

BAB I

PENDAHULUAN

A. Latar Belakang

Diabetes merupakan kondisi kronis yang berkembang ketika pankreas mengalami disfungsi dalam produksi insulin. Insulin, yang berperan sebagai hormon pengatur kadar glukosa darah, menjadi elemen sentral dalam proses metabolisme. *Hiperglikemia*, yang secara klinis dikenal sebagai peningkatan kadar glukosa dalam darah, merupakan konsekuensi umum dari diabetes yang tidak terkontrol. Dalam jangka waktu yang panjang, kondisi ini dapat menyebabkan kerusakan yang signifikan pada berbagai sistem organ, terutama sistem saraf dan pembuluh darah (WHO, 2024).

Berdasarkan data dari Edisi ke-10 Atlas Diabetes *IDF (International Diabetes Federation)*, prevalensi diabetes terus meningkat di seluruh dunia, dengan perkiraan 537 juta orang dewasa (usia 20-79 tahun) hidup dengan diabetes pada tahun 2021. Pada tahun 2030, jumlah ini diprediksi meningkat menjadi 643 juta, dan pada tahun 2045 diperkirakan mencapai 783 juta orang dewasa dengan diabetes. Meskipun populasi dunia diperkirakan akan tumbuh sebesar 20% selama periode tersebut, jumlah penderita diabetes diperkirakan akan meningkat sebesar 46%. Indonesia sendiri menempati peringkat kelima dunia dengan 19,47 juta penderita diabetes pada tahun 2021 (IDF, 2021). Lonjakan ini mengindikasikan bahwa sistem kesehatan perlu melakukan transformasi signifikan dalam upaya pencegahan dan deteksi dini.

Diabetes tidak hanya disebabkan oleh faktor genetik, tetapi juga dipengaruhi oleh pola hidup, seperti kurangnya pemeriksaan kadar glukosa darah secara berkala, konsumsi makanan tinggi karbohidrat dan lemak, minimnya aktivitas fisik, serta kebiasaan mengonsumsi makanan dan minuman tidak sehat. Keterlambatan dalam deteksi dini juga turut memperburuk kondisi penderita (RS Pondok Indah, 2024). Oleh karena itu, diperlukan pendekatan

berbasis teknologi untuk membantu memprediksi risiko diabetes sejak dini agar pencegahan dapat dilakukan secara optimal.

Seiring dengan kemajuan teknologi digital, pemanfaatan sistem informasi dalam bidang kesehatan berkembang pesat. Salah satu solusi potensial adalah pengembangan sistem prediksi risiko diabetes berbasis web yang memungkinkan masyarakat melakukan evaluasi risiko secara mandiri berbasis data medis secara cepat, praktis, dan mudah diakses. Namun, sebagian besar sistem yang ada saat ini hanya fokus pada memberikan hasil prediksi, tanpa menyertakan elemen edukatif yang membantu pengguna memahami faktor risiko dan langkah-langkah pencegahan. Menurut Arminarahmah dan Mahalisa (2024) sistem prediksi penyakit diabetes berpotensi memberikan dampak positif dalam pengelolaan diabetes. Penyediaan informasi yang cepat dan akurat kepada pengguna dapat mendukung pengambilan keputusan yang lebih tepat terkait perawatan serta penyesuaian gaya hidup (Arminarahmah and Mahalisa, 2024). Dengan demikian, perancangan sistem prediksi penyakit diabetes berbasis web menggunakan algoritma *Random Forest* dapat menjadi solusi yang efektif. Model prediksi dapat digunakan sebagai sarana bantu bagi profesional medis maupun masyarakat untuk memperkirakan kemungkinan seseorang mengidap diabetes.

Algoritma *Random Forest* merupakan salah satu algoritma *Machine Learning* yang menggabungkan beberapa pohon keputusan (*Decision Trees*) untuk menghasilkan prediksi yang lebih akurat dan mengurangi *overfitting*. (Sriyanto and Supriyatma, 2023). Dalam penelitian ini menggunakan data sekunder dari situs web *Kaggle*, yaitu *Pima Indians Diabetes Database*, yang berisi 768 entri dan 9 atribut yang berhubungan dengan faktor risiko diabetes (*UCI Machine Learning, n.d*). Namun, dataset ini memiliki tantangan berupa ketidakseimbangan data, di mana jumlah data pasien tidak diabetes lebih banyak dibandingkan pasien diabetes. Kondisi ini dapat menimbulkan bias model dalam mengenali kelas mayoritas dan melemahkan kemampuan mendeteksi pasien diabetes. Untuk mengatasi ketidakseimbangan data, digunakan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) agar

distribusi data lebih seimbang. Selain itu, dilakukan *hyperparameter tuning* untuk mengoptimalkan performa model. Model prediksi yang telah dilatih kemudian akan diimplementasikan dalam bentuk aplikasi web menggunakan *Streamlit Cloud* yang terhubung dengan *GitHub*, sehingga dapat diakses secara publik oleh pengguna.

Studi sebelumnya oleh Salsabil, Azizah, dan Eviyanti (2024) yang juga menggunakan dataset ini menunjukkan bahwa algoritma *Random Forest* dengan akurasi 74%, hanya terbatas pada implementasi skrip *Python* (Salsabil *et al.* 2024). Sedangkan penelitian Muhadi dan Amir Ali (2025) menggunakan dataset yang sama telah mengembangkan sistem web sederhana berbasis *Random Forest* dengan akurasi 76%, namun masih terbatas pada pemberian hasil prediksi positif/negatif belum dilengkapi fitur edukatif dan saran pencegahan (Muhadi *et al.* 2025). Kedua penelitian tersebut menunjukkan bahwa meskipun algoritma yang digunakan cukup akurat, masih terdapat celah dalam hal aksesibilitas sistem, penyajian informasi edukatif, dan kemudahan penggunaan bagi masyarakat umum. Performa model juga sangat bergantung pada teknik pemrosesan data yang digunakan yang menunjukkan masih adanya ruang untuk eksplorasi dan peningkatan performa.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk merancang sistem prediksi risiko diabetes berbasis web menggunakan algoritma *Random Forest* dengan teknik SMOTE. Sistem ini tidak hanya dirancang untuk memberikan hasil prediksi secara akurat, tetapi juga dilengkapi dengan fitur edukasi diabetes secara umum dan rekomendasi pencegahan yang disesuaikan dengan hasil prediksi masing-masing (risiko rendah/tinggi diabetes). Sistem ini diharapkan dapat menjadi alat bantu deteksi dini dan sebagai media edukasi yang mendukung peningkatan literasi kesehatan masyarakat dalam upaya pencegahan diabetes secara berkelanjutan.

B. Perumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan mengenai tingginya prevalensi penyakit diabetes, kurangnya deteksi dini, serta pentingnya pemanfaatan teknologi prediktif yang mudah diakses oleh masyarakat luas, maka dalam penelitian ini dirumuskan beberapa permasalahan yang menjadi fokus utama kajian, yaitu:

1. Bagaimana merancang dan mengembangkan sistem prediksi penyakit diabetes berbasis web menggunakan algoritma *Random Forest*?
2. Sejauh mana akurasi sistem prediksi yang dihasilkan menggunakan algoritma *Random Forest* dengan teknik SMOTE dalam memprediksi risiko diabetes menggunakan *Pima Indian Diabetes Database*?

C. Tujuan

Penelitian ini diharapkan dapat memberikan kontribusi yang selaras dengan rumusan masalah yang telah dirumuskan sebelumnya, diantaranya yaitu:

1. Merancang dan mengembangkan sistem prediksi diabetes berbasis web yang dapat diakses dengan mudah oleh tenaga medis maupun masyarakat untuk mendeteksi kemungkinan risiko diabetes secara dini, dilengkapi dengan fitur edukasi.
2. Mengukur akurasi dan efektivitas algoritma *Random Forest* dengan teknik SMOTE dalam memprediksi risiko diabetes menggunakan *Pima Indians Diabetes Database*.

D. Manfaat

Penelitian ini diharapkan dapat memberikan manfaat tentang Perancangan Sistem Prediksi Penyakit Diabetes Berbasis Web menggunakan algoritma *Random Forest*:

1. Bagi Penulis
 - a. Sebagai syarat untuk memperoleh gelar Sarjana Informatika Medis.
 - b. Menjadi pengalaman dalam menerapkan ilmu yang diperoleh selama masa studi ke dalam penelitian nyata.
 - c. Meningkatkan pemahaman dan keterampilan dalam pengembangan sistem berbasis web menggunakan algoritma *Random Forest*.
2. Bagi Institusi Pendidikan
 - a. Menambah referensi penelitian dalam bidang Informatika Medis terkait penerapan *Machine Learning* untuk analisis kesehatan.
 - b. Mendorong mahasiswa untuk mengembangkan penelitian berbasis teknologi dalam upaya meningkatkan kualitas kesehatan.
3. Bagi Tenaga Kesehatan
 - a. Mempermudah analisis risiko diabetes dengan hasil prediksi yang cepat dan berbasis data.
 - b. Menyediakan alat bantu dalam proses diagnosis dini dan pengambilan keputusan terkait pencegahan serta penanganan diabetes.

E. Keaslian Penelitian

Penelitian ini berfokus pada perancangan sistem prediksi penyakit diabetes berbasis web menggunakan algoritma *Random Forest*. Meskipun terdapat beberapa penelitian sebelumnya yang menggunakan algoritma *Random Forest* untuk prediksi diabetes, penelitian ini memiliki perbedaan signifikan, terutama dalam integrasi algoritma tersebut kedalam platform berbasis web yang dapat diakses oleh masyarakat umum dan tenaga medis. Berikut adalah matriks perbandingan antara penelitian ini dengan penelitian terdahulu:

Tabel 1. 1 Keaslian Penelitian

No	Judul Penelitian dan Nama Penulis	Metode Penelitian	Hasil Penelitian	Perbedaan dengan Penelitian Penulis
1	Implementasi <i>Data Mining</i> Dalam Melakukan Prediksi Penyakit Diabetes Menggunakan Metode <i>Random Forest</i> Dan <i>Xgboost</i> , Muhammad Salsabil; Nuril Lutvi Azizah dan Ade Eviyanti.	<i>Random Forest, Xgboost</i>	Penelitian tersebut hanya mengukur akurasi dengan skrip <i>Python</i> . Hasil dari implementasi kedua teknik <i>Data Mining</i> tersebut dengan nilai akurasi <i>Random Forest</i> sebesar 74%, <i>XGBoost</i> sebesar 76% dengan menggunakan metode <i>Z-Score</i> untuk mengatasi <i>outlier</i> ; juga menerapkan <i>hyperparameter</i> (Salsabil <i>et al.</i> , 2024)	Dalam penelitian ini berfokus pengembangan sistem berbasis web, penulis meningkatkan performa algoritma <i>Random Forest</i> pada tahap <i>preprocessing</i> menggunakan teknik SMOTE untuk menyeimbangkan kelas pada data.
2	Sistem Informasi Model Prediksi Penyakit Diabetes Menggunakan Algoritma <i>Random Forest</i> di Fasyankes, Muhadi dan Amir Ali.	<i>Random Forest</i>	Penelitian tersebut menghasilkan model prediksi dengan tingkat akurasi 76%, tampilan sistem informasi hanya berbentuk form sistem prediksi yang menghasilkan <i>output</i> positif/negatif diabetes (Muhadi <i>and</i> Ali, 2025)	Pada penelitian ini mencari akurasi terbaik dengan <i>hyperparameter tuning</i> dan teknik SMOTE. Model prediksi diimplementasikan dalam bentuk sistem web interaktif yang dapat diakses oleh masyarakat secara luas dan dilengkapi dengan fitur edukasi kesehatan mengenai penyakit diabetes.

3	<i>Comparison Support Vector Machine and Random Forest Algorithms in Detect Diabetes</i> , Habib Alrasyid, Ahmad Homaidi, Zaehol Fatah	<i>SVM, Random Forest</i>	Pada penelitian tersebut melakukan perbandingan pada kedua algoritma pemodelan, kemudian model (SVM) dengan akurasi terbaik dirancang sebagai web prediksi berisi form prediksi dengan <i>output</i> positif/negatif diabetes (Alrasyid <i>et al.</i> , 2024)	Penelitian ini, melakukan perancangan sistem prediksi berbasis web menggunakan algoritma <i>Random Forest</i> yang menghasilkan <i>output</i> Risiko rendah/ Risiko Tinggi diabetes sebagai diagnosa awal dilengkapi dengan edukasi mengenai penyakit diabetes.
4	Implementasi Model <i>Machine Learning</i> pada Klasifikasi Status Penyakit Diabetes Berbasis <i>Streamlit</i> , Nur Arminarahmah dan Galih Mahalisa	<i>K-Nearest Neighbors (KNN)</i>	Implementasi model <i>K-Nearest Neighbors</i> dengan nilai akurasi 72%. Tampilan sistem hanya berupa <i>form</i> prediksi dengan <i>output</i> positif/negatif diabetes (Arminarahmah and Mahalisa, 2024).	Implementasi model <i>Random Forest</i> dengan tampilan web yang dilengkapi <i>ouput</i> edukasi.

Berdasarkan Tabel 1.1, dapat disimpulkan bahwa seluruh penelitian sebelumnya memiliki tujuan yang sama, yaitu membangun model prediksi penyakit diabetes menggunakan algoritma *Machine Learning*. Namun, terdapat beberapa perbedaan signifikan dari segi metode, implementasi, hingga fitur sistem yang dibangun, jika dibandingkan dengan penelitian ini.

Penelitian oleh Muhammad Salsabil; Nuril Lutvi Azizah dan Ade Eviyanti (2024) menggunakan algoritma *Random Forest* dan *XGBoost*, namun

hanya difokuskan pada perhitungan akurasi menggunakan skrip *Python* tanpa implementasi ke dalam sistem berbasis web. Meskipun menggunakan *hyperparameter* untuk mendapatkan akurasi yang optimal, namun penelitian tersebut tidak menerapkan teknik penyeimbangan data seperti SMOTE. Dalam penelitian ini, penulis tidak hanya membangun model *Random Forest*, tetapi juga melakukan *preprocessing* dengan SMOTE untuk menyeimbangkan data, serta mengembangkan sistem prediksi yang dapat diakses publik secara daring.

Pada penelitian oleh Muhadi dan Amir Ali (2025) telah membangun sistem informasi prediksi diabetes menggunakan algoritma *Random Forest* dan menghasilkan akurasi sebesar 76%. Penelitian tersebut belum menggunakan teknik SMOTE dan *hyperparameter tuning*, serta sistem yang dikembangkan masih terbatas hanya pada form prediksi positif/negatif, tanpa menyertakan fitur edukasi atau optimasi model lebih lanjut. Berbeda dengan itu, sistem yang penulis kembangkan tidak hanya menyajikan hasil prediksi, tetapi juga dilengkapi dengan informasi edukatif dan saran pencegahan, serta penerapan SMOTE hingga hyperparameter tuning untuk meningkatkan performa model.

Pada penelitian Habib Alrasyid, Ahmad Homaidi, Zaehol Fatah (2024), dilakukan perbandingan antara algoritma *Support Vector Machine* (SVM) dan *Random Forest*, dan sistem prediksi telah diimplementasikan dalam bentuk web. Meskipun demikian, sistem tersebut membangun sistem dengan algoritma *Support Vector Machine* (SVM) hanya memberikan hasil prediksi tanpa menyajikan edukasi kesehatan sebagai bagian dari *output*. Hal ini menjadi pembeda dari penelitian ini yang lebih menekankan aspek edukatif sebagai sarana peningkatan literasi kesehatan pengguna.

Kemudian penelitian oleh Nur Arminarahmah dan Galih Mahalisa menggunakan algoritma *K-Nearest Neighbors* (KNN) dan telah mengimplementasikannya menggunakan *Streamlit* namun hanya menghasilkan form prediksi tanpa edukasi tambahan. Nilai akurasi pun lebih rendah, yaitu sekitar 72%. Sedangkan penelitian ini menggunakan algoritma *Random Forest* untuk mendapatkan akurasi yang lebih tinggi, dan sistem web yang dibangun juga dilengkapi dengan konten edukasi sesuai hasil prediksi.

Dapat disimpulkan bahwa penelitian ini memiliki keunggulan utama dalam menggabungkan teknik SMOTE untuk menyeimbangkan data, *hyperparameter tuning* untuk meningkatkan performa model, serta pengembangan sistem web interaktif menggunakan *Streamlit*. Selain memberikan hasil prediksi, sistem juga menyajikan fitur edukasi kesehatan berdasarkan hasil prediksi pengguna. Dengan demikian, penelitian ini tidak hanya fokus pada aspek teknis akurasi, tetapi juga mendukung literasi kesehatan masyarakat melalui pendekatan teknologi yang mudah diakses.

F. Batasan Masalah

Dalam penelitian ini, terdapat beberapa batasan masalah yang perlu diperhatikan, antara lain:

1. Dataset yang digunakan terbatas pada data yang diperoleh dari platform *Kaggle* dan penelitian terkait yang digunakan dalam pengembangan sistem, dengan jumlah entri dan indikator yang terbatas.
2. Sistem prediksi yang dikembangkan hanya memprediksi risiko diabetes berdasarkan indikator yang ada dalam dataset, seperti *Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age, Outcome*.
3. Pengembangan sistem berbasis web hanya akan mencakup aspek perancangan dan implementasi algoritma *Random Forest* untuk memprediksi diabetes, tanpa mempertimbangkan integrasi dengan perangkat medis atau alat diagnostik lainnya.

Akurasi yang diukur akan terbatas pada pengujian model *Random Forest* dengan dataset yang digunakan, sehingga mungkin tidak mencakup semua kemungkinan variabel yang dapat mempengaruhi prediksi diabetes.