BAB I
PENDAHULUAN

1.1 Latar Belakang Masalah

Diabetes merupakan salah satu penyakit degeneratif kronis yang menjadi tantangan utama dalam dunia kesehatan. Menurut *World Health Organization* (WHO), diabetes terjadi akibat ketidakcukupan produksi hormon insulin oleh pankreas atau ketidakmampuan tubuh dalam menggunakan hormon insulin secara efektif. Kondisi ini ditandai dengan hiperglikemia atau peningkatan kadar glukosa darah sebagai indikator utama. Permasalahan utama dari penyakit ini adalah sifatnya yang sering kali tidak menunjukkan gejala khas, sehingga banyak penderita tidak menyadari bahwa mereka mengidap diabetes hingga timbul komplikasi serius. Oleh karena itu, diagnosis medis melalui pemeriksaan kadar glukosa darah menjadi langkah krusial dalam mendeteksi keberadaan penyakit ini secara akurat dan dini. Dampak jangka panjang dari diabetes yang tidak terdiagnosis atau tidak terkontrol mencakup berbagai komplikasi serius seperti kerusakan ginjal (nefropati), gangguan penglihatan (retinopati), penyakit jantung, stroke, serta amputasi akibat infeksi kronis. Hal ini menjadikan diabetes sebagai salah satu penyakit dengan beban ekonomi dan sosial yang sangat tinggi di masyarakat (Apriliah et al., 2021).

Diabetes merupakan kondisi yang ditandai oleh peningkatan kadar glukosa dalam darah (hiperglikemia), yang disebabkan oleh ketidakmampuan tubuh dalam memproduksi hormon insulin secara memadai atau gangguan dalam penggunaan hormon insulin secara optimal (Nurlan, 2023). Hormon Insulin adalah hormon yang memiliki peran vital dalam mengatur metabolisme gula darah, dengan cara membantu glukosa masuk ke dalam sel untuk digunakan sebagai energi. Selain itu, hormon insulin juga berfungsi menghambat produksi glukosa yang berlebihan, sehingga kadar gula darah tetap berada dalam batas normal.

Aspek genetik merupakan salah satu faktor penting yang perlu dipertimbangkan dalam kasus penyakit diabetes. Meskipun tergolong sebagai penyakit tidak menular (PTM), diabetes termasuk penyakit keturunan yang melibatkan banyak faktor genetik, yaitu dipengaruhi oleh interaksi beberapa gen, bukan hanya satu gen tunggal (Faida et al., 2020).

Di samping itu, pola hidup yang tidak sehat juga menjadi pemicu utama bertambahnya jumlah penderita diabetes. Oleh karena itu, penanganan dan pengelolaan diabetes secara tepat sangat diperlukan guna mencegah komplikasi yang dapat memperburuk kualitas hidup pasien dan membebani sistem layanan kesehatan (Yusnanda et al., 2018).

Pengelolaan yang optimal bagi individu dengan diabetes sangat krusial untuk mendukung deteksi dini penyakit ini. Identifikasi yang akurat sejak awal memungkinkan dilakukannya intervensi segera guna menstabilkan kadar glukosa darah dan meminimalkan risiko terjadinya komplikasi yang lebih berat (Elfaladonna & Rahmadani, 2019). Pemanfaatan teknologi seperti data *mining* memiliki potensi signifikan dalam mengekstraksi pola dan informasi berharga dari kumpulan data terkait diabetes. Salah satu teknik data *mining* yang dapat diterapkan untuk membantu proses diagnosis diabetes adalah metode klasifikasi. Klasifikasi merupakan proses pembelajaran mesin yang bertujuan untuk menempatkan data ke dalam kategori atau kelas tertentu yang telah ditetapkan sebelumnya (Nawawi et al., 2024). Melalui pendekatan ini, data pasien dapat dikelompokkan ke dalam beberapa kategori berdasarkan karakteristik atau fitur yang dimiliki.

Terdapat berbagai jenis algoritma yang dapat digunakan, salah satunya adalah algoritma *Random Forest*. Algoritma ini termasuk dalam keluarga algoritma pohon keputusan (*Decision Tree*) yang bekerja dengan membangun sejumlah pohon melalui proses pelatihan pada sampel data untuk membentuk suatu model prediktif (Febtiawan et al., 2024). Keunggulan dari algoritma *Random Forest* terletak pada tingkat akurasinya yang tinggi serta kemampuannya dalam mengolah data berukuran besar, selain itu *Random Forest* juga melakukan seleksi fitur secara otomatis, sehingga fitur-fitur paling relevan dapat dipilih untuk meningkatkan kinerja model klasifikasi (Fauzi et al., 2025).

Klasifikasi diabetes yang telah dilakukan dengan menerapkan algoritma *Random Forest*. Studi tersebut menghasilkan tingkat akurasi sebesar 74,78%. Terdapat tiga variabel utama yang dianggap memiliki pengaruh besar (*variable importance*) dalam proses klasifikasi. Pada penelitian yang akan dilakukan, peneliti berencana menggunakan pendekatan *Synthetic Minority Oversampling Technique* (SMOTE) maupun tanpa SMOTE dalam tahap *resampling* setelah proses pelatihan dan pengujian data. Teknik SMOTE digunakan untuk membantu menyeimbangkan distribusi data yang tidak seimbang (Hadi & Sirodj, 2023).

Dalam beberapa tahun terakhir, diabetes telah menjadi permasalahan kesehatan global dengan tingkat prevalensi dan angka kematian yang terus meningkat. Data menunjukkan bahwa pada tahun 2011, diabetes menyebabkan sekitar 4,6 juta kematian di seluruh dunia. Kadar glukosa darah yang tinggi juga berkontribusi terhadap 2,2 juta kematian pada tahun 2012, dan secara langsung menyebabkan 1,6 juta kematian pada tahun 2016 (Setiawan et al., 2024). Jumlah penderita diabetes pun meningkat drastis, dari 415 juta orang pada tahun 2015, dan diperkirakan akan mencapai 642 juta orang pada tahun 2040. Peningkatan ini dipicu oleh berbagai faktor seperti urbanisasi, perubahan pola makan, kurangnya aktivitas fisik, serta meningkatnya angka obesitas. Faktor keturunan (genetik) merupakan salah satu aspek penting yang perlu dipertimbangkan dalam kasus penyakit diabetes. Dalam konteks ini, faktor keturunan mengacu pada riwayat keluarga yang pernah menderita diabetes, seperti orang tua atau saudara kandung, yang dapat meningkatkan risiko terjadinya penyakit. Meskipun tergolong sebagai penyakit tidak menular (PTM), diabetes melibatkan interaksi banyak faktor genetik dan lingkungan, sehingga bukan hanya dipengaruhi oleh satu gen tunggal

Jumlah penderita diabetes secara global mengalami peningkatan yang signifikan dari tahun 2021 ke tahun 2045. Pada tahun 2021, Indonesia menempati peringkat kelima dunia dengan jumlah penderita diabetes sebanyak 19,5 juta orang. Angka ini diproyeksikan meningkat menjadi 28,6 juta pada tahun 2045, mempertahankan posisi kelima secara global. Peningkatan ini menunjukkan tren yang mengkhawatirkan, di mana prevalensi diabetes terus meningkat seiring waktu. Di sisi lain, negara-negara seperti China dan India juga menunjukkan lonjakan

penderita diabetes yang sangat besar, masing-masing diperkirakan akan mencapai 174,4 juta dan 124,9 juta penderita pada tahun 2045. Data ini menunjukkan bahwa diabetes bukan hanya menjadi masalah kesehatan nasional, tetapi juga global. Khususnya bagi Indonesia, tren ini menjadi peringatan serius terhadap sistem kesehatan nasional dan pentingnya strategi pencegahan serta prediksi risiko secara dini. Oleh karena itu, penelitian mengenai penerapan metode prediktif berbasis data klinis, seperti algoritma *Random Forest*, menjadi sangat relevan untuk membantu mengidentifikasi individu yang berisiko tinggi terkena diabetes, sehingga intervensi dapat dilakukan lebih awal guna menekan pertumbuhan jumlah penderita di masa depan.

Pada tahun 2021, jumlah penderita diabetes di Indonesia tercatat mencapai 19,5 juta orang. Angka ini diproyeksikan terus meningkat hingga mencapai 28,6 juta pada tahun 2045 (Zaininda & Utama, 2023). Diabetes menempati posisi ketiga sebagai penyebab kematian terbanyak di Indonesia, setelah stroke dan penyakit jantung, dengan kontribusi lebih dari 80% terhadap total kematian setiap tahunnya (Riskesdas, 2018).



Gambar 1.1 Jumlah Penderita Diabetes di Indonesia

Berdasarkan grafik "Penderita Diabetes di Indonesia", terlihat adanya tren peningkatan jumlah penderita diabetes yang signifikan dari tahun ke tahun. Berdasarkan data dari Kementerian Kesehatan, pada tahun 2019 jumlah penderita

diabetes di Indonesia tercatat sebesar 10,7 juta jiwa. Angka ini mengalami lonjakan tajam pada tahun 2020 menjadi 18,69 juta jiwa, yang menunjukkan peningkatan lebih dari 70% hanya dalam satu tahun. Kenaikan ini berlanjut pada tahun 2021, meskipun tidak setajam tahun sebelumnya, dengan jumlah penderita mencapai 19,5 juta jiwa. Tren peningkatan ini terus berlanjut hingga tahun 2023, di mana jumlah penderita diabetes mencapai angka tertinggi, yaitu 20,8 juta jiwa. Secara keseluruhan, dalam rentang waktu lima tahun dari 2019 hingga 2023, jumlah penderita diabetes di Indonesia hampir dua kali lipat. Data ini menunjukkan bahwa diabetes merupakan masalah kesehatan masyarakat yang semakin serius di Indonesia, dan hal ini memerlukan perhatian khusus baik dari segi pencegahan maupun penanganan melalui pendekatan berbasis data dan teknologi (Kemenkes, 2019).

Penyakit diabetes tidak hanya menjadi tantangan kesehatan global, tetapi juga krisis yang diam-diam berkembang di banyak negara berkembang, termasuk Indonesia. Berdasarkan data dari *International Diabetes Federation* (IDF), pada tahun 2021 tercatat bahwa lebih dari 10 juta orang di Indonesia menderita diabetes, menjadikannya salah satu dari sepuluh negara dengan penderita diabetes tertinggi di dunia. Akan tetapi, lebih dari separuh angka tersebut tidak menyadari bahwa mereka telah mengidap penyakit ini, sehingga terlambat dalam mendapatkan penanganan medis yang memadai. Kondisi ini menandakan lemahnya sistem deteksi dini serta rendahnya kesadaran masyarakat terhadap pentingnya pemeriksaan kesehatan secara berkala.

Di sisi lain, tantangan terbesar dalam manajemen penyakit diabetes adalah kemampuan untuk mengidentifikasi individu yang berisiko tinggi secara efisien sebelum individu tersebut menunjukkan gejala klinis. Dalam konteks ini, pendekatan berbasis teknologi prediktif menjadi sangat relevan. Salah satu pendekatan yang saat ini banyak dikembangkan adalah penggunaan algoritma *machine learning* untuk mendeteksi risiko diabetes berdasarkan data klinis. Dengan memanfaatkan data rekam medis yang tersedia, model prediksi berbasis *machine learning* dapat menemukan pola tersembunyi yang sulit dikenali oleh metode konvensional.

Dari berbagai algoritma yang tersedia, *Random Forest* menjadi salah satu yang paling menjanjikan. Algoritma ini tidak hanya mampu menangani data dalam jumlah besar dan kompleks, tetapi juga menawarkan fleksibilitas dalam menganalisis berbagai jenis variabel klinis. Penelitian sebelumnya menunjukkan bahwa *Random Forest* mampu menghasilkan akurasi prediksi yang sangat tinggi dalam konteks diagnosis penyakit, termasuk diabetes. Namun demikian, belum banyak penelitian yang mengkaji efektivitas algoritma ini secara mendalam di Indonesia, terutama dengan menggunakan data klinis yang mencerminkan keragaman karakteristik populasi.

Tahapan awal diabetes dikenal dengan istilah pradiabetes, yakni kondisi ketika kadar glukosa darah seseorang lebih tinggi dari batas normal. Pada fase ini, tubuh mengalami ketidakmampuan dalam merespons hormon insulin secara efektif, yang menyebabkan pankreas bekerja lebih keras untuk memproduksi hormon hormon insulin. Karena tidak menunjukkan gejala yang jelas, pradiabetes sering kali hanya dapat diketahui melalui pemeriksaan medis. Padahal, deteksi dini pada tahap ini sangat penting karena dapat mencegah progresi menuju diabetes melalui perubahan gaya hidup dan pola makan sehat (Iskandar et al., 2024). Selain itu, pada tahap pradiabetes, intervensi yang dilakukan lebih mudah, murah, dan berpotensi memberikan dampak jangka panjang yang signifikan terhadap kualitas hidup individu. Edukasi masyarakat mengenai pentingnya *skrining* rutin dan pengelolaan gaya hidup sehat menjadi kunci utama dalam upaya pencegahan diabetes secara menyeluruh.

Salah satu algoritma yang telah menunjukkan performa unggul dalam tugas klasifikasi dan prediksi diabetes adalah *Random Forest*, yaitu algoritma berbasis *ensemble learning* yang menggabungkan sejumlah pohon keputusan (*Decision Trees*) untuk menghasilkan prediksi yang lebih stabil dan akurat. *Random Forest* mampu menangani *dataset* besar dengan kompleksitas tinggi, serta dikenal tahan terhadap *overfitting* (Majid et al., 2025).

Meskipun demikian, *Random Forest* juga memiliki sejumlah kekurangan. Pertama, model ini cenderung lebih sulit diinterpretasikan dibandingkan pohon keputusan tunggal karena hasil prediksi berasal dari kombinasi ratusan bahkan

ribuan pohon, sehingga transparansi dan keterjelasan logika prediksi menjadi berkurang. Kedua, proses pelatihan dan prediksi pada *Random Forest* bisa memerlukan sumber daya komputasi yang lebih besar, terutama jika jumlah pohon yang dibangun sangat banyak. Selain itu, pada *dataset* dengan jumlah fitur yang sangat tinggi, *Random Forest* berpotensi menghasilkan model yang lebih kompleks dan kadang kurang efisien. (Mahmuda, 2024)

Dalam penelitian yang dilakukan oleh (Suryanegara et al., 2021), algoritma ini mencapai akurasi hingga 97,88% dalam mendeteksi diabetes. Kemampuannya dalam melakukan seleksi fitur serta kemudahan interpretasi menjadikan *Random* dan diagnosis penyakit, termasuk diabetes (Maulana et al., 2024); (Junus et al., 2023). Peningkatan jumlah penderita diabetes secara signifikan di Indonesia menjadi ancaman serius bagi sistem kesehatan nasional. Fakta bahwa sebagian besar kasus tidak terdeteksi pada tahap awal menegaskan lemahnya sistem deteksi dini yang ada. Keterlambatan diagnosis menyebabkan komplikasi serius, meningkatkan beban ekonomi, serta menurunkan kualitas hidup pasien. Dalam konteks ini, diperlukan pendekatan baru yang mampu melakukan prediksi risiko secara lebih cepat dan akurat. Algoritma *Random Forest* sebagai bagian dari teknologi *machine learning* menawarkan potensi besar dalam pengolahan data klinis untuk menghasilkan model prediktif yang efisien dan andal.

Selain membangun model prediksi, implementasi hasil penelitian ke dalam bentuk aplikasi berbasis web menjadi langkah penting agar manfaatnya dapat dirasakan secara langsung oleh pengguna. Aplikasi ini diharapkan dapat diakses oleh tenaga medis maupun masyarakat umum untuk melakukan prediksi risiko diabetes secara cepat dan praktis, hanya dengan memasukkan data klinis yang relevan. Pemanfaatan teknologi berbasis web memberikan keunggulan dari sisi aksesibilitas dan kemudahan penggunaan, sehingga deteksi dini dapat dilakukan kapan saja dan di mana saja. Dengan demikian, pengembangan prototipe aplikasi web yang terintegrasi dengan model prediksi *Random Forest* diharapkan mampu menjadi alat bantu yang efektif dalam mendukung pencegahan dan penanganan penyakit diabetes di Indonesia.

Perkembangan teknologi informasi telah membawa perubahan signifikan dalam bidang kesehatan, terutama pada pemanfaatan *machine learning* untuk mendukung proses deteksi dini dan prediksi penyakit. Algoritma *Random Forest* yang terbukti memiliki performa baik dalam klasifikasi risiko diabetes tidak hanya berhenti pada tahap analisis data, tetapi juga perlu diimplementasikan ke dalam sebuah sistem yang dapat diakses dengan mudah oleh tenaga medis maupun masyarakat. Implementasi model prediksi ke dalam aplikasi berbasis web merupakan langkah strategis untuk menjembatani hasil penelitian dengan praktik nyata di lapangan. (Sukamto et al., 2025)

Aplikasi berbasis web memiliki keunggulan dalam hal aksesibilitas, fleksibilitas, dan kemudahan penggunaan, sehingga dapat dimanfaatkan oleh berbagai pihak tanpa memerlukan perangkat khusus. Dengan integrasi model prediksi *Random Forest* ke dalam aplikasi web, pengguna hanya perlu memasukkan data klinis sederhana, seperti usia, indeks massa tubuh, tekanan darah, kadar glukosa, atau riwayat keluarga untuk memperoleh estimasi risiko diabetes. Hal ini tidak hanya mempermudah proses skrining awal, tetapi juga dapat menjadi alat bantu dalam pengambilan keputusan medis yang lebih cepat dan tepat. (Fithriyaningrum et al., 2021)

Variabel-variabel klinis yang digunakan dalam penelitian ini dipilih berdasarkan keterkaitan ilmiahnya dengan mekanisme patofisiologi diabetes melitus. Kadar glukosa darah puasa dan glukosa darah sewaktu merupakan indikator langsung hiperglikemia yang menjadi dasar diagnosis diabetes. HbA1c mencerminkan rata-rata kadar gula darah selama 2–3 bulan terakhir sehingga menjadi penanda kontrol glukosa jangka panjang. Tekanan darah tinggi berkaitan erat dengan resistensi insulin dan sering muncul bersama diabetes sebagai bagian dari sindrom metabolik. Usia menjadi faktor risiko karena kemampuan pankreas memproduksi insulin cenderung menurun seiring bertambahnya umur (Rahman et al., 2023). Kolesterol dan trigliserida menggambarkan gangguan metabolisme lipid yang mempercepat kerusakan pembuluh darah pada penderita diabetes, sedangkan indeks massa tubuh berkaitan dengan obesitas yang meningkatkan resistensi insulin (Yustian et al., 2023).

Penelitian terdahulu juga membuktikan bahwa jumlah variabel yang digunakan berpengaruh langsung terhadap akurasi model. ANISHA et al. (2023) menunjukkan bahwa model *Random Forest* yang dilatih menggunakan variabel klinis lengkap — meliputi usia, kadar glukosa, tekanan darah, dan BMI — mampu mencapai akurasi 96,7%, presisi 95,2%, dan *recall* 94,8%. Sementara itu, Chari et al. (2019) membuktikan bahwa ketika jumlah fitur dikurangi dari delapan menjadi lima melalui seleksi fitur, akurasi model justru meningkat dari 93,45% menjadi 97,12%. Namun hal ini dicapai karena fitur yang dibuang memang redundan, bukan karena variabel klinis penting dihilangkan. Temuan ini menegaskan bahwa pemilihan variabel yang tepat dan representatif sangat menentukan kualitas prediksi (Zuhri et al., 2025). Oleh karena itu, penelitian ini menggunakan kombinasi 13 variabel klinis yang mencakup aspek demografis, fisiologis, dan laboratoris guna memaksimalkan kemampuan model dalam mengenali pola risiko diabetes.

Namun, pemanfaatan algoritma ini dalam konteks lokal Indonesia, khususnya berbasis data klinis dari fasilitas kesehatan di daerah, masih sangat terbatas. Oleh karena itu, penelitian ini yang berjudul "Implementasi Algoritma *Random Forest* untuk Prediksi Risiko Diabetes Berbasis Data Klinis" menjadi penting untuk dilakukan guna mengembangkan sistem deteksi dini yang akurat dan relevan dengan karakteristik pasien di Indonesia. Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam meningkatkan efektivitas pencegahan diabetes serta mendukung pengambilan keputusan medis secara lebih cepat dan tepat.

1.2 Identifikasi Masalah

Identifikasi masalah ini adalah:

1. Penyakit diabetes sering kali tidak terdeteksi sejak dini karena gejala awal yang samar dan mirip dengan penyakit lain, sehingga banyak penderita tidak menyadari kondisinya sampai muncul komplikasi serius.
2. Sistem deteksi dini yang tersedia saat ini masih memiliki keterbatasan, terutama dari sisi akurasi dan efisiensi dalam mengidentifikasi individu berisiko tinggi secara tepat.

3. Meskipun algoritma *Random Forest* memiliki potensi besar dalam klasifikasi penyakit seperti diabetes, pemanfaatannya secara optimal dalam konteks data klinis lokal di Indonesia masih belum banyak diteliti dan dikembangkan.
4. Hasil penelitian yang ada belum banyak diimplementasikan ke dalam bentuk aplikasi berbasis web yang dapat digunakan langsung oleh tenaga medis maupun masyarakat umum untuk mendukung deteksi dini.

1.3 Ruang Lingkup Masalah

Penelitian ini secara spesifik memfokuskan pada pengembangan model prediksi diabetes menggunakan algoritma *Random Forest*, yang dipilih karena kemampuannya menangani data kompleks dan mengurangi risiko *overfitting*. *Random Forest* merupakan algoritma pembelajaran mesin berbasis *ensemble* yang memiliki keunggulan dalam mengolah data dengan jumlah fitur yang besar serta menghasilkan prediksi yang stabil dan akurat, bahkan ketika dihadapkan pada data yang heterogen atau tidak terdistribusi secara merata. Dalam konteks prediksi penyakit kronis seperti diabetes, ketahanan algoritma ini terhadap *noise* dan kemampuannya dalam menangani data *non-linear* menjadi nilai tambah yang signifikan.

Data yang digunakan terbatas pada *dataset* rekam medis yang tersedia untuk umum dan diperoleh dari Rumah Sakit/pusat kesehatan, tanpa melakukan pengumpulan data baru. Dengan demikian, penelitian ini tidak melibatkan kegiatan survei, wawancara, atau eksperimen laboratorium yang memerlukan interaksi langsung dengan pasien atau subjek manusia. Penggunaan data sekunder ini dilakukan dengan tetap memperhatikan etika penelitian, terutama dalam menjaga kerahasiaan informasi pasien serta kepatuhan terhadap prinsip keamanan data dan perlindungan privasi. Selain itu, data yang digunakan akan melalui proses pembersihan dan praproses (*preprocessing*) terlebih dahulu agar siap digunakan dalam pelatihan model prediksi.

Ruang lingkup penelitian ini dibatasi pada validasi dan optimasi model prediksi *Random Forest* saja, dengan tujuan meningkatkan akurasi dan efisiensi

prediksi, dan tidak mencakup pembahasan implementasi model dalam aplikasi klinis atau sistem kesehatan yang lebih luas, integrasi model ke alur kerja klinis, atau kajian dampak implementasi model terhadap praktik medis. Penelitian ini juga tidak akan membahas aspek legal, ekonomi, atau kebijakan yang berkaitan dengan penerapan teknologi *machine learning* dalam sistem pelayanan kesehatan. Fokus utama ditekankan pada tahap merancang dan evaluasi teknis dari model prediktif, termasuk analisis performa menggunakan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Pembatasan ruang lingkup ini dimaksudkan agar penelitian dapat berjalan secara terfokus, mendalam, dan sistematis dalam menjawab pertanyaan penelitian terkait efektivitas algoritma *Random Forest* dalam mendeteksi risiko diabetes berdasarkan data yang tersedia.

1.4 Batasan Masalah

Agar penelitian ini tetap terfokus dan tidak meluas ke hal-hal yang tidak relevan, berikut adalah batasan masalah yang ditetapkan:

1. Penelitian ini hanya menggunakan algoritma *Random Forest* untuk prediksi diabetes.
2. *Dataset* yang digunakan adalah *dataset* rekam medis pasien yang tersedia untuk umum atau diperoleh dari Rumah Sakit/pusat kesehatan.
3. Implementasi model prediksi dilakukan dalam bentuk prototipe aplikasi berbasis web, namun tidak mencakup integrasi penuh dengan sistem klinis atau pembahasan aspek legal dan kebijakan.

1.5 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan di atas, dapat dirumuskan suatu pokok permasalahan yaitu:

1. Apa saja faktor-faktor klinis yang paling berpengaruh dalam proses prediksi risiko diabetes berdasarkan hasil klasifikasi dengan algoritma *Random Forest*?
2. Sejauh mana efektivitas algoritma *Random Forest* dalam mendeteksi risiko diabetes berdasarkan data rekam medis pasien?

3. Bagaimana performa algoritma *Random Forest* dalam membangun model prediksi risiko diabetes yang akurat dan dapat diandalkan?
4. Bagaimana mengimplementasikan model prediksi tersebut ke dalam aplikasi berbasis web yang dapat digunakan sebagai alat bantu deteksi dini?

1.6 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijelaskan sebelumnya, tujuan dari penelitian ini adalah sebagai berikut:

1. Merancang model prediksi risiko diabetes berbasis algoritma *Random Forest* yang dapat dijadikan alat bantu deteksi dini secara akurat dan efisien.
2. Mengidentifikasi dan menganalisis variabel-variabel klinis utama yang berperan signifikan dalam prediksi risiko diabetes menggunakan algoritma *Random Forest*.
3. Mengevaluasi efektivitas algoritma *Random Forest* dalam memprediksi risiko diabetes berdasarkan data rekam medis dari pasien.
4. Mengimplementasikan model prediksi risiko diabetes ke dalam prototipe aplikasi berbasis web yang dapat digunakan oleh tenaga medis maupun masyarakat umum.

1.7 Manfaat Penelitian

Berdasarkan latar belakang dan tujuan penelitian yang telah dijelaskan sebelumnya, penelitian ini diharapkan memberikan beberapa manfaat sebagai berikut:

1. Bagi pengguna:

   Memberikan kontribusi terhadap pengembangan model prediksi penyakit berbasis *machine learning*, khususnya dalam deteksi dini penyakit diabetes.
2. Bagi peneliti:

   Memberikan kemudahan bagi tenaga medis dalam mengidentifikasi pasien dengan risiko tinggi diabetes secara lebih cepat dan akurat.

3. Bagi tenaga medis:

   Mempermudah proses identifikasi pasien dengan risiko tinggi diabetes secara cepat, akurat, dan dapat diakses melalui perangkat digital.