapan algoritma *Optimize Weights* (*Evolutionary*) selain pendekatan jaringan saraf untuk klasifikasi stroke otak. Algoritme ini menggunakan optimasi Evolutinary untuk mengoptimalkan bobot fitur yang ditentukan guna meningkatkan akurasi keseluruhan model klasifikasi. Menurut temuan penelitian mereka, algoritma *Optimize Weights (Evolutionary)* [23]. Optimasi bobot menggunakan algoritma Optimalisasi ke depan terbukti berhasil meningkatkan nilai akurasi algoritma pembelajaran mendalam saat digunakan untuk meningkatkan akurasi prediksi pembelajaran mendalam untuk permintaan sepeda bersama. Ketika para peneliti menggunakan teknik bobot *(Forward)* yang optimal, mereka menemukan bahwa nilai akurasinya meningkat [24]. Tujuan dari penelitian ini adalah untuk mengevaluasi dan membandingkan kinerja dari masing-masing fitur optimasi dalam menentukan fitur yang optimal untuk klasifikasi. Dengan demikian, diharapkan dapat ditemukan metode optimasi selection yang paling efektif dan efisien untuk menghasilkan hasil klasifikasi yang optimal.

3.1.6 Optimasi *Forward Selection*

Salah satu strategi pemodelan yang digunakan untuk mengidentifikasi kombinasi variabel yang optimal dari sekumpulan variabel adalah optimasi *Forward Selection*. Variabel secara bertahap dimasukkan ke dalam persamaan dalam fase pemilihan maju dan tidak dapat dihapus. Metode ini dapat digunakan untuk meningkatkan tingkat klasifikasi [25]. Metode *Forward Selection* pada algoritma *Random Forest* dapat meningkatkan performansi secara signifikan dibandingkan dengan *Random Forest* tanpa menggunakan metode tersebut. Dengan demikian, penggunaan algoritma *Forward*

 ISSN:2302-8149
e-ISSN:2540-9719

Selection dapat membantu dalam memilih subset variabel yang paling berpengaruh pada performa model klasifikasi, sehingga dapat meningkatkan akurasi dan performa model secara keseluruhan [26].

3.1.7 Evaluasi

Dalam tahap evaluasi ini nilai Accuracy yang terbaik dalam klasifikasi penyakit obesitas akan diketahui. Perbandingan hasil Accuracy dengan menggunakan split ratio 0,5 hingga 0,9 dari algoritma *Random Forest* dan Algoritma *Random Forest* Berbasis *Optimize* Selection Fitur diamati dan *T-test paired two sample for means* menggunakan Microsoft Excel untuk menguji perbedaan akurasi obesitas sebelum dan sesudah optimasi. Confusion matrix berisi informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi default. Confusion matrix terdiri dari empat bagian yaitu *True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN).* Pada umunya didalam confussion matrix ditampilkan 3 metode *Testing* yaitu:

1. *Accuracy*

   Accuracy adalah cara menguji suatu algoritma berdasarkan tingkat kesamaan antara nilai prediksi dan aktual. Rumus menghitung *accuracy*, yaitu:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} X\ 100 \quad (1)$$

   Rumus di atas digunakan untuk menghitung akurasi dalam konteks klasifikasi. Akurasi membandingkan jumlah prediksi yang benar *(True Positive dan True Negative)* dengan jumlah total data yang diperiksa *(True Positive, True Negative, False Positive, dan False Negative).* Hasil akhir rumus dihitung sebagai persentase, dengan persentase yang lebih besar menandakan tingkat keberhasilan model yang lebih tinggi dalam mengklasifikasikan secara akurat. Akurasi adalah parameter evaluasi kritis untuk menilai kualitas dan kinerja model klasifikasi dalam berbagai aplikasi dan penelitian.

2. *Precision*

   Precision, terkadang dikenal sebagai pengujian presisi, adalah cara menguji algoritma yang membandingkan data yang benar yang dikumpulkan oleh sistem dengan jumlah keseluruhan data yang diambil oleh sistem, baik yang benar maupun yang salah. Rumus menentukan presisi, yaitu:

$$Precision = \frac{TP}{TP+FP} X\ 100 \quad (2)$$

   Rumus di atas digunakan untuk menghitung presisi dalam konteks klasifikasi. Presisi adalah parameter evaluasi yang penting untuk menentukan seberapa akurat model klasifikasi dalam memprediksi kelas positif. Menggunakan jumlah *True Positives (TP)*, yang merupakan kasus positif yang diprediksi dengan benar oleh model, dan jumlah *False Positives (FP)*, yaitu kasus negatif yang salah diprediksi sebagai positif oleh model, presisi dihitung dengan membagi TP dengan prediksi positif total (TP + FP) dan hasilnya dikalikan dengan 100 untuk mendapatkan persentase. Semakin tinggi angka presisi, semakin akurat model dalam mengkategorikan data ke dalam klasifikasi positif dan negatif. dan presisi membatu dalam menentukan seberapa baik model menghasilkan temuan prediksi yang akurat dan bermakna untuk analisis dan pengambilan keputusan.

3. *Recall*

   Recall adalah metode pengujian algoritma yang membandingkan jumlah dukungan data yang tepat yang diambil atau tidak diambil oleh sistem. Rumus *recal*l adalah sebagai berikut:

$$Recall = \frac{TP}{TP+FN} X\ 100 \quad (3)$$

   Rumus di atas digunakan untuk menghitung Recall dalam konteks klasifikasi. Recall adalah parameter evaluasi penting yang mengukur seberapa baik model klasifikasi dapat mengenali (mendeteksi ulang) kasus positif dari semua data positif. Dalam rumus ini, TP adalah *True Positive*, atau jumlah kasus positif yang diprediksi dengan benar oleh model, dan FN adalah False Negative, atau jumlah kasus positif yang secara keliru diprediksi sebagai negatif oleh model. Untuk menghitung perolehan, bagi jumlah *True Positives* dengan jumlah total *True Positives (True Positives + False Negatives)*, dan kalikan hasilnya dengan 100 untuk mendapatkan persentase. Penarikan kembali memberikan ikhtisar tentang sensitivitas keseluruhan model dalam menemukan

 ISSN:2302-8149
e-ISSN:2540-9719

kasus positif, dan semakin besar angka penarikan, semakin baik model dalam mendeteksi kembalinya situasi positif secara akurat, yang sangat penting untuk menghindari pengabaian kasus yang benar-benar positif.

4 Hasil dan Pembahasan

Fase ini menyajikan hasil penelitian uji klasifikasi penyakit obesitas. Langkah pertama adalah mengidentifikasi masalah. Diketahui bahwa teknik atau algoritma pemodelan yang paling sesuai diperlukan untuk mengklasifikasikan penyakit obesitas. Langkah selanjutnya yang dilakukan dataset yang digunakan dalam penelitian ini merupakan dataset publik yang dapat diakses melalui halaman web https://archivebeta.ics.uci.edu/dataset/544/estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition. Setelah dilakukan pengumpulan, dataset tersebut terdiri dari 2.111 record data dengan 17 atribut dimana terdapat 1 label yaitu NObeyesdad. Hal ini ditunjukkan pada Tabel 1.
Setelah pengumpulan data, peneliti mengolah data terlebih dahulu. Untuk menghapus data yang tidak sesuai, data difilter pada level ini dengan menggunakan operator filter contoh dan memeriksa nilai yang hilang dalam data. Missing atau nilai yang hilang pada data yang seharusnya memiliki nilai disebut dengan missing value. Nilai yang hilang dapat disebabkan oleh berbagai sumber, termasuk kesalahan input, kerusakan file, atau ketidaktahuan responden. Setelah prosedur preprocessing data selesai, langkah selanjutnya adalah melakukan perbandingan algoritma untuk membandingkan keempat algoritma yang diuji dalam penelitian ini. *Random Forest (RF), Decision Tree (DT), K-Nearest Neighbor (K-NN), dan Nave Bayes (NB)* adalah algoritma yang digunakan. Untuk mengetahui performansi keempat algoritma tersebut. Metode *Split Validation* digunakan untuk validasi data, menghasilkan angka akurasi, presisi, dan recall. Untuk hasil evaluasi algoritma *Random Forest* dalam klasifikasi penyakit obesitas, untuk mencari hasil akurasi terbaik dari algoritma, algortima *Random Forest* memberikan hasil akurasi yang tinggi yaitu 89.53%. Nilai akurasi yang dihasilkan oleh masing-masing algoritma ditunjukkan pada Tabel 3.

Tabel 3. Komparasi Algoritma

Comparasi Algoritma		
Algoritma	Validation	Accuracy
Random Forest (RF)	*Cross Validation*	89,53%
Decision Tree (DT)	*Cross Validation*	88,54%
K-Nearest Neighbor (K-KNN)	*Cross Validation*	88,11%
Naïve Bayes (NB)	*Cross Validation*	64,76%

Berdasarkan hasil komparasi algoritma tersebut diketahui bahwa grafik akurasi dari ke empat algoritma yang di uji algoritma *Random Forest* (RF) memiliki nilai performa algoritma tertinggi dibandingkan algoritma yang lainnya yaitu sebesar 89,53% untuk nilai accuracy. Selanjutnya, *Confusion Matrix* adalah alat untuk mentabulasi keakuratan *Data Mining*. *Confusion Matrix* yang dibuat dengan teknik klasifikasi Random Forest (RF) dapat dilihat pada Tabel 4.

Tabel 4. Confusion Matrix RF

	TNW	TOL I	T O L II	TOT I	TIW I	TOT II	TOT III	Class Precision
PNW	238	24	12	4	14	1	0	81.23%
POL I	21	216	6	1	0	0	0	88.52%
POL II	6	46	241	15	0	0	0	78.25%
POT I	0	4	30	321	0	2	1	89.66%
PIW	22	0	0	0	258	0	0	92.14%
POT II	0	0	1	10	0	293	0	96.38%
POT III	0	0	0	0	0	1	323	99.69%
Class Recall	82.93%	74.48%	83.10%	91.45%	94.85%	98.65%	99.69%	

 ISSN:2302-8149
e-ISSN:2540-9719

Berdasarkan komparasi algoritma tersebut diketahui bahwa algoritma *Random Forest* (RF) memiliki nilai akurasi paling tinggi dibanding algoritma lainnya yaitu 89,53%. Setelah diketahui performa kinerja terbaik algoritma *Random Forest* (RF) dalam melakukan klasifikasi penyakit obesitas. Selanjutnya data harus dipisahkan menjadi dua bagian, yaitu data *Training* dan data *Testing*. Data *Training* digunakan untuk melatih model RF, sedangkan data *Testing* digunakan untuk menguji model RF. Pemisahan data dapat dilakukan dengan menggunakan *Split Validation*. Split validation dilakukan dengan menggunakan sebaran data 90% *Training* dan 10% *Testing* sampai dengan sebaran data 50% *Training* dan 50% *Testing*, atau dengan kata lain, rasio data yang digunakan 0,9 sampai dengan 0,5. Hal ini dilakukan untuk mengetahui seberapa baik model RF dapat mengklasifikasikan penyakit obesitas yang belum diketahui sebelumnya. Dengan menggunakan split validation, dapat dilihat apakah model RF memiliki kemampuan generalisasi yang baik atau tidak dan melihat model hasil klasifikasi yang paling tinggi akurasi. Berikut ini adalah sebaran data *Testing* dan *Training* menggunakan *Split Validation* dengan rasio 0,5 hingga 0,9. Jika diperhatikan dari Tabel 5, dapat dilihat bahwa nilai akurasi yang paling tinggi pada saat menggunakan split validation algoritma *Random Forest* (RF) dengan rasio 0,8 sebesar 90,02%. Nilai Akurasi yang tinggi menunjukkan bahwa model klasifikasi memiliki tingkat akurasi yang tinggi. Hal ini berarti bahwa data *Training* yang digunakan untuk melatih model sebesar 80% dari data keseluruhan, dan data uji yang digunakan untuk menguji model sebesar 20% dari data keseluruhan, memberikan hasil klasifikasi yang paling baik. Berikut dapat kita lihat pada Tabel 5.

Tabel 5. Split Ratio 0,5 – 09 RF

Algoritma	Validation	Ratio	Accuracy
Random Forest (RF)	*Split Validation*	0,5	87,01%
Random Forest (RF)	*Split Validation*	0,6	89,47%
Random Forest (RF)	*Split Validation*	0,7	87,52%
Random Forest (RF)	*Split Validation*	0,8	90,02%
Random Forest (RF)	*Split Validation*	0,9	88,63%
Average			88,53%

Berdasarkan hasil *Split Validation* dengan menggunakan split ratio 0,5 hingga 0,9 dapat diketahui bahwa algoritma *Random Forest* dengan split ratio 0,5 hingga 0,9 memiliki nilai *average* 88,53% untuk akurasi. Validasi dengan split ratio 0,8 memiliki nilai accuracy tertinggi sebesar 90,02%. Nilai Akurasi yang tinggi menunjukkan bahwa model klasifikasi memiliki tingkat akurasi yang tinggi untuk nilai accuracy. Untuk meningkatkan nilai akurasi pada algoritma *Random Forest* (RF) maka digunakan fitur Optimasi. Pada penelitian ini melakukan komparasi fitur optimasi yaitu Optimize Selection dan *Optimize Weights*. Validasi yang dilakukan pada Tabel 6 adalah tabel perbandingan akurasi algoritma *Random Forest* (RF) berbasis *Optimize Weights Evolunationary*, *Optimize Weights* PSO dan *Optimize Forward* dengan menggunakan cross validasi Berikut dapat kita lihat pada Tabel 6.

Tabel 6. RF + Feature *Optimize Weights*

Optimize Algoritma RF + Weights			
Algortima	Metode	Valdation	Accuracy
Random Forest (RF)	*Optimize Weights* Evolunationary	*Cross Validation*	91,09%
Random Forest (RF)	*Optimize Weights* PSO	*Cross Validation*	92,66%
Random Forest (RF)	*Optimize Weights* Forward	*Cross Validation*	94,13%

Dapat dilihat pada Tabel 6 hasil evaluasi algoritma *Random Forest* (RF) berbasis *Optimize Weights* Forward memberikan hasil akurasi yang tinggi yaitu *Optimize Weights* Forward sebesar 94,13%. Untuk mempermudah dalam memahami perbedaan nilai akurasi antara *Random Forest* yang menggunakan *Optimize Weights* dengan cross validation, dapat membuat grafik. Hasil evaluasi algoritma *Random Forest* (RF) berbasis *Optimize Weights* Forward memberikan hasil akurasi yang tinggi yaitu *Optimize*

 ISSN:2302-8149
e-ISSN:2540-9719

Forward Selection sebesar 94,13%, *Optimize Weights* Evolutionarya sebesar 91,09%, dan *Optimize Weights* PSO sebesar 92,66%. Selanjutnya, Setelah diketahui performa kinerja terbaik *Optimize* Selection Forward algoritma *Random Forest* dalam melakukan klasifikasi Penyakit Obesitas, dilakukan validasi data menggunakan *Split Validation* untuk menguji algoritma tersebut. Berikut ini adalah hasil dari validasi data menggunakan *Split Validation* dengan ratio 0,5 hingga 0,9 yang terdapat pada Tabel 8. Untuk meningkatkan nilai akurasi pada algoritma *Random Forest* (RF) maka digunakan fitur Optimasi. Pada penelitian ini melakukan komparasi fitur optimasi yaitu Optimize Selection dan *Optimize Weights*. Validasi yang dilakukan pada Tabel 7 adalah tabel perbandingan akurasi algoritma *Random Forest* (RF) berbasis Optimize *Forward Selection* dan Optimize Evolutionary Selection dengan menggunakan cross validasi Berikut dapat kita lihat pada Tabel 7.

Tabel 7. RF + *Optimize* Fitur Selection

Optimize Algoritma RF + Fitur Selection			
Algoritma	Metode	Validation	Accuracy
Random Forest (RF)	*Optimize Forward Selection*	*Cross Validation*	94,32%
Random Forest (RF)	*Optimize* Selection Evolunationary	*Cross Validation*	92,94%

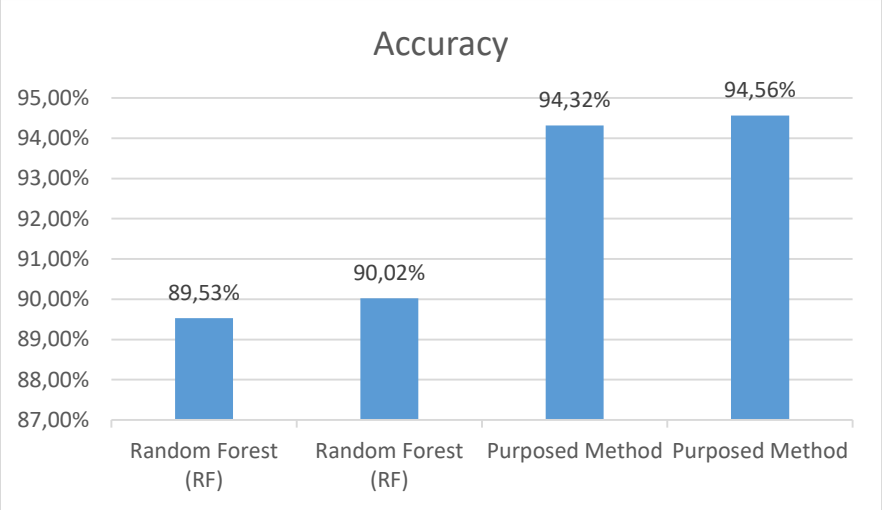
Dapat dilihat pada Tabel 7 hasil evaluasi algoritma *Random Forest* (RF) berbasis *Optimize Forward Selection* memberikan hasil akurasi yang tinggi yaitu *Optimize Forward Selection* sebesar 94,32%. Untuk mempermudah dalam memahami perbedaan nilai akurasi antara *Random Forest* yang menggunakan *Optimize* fitur dengan variasi cross validation, dapat membuat grafik. Hasil evaluasi algoritma *Random Forest* (RF) berbasis *Optimize Forward Selection* memberikan hasil akurasi yang tinggi yaitu *Optimize Forward Selection* sebesar 94,32%, dan *Optimize* Evolutionary Selection akurasinya yaitu 92,94%. Selanjutnya, Untuk meningkatkan nilai akurasi pada algoritma *Random Forest* (RF) maka digunakan fitur Optimasi. Pada penelitian ini *Optimize Weights*. Validasi yang dilakukan pada Tabel 7 adalah tabel perbandingan akurasi algoritma *Random Forest* (RF) berbasis *Optimize Weights* Evolutinary, *Optimize Weights* PSO, dan *Optimize Weights* Forward.
Dalam penelitian ini diusulkan suatu metode untuk klasifikasi Obesitas Optimasi Algorithm Selection Forward sebagai fitur seleksi dan algoritma *Random Forest* sebagai klasifikasi penyakit obesitas. Berikut *proposed method* atau metode yang di usulkan pada Tabel 8.

Tabel 8. Komparasi Algoritma *Random Forest* + Optimize *Forward Selection*

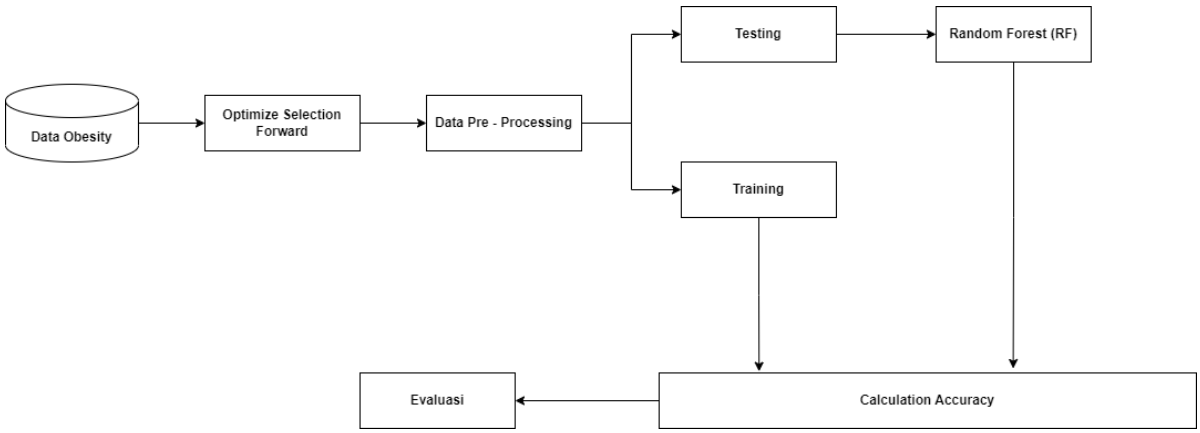
Algoritma	Validation	Accuracy
Random Forest (RF)	*Cross Validation*	89,53%
Random Forest (RF)	*Split Validation 0,8*	90,02%
Purposed Method	*Cross Validation*	94,32%
Purposed Method	*Split Validation 0,6*	94,56%

Berdasarkan hasil Tabel 8 komparasi algoritma *Random Forest* berbasis Optimasi *Forward Selection* tersebut diketahui bahwa hasil sebelum di optimasi dan sesudah memiliki nilai sebesar 89,53% Berdasarkan hasil Split Validation dengan menggunakan split ratio 0,5 hingga 0,9 dapat diketahui bahwa algoritma *Random Forest* dengan split ratio 0,8 memiliki nilai accuracy tertinggi sebesar 90,02%. *Random Forest* (RF) berbasis Optimize *Forward Selection* memberikan hasil akurasi yang tinggi yaitu sebesar 94,32%, Algoritma *Random Forest* berbasis *Optimize Forward Selection* dengan split ratio 0,5 hingga 0,9 memiliki nilai sebesar 94,56% dengan split ratio 0,6 memiliki nilai accuracy tertinggi. Berikut grafik dapat kita lihat pada Gambar 3.

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 2. Komparasi Algoritma Sebelum dan Sesudah



Gambar 3. *Proposed Method*

Pengumpulan dataset Obesitas merupakan tahap pertama dalam penelitian ini. Langkah selanjutnya adalah membagi data menjadi set *Training* dan *Testing*. Data *Training* digunakan untuk menghasilkan model algoritma *Random Forest*, sedangkan *Testing* dataset digunakan untuk menghasilkan hasil akurasi. Langkah selanjutnya dilakukan komparasi algoritma. Komparasi algoritma dilakukan untuk membandingkan beberapa algoritma dalam melakukan pengklasifikasian sehingga didapatkan model algoritma terbaik. Seleksi fitur yang digunakan dalam penelitian menggunakan Optimasi *Forward Selection*. Optimasi Algoritma *Forward Selection* membuat populasi terdiri dari banyak individu yang dipilih dengan paling banyak nilai yang relevan dengan klasifikasi sehingga dapat meningkatkan kinerja nilai akurasi klasifikasi Penyakit Obesitas. Selanjutnya, fitur telah dipilih oleh fitur seleksi algoritma *Forward Selection* diklasifikasikan menggunakan algoritma *Random Forest* (RF.

Tabel 8. Split Ratio 0,5 – 09 RF + *Optimize Forward Selection*

Ratio	Algoritma	
	RF	Accuracy (RF+*Optimize*)
Cross	89,53%	90,02%
0,5	87,01%	94,31%
0,6	89,47%	94,56%
0,7	87,52%	94,15%
0,8	90,02%	94,30%
0,9	88,63%	93,84%
Average	88,53%	94,23%

 ISSN:2302-8149
e-ISSN:2540-9719

Berdasarkan hasil *Split Validation* dengan menggunakan split cross ratio terbaik sebesar 90,02%, Hasil Split ratio 0,5 hingga 0,9 dapat diketahui bahwa algoritma *Random Forest* dengan split ratio 0,5 hingga 0,9 memiliki nilai *average* 94,23% untuk akurasi. Validasi dengan split ratio 0,6 memiliki nilai accuracy tertinggi sebesar 94,56%. Nilai Akurasi yang tinggi menunjukkan bahwa model klasifikasi memiliki tingkat akurasi yang tinggi. untuk nilai accuracy.

T-Test Paired Two Sample akan dilakukan pada nilai akurasi sebelum dan setelah optimasi untuk menyimpulkan penelitian ini. Analisis ini bertujuan untuk menentukan apakah ada perbedaan signifikan secara statistik antara nilai akurasi sebelum dan setelah optimasi. Nilai signifikansi yang digunakan dalam penelitian ini adalah 0.05. *T-Test Paired Two Sample* menunjukkan bahwa nilai dengan two-tail adalah 0,0000801. Dengan nilai signifikansi yang lebih rendah dari alpha (0.05), hasil ini menunjukkan bahwa nilai akurasi sebelum dan setelah optimasi berbeda secara signifikant. Ini menunjukkan bahwa optimasi Selection memiliki efek substansial pada akurasi dan kualitas klasifikasi. Hasilnya memberikan dukungan empiris untuk efektivitas metode optimalisasi *Forward Selection* dalam meningkatkan kinerja model dalam mengklasifikasikan penyakit obesitas. Hasil T-Test Paired Two sample pada Tabel 9.

Tabel 9. T-Test Paired Two Sample

	Accuracy	*Accuracy*
Mean	0,886966667	0,942466667
Variance	0,000145551	5,70267E-06
Observations	6	6
Pearson Correlation	0,283256279	
Hypothesized Mean Difference	0	
Df	5	
t Stat	-11,70337523	
P(T<=t) one-tail	0,00004002508	
t Critical one-tail	2,015048373	
P(T<=t) two-tail	0,0000801	
t Critical two-tail	2,570581836	

Kombinasi metode *Forward Selection* Optimization dan algoritma klasifikasi *Random Forest* untuk mengidentifikasi dan mengkarakterisasi obesitas merupakan inovasi penelitian dan kontribusi terhadap pengetahuan. Menggunakan teknik ini mengatasi kesulitan memilih fitur penting dalam analisis data obesitas sekaligus meningkatkan akurasi klasifikasi penyakit. Penelitian ini juga memajukan pemahaman kita tentang penerapan algoritma pembelajaran mesin dalam perawatan kesehatan, yaitu dalam identifikasi dan pengobatan obesitas anak dan remaja. Hasilnya, penelitian kami memberikan wawasan baru dan secara signifikan meningkatkan pemahaman kami tentang penggunaan teknologi penambangan data untuk mengatasi masalah kesehatan yang kompleks seperti obesitas.

Adopsi optimasi *Forward Selection Optimization* yang dapat meningkatkan kinerja klasifikasi dan akurasi dalam menentukan tingkat obesitas merupakan inovasi penelitian ini. Selanjutnya, menggunakan algoritma klasifikasi *Random Forest* untuk mengklasifikasikan data obesitas membantu meningkatkan pemahaman tentang faktor-faktor yang menyebabkan kondisi ini. Penelitian ini juga menyoroti penerapan algoritma pembelajaran mesin dalam pengaturan kesehatan, yang berpotensi meningkatkan diagnosis, pengobatan, dan pencegahan obesitas. Akibatnya, pembaruan ini dapat memberikan landasan yang kuat untuk pengembangan pendekatan baru untuk mengobati obesitas sekaligus meningkatkan pemahaman kita tentang hubungan antara karakteristik klinis dan obesitas.

5 Kesimpulan

Dalam penelitian ini, peneliti menggunakan dataset obesitas dari UCI *Machine Learning* Repository dan menguji empat algoritma: *Random Forest*, Naive Bayes, K-Nearest Neighbor, dan

Decision Tree. Hasil *Testing* awal menunjukkan bahwa algoritma *Random Forest* memiliki tingkat akurasi tertinggi sebesar 89,53%. Namun, setelah dilakukan *Testing* menggunakan metode *Split Validation* dengan variasi rasio 0,5 hingga 0,9, didapatkan tingkat akurasi tertinggi sebesar 90,02% dengan rasio 0,8. Selanjutnya, peneliti melakukan *Testing* dengan lima optimasi fitur dan bobot, yaitu *Optimize Forward Selection*, *Optimize* Evolutionary Selection, *Optimize Weights* PSO, *Optimize Weights* Evolutionary, dan *Optimize Weights* Forward. Hasilnya menunjukkan bahwa optimasi *Forward Selection* menghasilkan tingkat akurasi tertinggi sebesar 94,32%. Setelah dilakukan uji *Split Validation* dengan rasio 0,6, akurasi tertinggi mencapai 94,56%. Oleh karena itu, dapat disimpulkan bahwa dengan menggabungkan algoritma *Random Forest*, metode *Split Validation*, dan optimasi *Forward Selection*, penelitian ini mencapai tingkat akurasi yang tinggi dalam klasifikasi obesitas.

Referensi

[1] R. T. Aldisa, S. Alfarisi, and M. A. Abdullah, "Penerapan Metode Naïve Bayes Dalam Mendiagnosa Penyakit Leptospirosis," *J. Comput. Syst. Informatics*, vol. 3, no. 4, pp. 521–526, 2022, doi: 10.47065/josyc.v3i4.2205.

[2] Y. Susindra *et al.*, "Pengaruh Media Pembelajaran Infografis Berbasis Aplikasi Android Terhadap Tingkat Pengetahuan Mengenai Obesitas Pada Remaja Putri," vol. 4, no. 2, pp. 81–86, 2023.

[3] A. H. Santoso *et al.*, "Penapisan Hiperuresemia dan Obesitas Pada Remaja di Jakarta Barat," vol. 3, no. 2, 2023.

[4] F. li. D. Cahyanti, F. Sarasati, W. Astuti, and E. Firasari, "Klasifikasi Data Mining dengan Algoritma Machine Larning Untuk Prediksi Penyakit Liver," vol. 14, no. 2, 2023.

[5] L. Sari, A. Romadloni, and R. Listyaningrum, "Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random," vol. 14, no. 01, pp. 155–162, 2023, doi: 10.35970/infotekmesin.v14i1.1751.

[6] A. M. Widodo *et al.*, "Komparasi Performansi Algoritma Pengklasifikasi KNN , Bagging," pp. 367–372, 2021.

[7] B. A. R. P. Wahyu, A. F. Farozi, C. P. Mahendra, and R. K. Hapsari, "Klasifikasi Penderita Penyakit Diabetes Berdasarkan Decision Tree," no. Widyasari 2017, pp. 80–89, 2019.

[8] Y. Ramdhani, "Komparasi Algoritma LDA Dan Naïve Bayes Dengan Optimasi Fitur Untuk Klasifikasi Citra Tunggal Pap Smear," *Informatika*, vol. II, no. 2, pp. 434–441, 2015, [Online]. Available: https://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/view/130%0Ahttps://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/download/130/105

[9] E. N. Candra, I. Cholissodin, and R. C. Wihandika, "Klasifikasi Status Gizi Balita menggunakan Metode Optimasi Random Forest dengan Algoritme Genetika (Studi Kasus : Puskesmas Cakru)," vol. 6, no. 5, pp. 2188–2197, 2022.

[10] Y. Religia, A. Nugroho, and W. Hadikristanto, "JURNAL RESTI Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk," vol. 1, no. 10, pp. 187–192, 2021.

[11] M. R. Fanani, "Algoritma Naïve Bayes Berbasis Forward Selection Untuk Prediksi Bimbingan Konseling Siswa," *J. DISPROTEK*, vol. 11, no. 1, pp. 13–22, 2020, doi: 10.34001/jdpt.v11i1.952.

[12] H.- Harafani, "Forward Selection pada Support Vector Machine untuk Memprediksi Kanker Payudara," *J. Infortech*, vol. 1, no. 2, pp. 131–139, 2020, doi: 10.31294/infortech.v1i2.7398.

[13] W. Muslehatin and M. Ibnu, "Penerapan Naïve Bayes Classification untuk Klasifikasi Tingkat Kemungkinan Obesitas Mahasiswa Sistem Informasi UIN Suska Riau," *Semin. Nas. Teknol. Informasi, Komun. dan Ind.*, pp. 2579–5406, 2017.

[14] A. Nurmasani and Y. Pristyanto, "Algoritme Stacking Untuk Klasifikasi Penyakit Jantung Pada Dataset Imbalanced Class," *Pseudocode*, vol. 8, no. 1, pp. 21–26, 2021, doi: 10.33369/pseudocode.8.1.21-26.

[15] S. Devella, Y. Yohannes, and F. N. Rahmawati, "Implementasi Random Forest Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 2, pp. 310–320, 2020, doi: 10.35957/jatisi.v7i2.289.

[16] P. Rosyani, S. Saprudin, and R. Amalia, "Klasifikasi Citra Menggunakan Metode Random

 ISSN:2302-8149
e-ISSN:2540-9719

Forest dan Sequential Minimal Optimization (SMO)," *J. Sist. dan Teknol. Inf.*, vol. 9, no. 2, p. 132, 2021, doi: 10.26418/justin.v9i2.44120.

[17] N. M. Putry, "Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus," *EVOLUSI J. Sains dan Manaj.*, vol. 10, no. 1, 2022, doi: 10.31294/evolusi.v10i1.12514.

[18] U. Erdiansyah, A. Irmansyah Lubis, and K. Erwansyah, "Komparasi Metode K-Nearest Neighbor dan Random Forest Dalam Prediksi Akurasi Klasifikasi Pengobatan Penyakit Kutil," *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 208, 2022, doi: 10.30865/mib.v6i1.3373.

[19] L. Qadrini, A. Sepperwali, and A. Aina, "Decision Tree dan Adaboostpada Klasifikasi Penerima Program Bantuan Sosial," *Decis. Tree Dan Adab. Pada Klasifikasi Penerima Progr. Bantu. Sos.*, vol. 2, no. 7, pp. 1959–1966, 2021.

[20] M. Rizal, M. Z. Syahaf, S. R. Priyambodo, and Y. Rhamdani, "Optimasi Algoritma Naïve Bayes Menggunakan Forward Selection Untuk Klasifikasi Penyakit Ginjal Kronis," *Naratif J. Nas. Riset, Apl. dan Tek. Inform.*, vol. 5, no. 1, pp. 71–80, 2023, doi: 10.53580/naratif.v5i1.200.

[21] A. Fauzi and Tukiyat, "Analisis Potensi Dana Retail pada Nasabah PT . Bank Tabungan Negara (Persero), Tbk . Dengan Metode Decision Tree dan Naive Bayes Berbasis Optimize Selection (Evolutionary)," *J. Adm. Dan Manjemen*, vol. 9, no. 1, pp. 30–36, 2019.

[22] K. F. Irnanda, A. P. Windarto, and I. S. Damanik, "Optimasi Particle Swarm Optimization Pada Peningkatan Prediksi dengan Metode Backpropagation Menggunakan Software RapidMiner," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 1, p. 122, 2022, doi: 10.30865/jurikom.v9i1.3836.

[23] S. Agustin *et al.*, "Optimasi Feature Selection Menggunakan Algoritma Neural Network Untuk Klasifikasi Brain Stroke," *J. Penelit. Rumpun Ilmu Tek.*, vol. 2, no. 3, pp. 66–74, 2023, [Online]. Available: https://doi.org/10.55606/juprit.v2i3.2009

[24] S. Hidayatulloh *et al.*, "Penggunaan Otimasi Atribut Dalam Peningkatan Akurasi Prediksi Deep Learning Pada Bike Sharing Demand," vol. 9, no. 1, pp. 54–61, 2023.

[25] E. Nurlia and U. Enri, "Penerapan Fitur Seleksi Forward Selection Untuk Menentukan Kematian Akibat Gagal Jantung Menggunakan Algoritma C4.5," *J. Tek. Inform. Musirawas) Elin Nurlia*, vol. 6, no. 1, p. 42, 2021.

[26] T. B. Sasongko and O. Arifin, "Implementasi Metode Forward Selection pada Algoritma Support Vector Machine (SVM) dan Naive Bayes Classifier Kernel Density (Studi Kasus Klasifikasi Jalur Minat SMA)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 4, pp. 383–388, 2019, doi: 10.25126/jtiik.201961000.