

BAB I

PENDAHULUAN

1.1 Latar Belakang

Menurut World Health Organization (WHO), pada tahun 2016, sekitar 39% populasi dewasa global mengalami kelebihan berat badan, dengan 13% diantaranya mengalami obesitas. Prevalensi ini sedikit lebih tinggi pada wanita, dan angka kematian akibat penyakit tidak menular yang terkait obesitas terus meningkat, terutama di negara-negara berkembang (WHO, 2020). Di Indonesia, data Riset Kesehatan Dasar (Riskesdas) menunjukkan tren yang meningkat pada tahun 2007 prevalensi obesitas dewasa tercatat 10,5%, meningkat menjadi 14,8% pada 2013, dan mencapai 21,8% pada 2018. Salah satu penyebab utama tingginya angka obesitas ini adalah pola makan berlebihan yang tidak seimbang serta rendahnya aktivitas fisik. (Gloria Doloksaribu & Rumida, 2021)

Prevalensi obesitas sentral di Indonesia mencapai 31%, dan angka ini terus meningkat setiap tahunnya, khususnya pada remaja. Terdapat kenaikan kasus sebesar 2,9% dari tahun 2007 hingga 2010, kemudian naik lagi sebesar 4,9% dari tahun 2010 hingga 2013, dan bertambah sebesar 4,4% pada periode 2013 hingga 2018. Provinsi DKI Jakarta menempati urutan kedua dengan 42% penduduk usia ≥ 15 tahun yang mengalami obesitas sentral. Secara umum, prevalensi obesitas lebih tinggi pada remaja perempuan dibandingkan remaja laki-laki, dengan proporsi 1,5% pada remaja perempuan dan 1,3% pada remaja laki-laki. Obesitas sentral pada remaja disebabkan oleh banyak faktor yang saling berinteraksi (multifaktorial), termasuk peningkatan konsumsi junk food yang mudah diakses. (Putri et al., 2022)

Pada tahun 2018, prevalensi obesitas di Provinsi Jawa Timur tercatat cukup tinggi, bahkan melebihi rata-rata nasional. Berdasarkan data dari Kementerian Kesehatan (2018), sekitar 22,36% dari populasi di Jawa Timur mengalami obesitas, yang angkanya melampaui prevalensi obesitas di tingkat nasional. Kota Malang, khususnya, menunjukkan peningkatan signifikan dalam prevalensi obesitas antara tahun 2013 hingga 2018, menjadikannya kota dengan angka obesitas tertinggi

ketujuh di Jawa Timur setelah Surabaya. Di kalangan orang dewasa berusia di atas 15 tahun, prevalensi obesitas di Malang mencapai 38%, mendekati angka tertinggi di provinsi ini sebesar 44%. (Najib et al., 2021)

Pada penelitian yang dilakukan pada tahun 2021 sistem pakar dikembangkan untuk mendiagnosis obesitas pada orang dewasa dengan menggunakan metode backward chaining. Metode ini diharapkan dapat membantu tenaga medis dalam proses diagnosis obesitas dengan cara yang lebih efisien dan akurat. Dalam pengembangannya, sistem pakar ini memanfaatkan basis pengetahuan yang dibangun dari pengalaman dan keahlian pakar di bidang kesehatan. Meskipun demikian, hasil penelitian menunjukkan bahwa sistem ini dapat memberikan diagnosis obesitas dengan tingkat akurasi yang memadai, tetapi belum mencantumkan angka akurasi yang dijelaskan secara rinci atau divalidasi dengan angka yang lebih konkret, sehingga perlu evaluasi lebih lanjut. (Rahmah et al., 2021)

Pada penelitian lain yang bertujuan untuk mengklasifikasikan tingkat obesitas pada pria dengan menggunakan algoritma Naïve Bayes berdasarkan data antropometri. Sebanyak 15 variabel dan 252 data digunakan, yang kemudian dibagi untuk pelatihan dan pengujian model. Hasil penelitian menunjukkan bahwa model Naïve Bayes berhasil mengklasifikasikan obesitas dengan akurasi 88%, serta memberikan performa yang baik dalam hal precision, recall, dan F1-score. Dengan akurasi yang bisa dibilang relatif baik penelitian ini seharusnya bisa mengembangkan akurasi tersebut dengan algoritma lain seperti SVM (Support Vector Machine) dan random forest untuk akurasi diatas 90%. (Wie & Siddik, 2022)

Penelitian ini akan membandingkan kinerja dua algoritma machine learning, yaitu Random Forest (RF) dan Gradient Boosting Machine (GBM), dalam memprediksi tingkat obesitas menggunakan dataset yang diambil dari platform Kaggle. Untuk meningkatkan akurasi model, kedua algoritma ini akan diterapkan dengan menggunakan metode Feature Selection SelectKBest, yang berfungsi untuk memilih fitur-fitur yang paling relevan dan signifikan dalam dataset. SelectKBest

akan menggunakan ANOVA F-test (f_{classif}), yaitu metode statistik yang mengevaluasi perbedaan rata-rata antar kelompok untuk menentukan fitur dengan hubungan paling kuat terhadap target variabel.

Pada tahap pertama, algoritma Random Forest akan diterapkan dengan memilih fitur terbaik menggunakan SelectKBest, kemudian diuji untuk menilai akurasi dalam memprediksi tingkat obesitas yang dibagi menjadi enam kategori: Underweight, Normal, Overweight, Obesity I, Obesity II, dan Obesity III. Pengujian ini akan mengukur beberapa metrik kinerja, seperti akurasi, precision, recall, dan F1-score, untuk menilai seberapa baik algoritma Random Forest dalam memberikan prediksi yang tepat. Begitu pula dengan algoritma GBM yang juga akan diterapkan menggunakan SelectKBest untuk memilih fitur-fitur yang relevan. Akurasi model GBM akan diuji dan dibandingkan dengan Random Forest menggunakan metrik yang sama, yakni akurasi, precision, recall, dan F1-score. Melalui pengujian ini, kedua algoritma akan dievaluasi untuk menentukan mana yang memberikan hasil terbaik dalam memprediksi tingkat obesitas.

Setelah melakukan pengujian dan membandingkan akurasi keduanya, algoritma dengan akurasi tertinggi akan dipilih untuk dikembangkan lebih lanjut menjadi aplikasi prediksi tingkat obesitas. Aplikasi ini, yang dibangun dengan menggunakan Python dan framework Flask, memungkinkan pengguna untuk memasukkan data terkait pola makan, aktivitas fisik, dan faktor kesehatan lainnya, yang kemudian akan diproses oleh algoritma terpilih untuk memberikan prediksi tentang tingkat obesitas pengguna. Aplikasi ini diharapkan dapat membantu meningkatkan kesadaran kesehatan dan mencegah risiko penyakit yang terkait dengan obesitas.

1.2 Identifikasi masalah

Berikut adalah identifikasi masalah dari penelitian ini yaitu:

1. Angka prevalensi obesitas, termasuk obesitas sentral, terus mengalami peningkatan di Indonesia, khususnya di kalangan remaja dan dewasa di beberapa daerah seperti DKI Jakarta dan Jawa Timur. Peningkatan ini berhubungan dengan tingginya angka kematian akibat penyakit tidak menular

terkait obesitas, seperti diabetes dan hipertensi. Faktor-faktor yang menyebabkan obesitas, seperti pola makan yang buruk dan kurangnya aktivitas fisik, memerlukan pendekatan yang lebih efektif dalam deteksi dan pengelolaannya.

2. Meskipun berbagai metode dan sistem pakar, seperti backward chaining dan algoritma Naïve Bayes, telah digunakan untuk mendiagnosis obesitas, akurasi dan konsistensinya masih perlu diperbaiki, sehingga penelitian ini bertujuan mengembangkan aplikasi prediksi obesitas berbasis machine learning menggunakan algoritma Gradient Boosting Machine (GBM) untuk menghasilkan prediksi yang lebih akurat dan meningkatkan kesadaran kesehatan masyarakat.

1.3 Rumusan Masalah

Berikut adalah kajian rumusan masalah dari penelitian ini yaitu:

1. Bagaimana proses penerapan algoritma Random Forest (RF) dan Gradient Boosting Machine (GBM) untuk memprediksi tingkat obesitas berdasarkan data kesehatan individu?
2. Bagaimana penerapan metode Feature Selection SelectKBest dengan ANOVA F-test (f_{classif}) mempengaruhi kinerja kedua algoritma (Random Forest dan GBM) dalam meningkatkan akurasi prediksi tingkat obesitas?
3. Di antara algoritma Random Forest (RF) dan Gradient Boosting Machine (GBM), manakah yang memberikan akurasi lebih tinggi dalam memprediksi tingkat obesitas setelah penerapan Feature Selection SelectKBest dengan ANOVA F-test (f_{classif})?
4. Variabel kesehatan apa saja yang berpengaruh signifikan dalam prediksi tingkat obesitas menggunakan algoritma Random Forest (RF) dan Gradient Boosting Machine (GBM) setelah menggunakan SelectKBest dengan ANOVA F-test (f_{classif})?

5. Bagaimana aplikasi prediksi tingkat obesitas, yang dibangun menggunakan algoritma Random Forest (RF) atau Gradient Boosting Machine (GBM), dapat meningkatkan kesadaran masyarakat tentang pentingnya menjaga kesehatan dan mengurangi risiko obesitas?

1.4 Tujuan Penelitian

Berikut adalah tujuan dari penelitian yang akan dilakukan yaitu:

1. Mengembangkan dan menerapkan algoritma Random Forest (RF) dan Gradient Boosting Machine (GBM) untuk memprediksi tingkat obesitas berdasarkan data kesehatan individu serta menganalisis kinerja kedua algoritma dalam memprediksi obesitas.
2. Menganalisis dampak penggunaan metode Feature Selection SelectKBest dengan ANOVA F-test (f_{classif}) terhadap kinerja algoritma Random Forest (RF) dan Gradient Boosting Machine (GBM) dalam meningkatkan akurasi prediksi tingkat obesitas.
3. Membandingkan akurasi kedua algoritma, yaitu Random Forest (RF) dan Gradient Boosting Machine (GBM), setelah penerapan Feature Selection SelectKBest dengan ANOVA F-test (f_{classif}) untuk menentukan algoritma mana yang lebih akurat dalam memprediksi tingkat obesitas.
4. Mengidentifikasi variabel-variabel kesehatan yang memiliki pengaruh signifikan dalam prediksi tingkat obesitas menggunakan algoritma Random Forest (RF) dan Gradient Boosting Machine (GBM) setelah penerapan SelectKBest dengan ANOVA F-test (f_{classif}).

Mengembangkan aplikasi prediksi tingkat obesitas menggunakan algoritma Random Forest (RF) atau Gradient Boosting Machine (GBM) yang dapat meningkatkan kesadaran masyarakat tentang pentingnya menjaga kesehatan dan mencegah obesitas.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai Berikut:

1. Memberikan kontribusi dalam meningkatkan akurasi prediksi tingkat obesitas dengan membandingkan kinerja algoritma Random Forest (RF) dan Gradient Boosting Machine (GBM), serta penerapan metode Feature Selection SelectKBest dengan ANOVA F-test untuk memilih fitur-fitur yang relevan dalam dataset.
2. Menjadi referensi untuk penelitian selanjutnya yang mengkaji penerapan algoritma machine learning dalam prediksi masalah kesehatan, khususnya obesitas, serta aplikasi prediksi berbasis data untuk meningkatkan pemahaman kesehatan masyarakat.
3. Menghasilkan aplikasi prediksi tingkat obesitas yang dapat digunakan sebagai prototipe untuk pengembangan aplikasi kesehatan berbasis machine learning lainnya, yang dapat dikembangkan lebih lanjut untuk prediksi masalah kesehatan lainnya menggunakan algoritma serupa.
4. Meningkatkan kesadaran masyarakat tentang pentingnya pola hidup sehat dan pencegahan obesitas serta penyakit yang terkait dengan obesitas, dengan memberikan prediksi tingkat obesitas secara praktis melalui aplikasi yang dapat digunakan oleh pengguna.

1.6 Kontribusi

Penelitian ini memberikan dampak signifikan dalam bidang aplikasi kesehatan berbasis machine learning, khususnya dalam prediksi obesitas. Beberapa kontribusinya antara lain:

1. Penerapan algoritma Random Forest (RF) dan Gradient Boosting Machine (GBM) dalam memprediksi tingkat obesitas, yang memberikan pendekatan baru untuk meningkatkan akurasi prediksi obesitas, serta membuka potensi penerapan pada masalah kesehatan lain di masa depan.
2. Pengembangan aplikasi prediksi obesitas berbasis machine learning yang memungkinkan masyarakat untuk lebih memahami kondisi tubuh mereka dan

memberikan wawasan mengenai pentingnya pola hidup sehat dalam pencegahan obesitas dan penyakit terkait.

3. Pemanfaatan data yang dihasilkan dari aplikasi untuk program-program kesehatan masyarakat yang lebih terarah dan berbasis data, yang fokus pada pencegahan obesitas dengan pendekatan berbasis teknologi.
4. Pengembangan prototipe aplikasi berbasis Flask yang dapat diperluas ke platform lain, memberikan akses yang lebih luas kepada masyarakat untuk memanfaatkan prediksi kesehatan yang berbasis algoritma machine learning.

1.7 Batasan Masalah

Dari identifikasi masalah yang diterapkan pada penelitian ini, maka pembatasan masalah perlu dilakukan berikut adalah Batasan masalah:

1. Dataset yang digunakan berasal dari platform Kaggle, yang mungkin memiliki keterbatasan dalam variasi dan representasi data. Oleh karena itu, akurasi prediksi mungkin lebih relevan untuk kelompok data tertentu dan tidak bisa digeneralisasi untuk seluruh populasi.
2. Variabel kesehatan yang digunakan dalam model prediksi hanya terbatas pada data yang tersedia dalam dataset. Faktor-faktor risiko obesitas lain yang relevan namun tidak tercakup dalam dataset tidak akan dipertimbangkan dalam aplikasi.
3. Prediksi tingkat obesitas hanya mencakup enam kategori, yaitu Underweight, Normal, Overweight, Obesity I, Obesity II, dan Obesity III. Aplikasi ini tidak akan memprediksi kategori obesitas di luar enam tersebut.
4. Algoritma yang digunakan dalam penelitian ini terbatas pada Random Forest (RF) dan Gradient Boosting Machine (GBM). Penelitian ini tidak membandingkan algoritma lain, sehingga hasil yang diperoleh hanya berlaku untuk kedua algoritma tersebut.
5. Metode Feature Selection SelectKBest dengan ANOVA F-test diterapkan untuk memilih fitur yang paling relevan dalam dataset, dengan tujuan