

BAB I PENDAHULUAN

1.1 Latar Belakang

Machine learning merupakan salah satu cabang kecerdasan buatan yang mampu mengolah data dalam jumlah besar untuk menemukan pola, melakukan klasifikasi, maupun prediksi. Dalam bidang kesehatan, machine learning banyak dimanfaatkan untuk menganalisis data yang kompleks, sehingga dapat mendukung pengambilan keputusan berbasis data. Pemanfaatan teknik ini, seperti data mining, memungkinkan identifikasi pola-pola tersembunyi yang berguna untuk deteksi dini risiko penyakit maupun prediksi kondisi kesehatan di masa depan [1].

Salah satu metode yang sering digunakan dalam machine learning adalah regresi, khususnya untuk memprediksi nilai kontinu. Metode ini memiliki keunggulan dalam menjelaskan hubungan antarvariabel sehingga dapat memberikan interpretasi yang lebih jelas terkait faktor-faktor yang memengaruhi kondisi kesehatan [2]. Namun, penerapan regresi sering kali dihadapkan pada masalah banyaknya fitur yang tidak semuanya relevan, sehingga model dapat menghasilkan nilai Mean Squared Error (MSE) yang tinggi dan akurasi prediksi yang kurang optimal. Untuk mengatasi tantangan tersebut, teknik seleksi fitur seperti Recursive Feature Elimination (RFE) dapat diterapkan. Sebelum diterapkannya RFE, model cenderung memiliki performa yang kurang baik akibat adanya fitur-fitur yang tidak memberikan kontribusi signifikan. Setelah RFE diterapkan, fitur yang paling relevan dapat diidentifikasi sehingga model menjadi lebih sederhana, stabil, serta mampu meningkatkan akurasi prediksi sekaligus menurunkan nilai MSE [3].

Dalam konteks kesehatan masyarakat, isu yang mendapat perhatian global adalah Penyakit Tidak Menular (PTM) yang menjadi penyebab utama kematian sekaligus ancaman bagi pertumbuhan ekonomi. World Health Organization (WHO) pada tahun 2015 melaporkan bahwa PTM menyebabkan kematian sekitar

40 juta orang setiap tahunnya, atau setara dengan 70% kematian di seluruh dunia [4]. Di Indonesia, prevalensi PTM terus meningkat. Berdasarkan data Riset Kesehatan Dasar (Riskesdas) 2018, prevalensi hipertensi pada populasi dewasa Indonesia mencapai 34,1%, sementara diabetes melitus sebesar 10,9% [4]. Kondisi ini diperburuk oleh rendahnya kesadaran masyarakat terhadap pentingnya deteksi dini. Oleh karena itu, pemanfaatan pendekatan berbasis teknologi, termasuk *machine learning* dan regresi dengan dukungan seleksi fitur, menjadi strategi penting dalam mendukung analisis prediktif kesehatan Masyarakat [1].

Penelitian ini membangun model prediksi skor kesehatan individu menggunakan regresi yang dipadukan dengan RFE untuk memilih fitur paling relevan, berdasarkan data gaya hidup dan kondisi fisik yang mencakup faktor usia, indeks massa tubuh, status merokok dan lain sebagainya [5]. Model dievaluasi menggunakan Mean Squared Error (MSE) dan R Squared (R^2), sehingga mampu menunjukkan kemampuan prediksi yang optimal. Hasil penelitian ini diharapkan dapat mendukung pengembangan platform berbasis web yang memungkinkan masyarakat memantau kondisi kesehatannya secara cepat, praktis dan akurat. [6]

1.2 Rumusan Masalah

Masalah yang akan diidentifikasi dalam penelitian ini adalah :

1. Bagaimana penerapan metode regresi dalam membangun model prediksi skor kesehatan berdasarkan data gaya hidup dan kondisi fisik?
2. Bagaimana seleksi fitur Recursive Feature Elimination (RFE) dalam menentukan variabel paling relevan untuk meningkatkan akurasi prediksi skor kesehatan?
3. Bagaimana hasil evaluasi kinerja model prediksi skor kesehatan menggunakan indikator Mean Squared Error (MSE) dan R Squared (R^2)?
4. Bagaimana pengembangan platform berbasis web dapat mendukung penerapan model prediksi skor kesehatan agar mudah diakses oleh

masyarakat?

1.3 Batasan Masalah

Dalam penelitian ini ditetapkan agar fokus kajian lebih terarah dan sesuai dengan tujuan yang ingin dicapai. Batasan tersebut meliputi:

1. Model prediksi yang digunakan hanya berfokus pada metode regresi dengan dukungan seleksi fitur Recursive Feature Elimination (RFE).
2. Evaluasi model dibatasi pada metrik Mean Squared Error (MSE) dan R Squared (R^2).
3. Penelitian hanya menggunakan data skunder dan belum melibatkan data primer.
4. Penelitian ini hanya berfokus pada gaya hidup dan kondisi fisik masyarakat Indonesia, sehingga temuan yang diperoleh belum dapat digeneralisasikan ke negara lain.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah :

1. Menerapkan metode regresi untuk membangun model prediksi skor kesehatan berdasarkan data gaya hidup dan kondisi fisik.
2. Menganalisis efektivitas seleksi fitur Recursive Feature Elimination (RFE) dalam memilih variabel paling relevan untuk prediksi skor kesehatan.
3. Mengevaluasi kinerja model prediksi menggunakan metrik Mean Squared Error (MSE) dan R Squared (R^2) untuk memperoleh hasil yang optimal.
4. Mengembangkan rancangan awal platform berbasis web yang dapat memanfaatkan model prediksi skor kesehatan untuk pemantauan kesehatan secara praktis.

1.5 Manfaat Penelitian

1.5.1 Manfaat Teoritis

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan ilmu pengetahuan, khususnya di bidang teknologi informasi dan data science, terkait penerapan algoritma machine learning dengan metode regresi. Dukungan seleksi fitur Recursive Feature Elimination (RFE) digunakan untuk meningkatkan akurasi dengan memilih variable paling relevan, sehingga model dihasilkan lebih optimal. Selain itu penelitian ini dapat menjadi referensi bagi peneliti selanjutnya dalam pengembangan sistem prediksi kesehatan berbasis data gaya hidup dan fisik, khususnya untuk mendukung upaya deteksi dini penyakit tidak menular dan efisien.

1.5.2 Manfaat Praktis

Sistem prediksi yang dikembangkan dalam penelitian ini diharapkan dapat membantu masyarakat dan tenaga kesehatan dalam mengidentifikasi risiko penyakit tidak menular secara lebih cepat dan akurat. Dengan informasi yang dihasilkan dari model berbasis regresi yang didukung seleksi fitur Recursive Feature Elimination (RFE), langkah pencegahan maupun penanganan dapat dilakukan sejak dini, seperti perubahan pola hidup atau pemeriksaan lanjutan. Hal ini berpotensi menurunkan risiko komplikasi serta mendukung peningkatan kualitas kesehatan masyarakat di Indonesia.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN: Bab ini berisi kajian teoritis yang mendukung penelitian yang dilakukan. Isi bab ini meliputi latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, serta batasan masalah. Bab ini memberikan kerangka awal mengenai alasan penelitian dilakukan dan fokus yang akan dicapai.

BAB II TINJAUAN PUSTAKA, Bab ini membahas terkait studi literatur yang membandingkan ke enam referensi yang memiliki kemiripan dengan penelitian ini, untuk referensinya itu maksimal 5 tahun terakhir. Selain itu, disertakan tabel

keaslian penelitian yang memuat judul, nama penulis, tahun publikasi, isi penelitian, dan perbandingannya dengan penelitian ini. Bab ini juga membahas dasar teori yang menjadi landasan metodologi dan analisis yang diterapkan dalam penelitian.

BAB III METODE PENELITIAN, Bab ini menjelaskan tahapan rancangan penelitian secara rinci. Rancangan penelitian mencakup objek penelitian, yaitu masyarakat dan tenaga kerja, dengan karakteristik data seperti demografi dan gaya hidup (usia, jenis kelamin, BMI, frekuensi olahraga, jam tidur, dan kebiasaan merokok). Selanjutnya dijelaskan alur penelitian, yang mencakup pengumpulan data, pemilihan fitur, penerapan algoritma, hingga evaluasi dan interpretasi hasil, dengan penjabaran tiap tahapan agar proses penelitian mudah dipahami. Selain itu, diuraikan alat dan bahan yang digunakan, termasuk perangkat lunak dan instrumen pendukung, sebagai bagian dari pelaksanaan rancangan penelitian.

BAB IV HASIL DAN PEMBAHASAN, Bab ini membahas tahapan penelitian yang meliputi data yang digunakan, pengolahan data, pemilihan fitur, pembangunan model, serta evaluasi dan interpretasi hasil. Data yang digunakan terdiri dari variabel terkait karakteristik dan gaya hidup masyarakat serta tenaga kerja, seperti usia, BMI, frekuensi olahraga, jam tidur, dan kebiasaan merokok. Dari keseluruhan variabel, dilakukan seleksi fitur menggunakan Recursive Feature Elimination (RFE) untuk menentukan fitur paling relevan yang mempengaruhi skor kesehatan. Selanjutnya, data yang terpilih diolah dan digunakan untuk membangun model prediksi menggunakan Polynomial Regression, kemudian dievaluasi performanya menggunakan metrik MSE, dan R^2 . Hasil prediksi divisualisasikan melalui grafik atau scatter plot, dan model kemudian diimplementasikan ke dalam bentuk website agar dapat digunakan secara interaktif. Pembahasan difokuskan pada interpretasi performa model, kontribusi masing-masing fitur terhadap skor kesehatan, serta penerapan model dalam website agar proses analisis dan penggunaan aplikasi dapat dipahami secara menyeluruh.

BAB V PENUTUP, Bab ini berisi kesimpulan dan saran dari seluruh rangkaian

penelitian yang telah dilakukan. Kesimpulan menjelaskan hasil temuan penelitian, termasuk efektivitas model Polynomial Regression dengan seleksi fitur Recursive Feature Elimination (RFE) dalam memprediksi skor kesehatan masyarakat dan tenaga kerja, fitur-fitur paling relevan, serta performa model berdasarkan metrik evaluasi seperti MSE, dan R^2 . Saran memberikan rekomendasi untuk penelitian selanjutnya, antara lain memperluas variasi data agar identifikasi fitur lebih akurat, mengeksplorasi penggunaan model hybrid untuk meningkatkan akurasi prediksi, serta menambahkan metrik evaluasi lain seperti MAE dan RMSE agar analisis performa model menjadi lebih komprehensif

