

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Diabetes *mellitus* (DM) yaitu penyakit mematikan yang sering dikenal “*silent killer*” dikarenakan sering tidak terdiagnosis oleh pengidapnya dan diketahui ketika telah terjadi komplikasi. Kadar gula darah tidak normal, yaitu 200 mg/dL atau lebih, dipicu oleh gangguan disekresi atau kerja insulin, atau keduanya dikenal sebagai DM. Hiperglikemia atau kadar gula darah tinggi dapat menyebabkan komplikasi serius seperti penyakit jantung, stroke, masalah penglihatan, kerusakan saraf, dan gangguan ginjal (Abdurrahman, 2022). Penyebab gula darah tinggi seperti autoimun, gaya hidup menjadi dasar pengelompokan jenis diabetes. DM terdiri dari beberapa jenis yaitu diabetes tipe 1, diabetes tipe 2, diabetes gestasional. Namun jenis diabetes yang paling umum yaitu diabetes tipe 1 dan tipe 2 (Marzel, 2021). Meningkatnya jumlah penderita diabetes dari waktu ke waktu disebabkan oleh beberapa faktor, antara lain pola hidup yang tidak sehat (tinggi kalori, pola makan kurang gizi, kurang aktivitas fisik dan obesitas), faktor genetik (riwayat diabetes dalam keluarga), dan penuaan yang berhubungan dengan penurunan fungsi pankreas dan resistensi insulin (Rosa Tiurma & Syahrizal, 2021). Selain itu, faktor lain yang menyebabkan diabetes yaitu usia (Arania et al., 2021), aktivitas fisik (Syahputri, 2020), GDA (gula darah acak), GDP (gula darah puasa), gula darah 2 jam pp, tekanan darah (Julianti K, 2020), BMI (indeks massa tubuh), HbA1c (Saputra et al., 2020).

Kondisi di atas menjadikan Indonesia salah satu negara dengan banyak penderita diabetes. Menurut laporan “*International Diabetes Federation*” tahun 2021, terdapat 19,47 juta penderita diabetes di Indonesia, menempatkan negara ini pada peringkat kelima dunia (Pahlevi, 2021). Jumlah tersebut diperkirakan akan meningkat menjadi 23,32 juta pada tahun 2030. Pada tahun 2021, IDF melaporkan bahwasannya sekitar 73,7% penderita diabetes di Indonesia yang berjumlah sekitar 14,34 juta belum ter-diagnosis. Selain itu, terdapat 236.711 kematian yang diakibatkan oleh diabetes (IDF, 2021). Pada tahun 2021, jumlah penderita diabetes di dunia mencapai 537 juta. Angka ini diperkirakan akan terus bertambah hingga

643 juta pada tahun 2030 dan 784 juta pada tahun 2045. Selain itu, diperkirakan 6,7 juta orang meninggal akibat diabetes pada tahun 2021 (IDF, 2021). Jawa Timur, salah satu provinsi terbesar di Indonesia yang memiliki jumlah penderita diabetes yang signifikan, yaitu sebanyak 863.686 kasus pada tahun 2021. Dari jumlah tersebut, sekitar 842.004 pasien telah terdiagnosis dan mendapat layanan kesehatan (Dinkes Jatim, 2022).

Diabetes remaja yaitu istilah untuk diabetes tipe 1 yang disebabkan oleh penghancuran sel  $\beta$  secara autoimun atau idiopatik yang mengakibatkan defisiensi insulin. Diabetes tipe 2 memiliki ciri khas resistensi terhadap insulin, membedakannya dari jenis diabetes lainnya. Diabetes tipe 1 umumnya dianggap sebagai penyakit yang menyerang anak-anak dan remaja (Faida & Santik, 2020). Penelitian terbaru menunjukkan bahwasannya 40% orang dewasa di atas usia 30 tahun yang sebenarnya menderita diabetes tipe 1 didiagnosa sebagai diabetes tipe 2. Tantangan utama dalam kesalahan diagnosis ini yaitu asumsi umum bahwasannya orang dewasa dengan diabetes kemungkinan besar menderita diabetes tipe 2, dan beberapa orang dewasa dengan diabetes tipe 1 kemungkinan tidak memerlukan insulin saat diagnosis sehingga gejalanya menyerupai diabetes tipe 2. Akibat penatalaksanaan yang salah dapat berdampak negatif pada kualitas hidup dan kelangsungan hidup pasien (Katongole et al., 2022). Cheheltani et al (2022) menunjukkan bahwasannya hampir setengah dari kasus diabetes tipe 1 yang baru didiagnosis terjadi pada orang dewasa, namun sering salah didiagnosis sebagai diabetes tipe 2. Cara untuk mengatasinya, penelitian ini mengembangkan model pembelajaran mesin yang mempergunakan data rekam medis elektronik untuk mengidentifikasi pasien diabetes tipe 1 yang salah didiagnosis dengan prediktor utama seperti usia, BMI, riwayat terapi, nilai HbA1c dan glukosa darah.

Melihat tingginya penderita DM, maka dibutuhkan tindakan pencegahan baik secara langsung melalui tenaga medis maupun pendekatan model kuantitatif matematis. Penyakit diabetes dapat dicegah dengan cara mendeteksi sejak dini sehingga masyarakat menerapkan pola kehidupan yang sehat (W. Nugraha & Sabaruddin, 2021). Salah satu pendekatan model kuantitatif matematis yaitu dengan menggunakan *machine learning* (Dritsas & Trigka, 2022). Menurut IBM, *machine learning* yaitu cabang dari AI yang menggunakan data dan algoritma

untuk meniru proses pembelajaran manusia, dengan tujuan meningkatkan akurasi dalam membuat prediksi yang tepat dan benar (Faiza et al., 2022). Data terkait faktor penyakit diabetes yang tersedia dapat digunakan dengan pendekatan *supervised learning*, yang mengacu pada teknik klasifikasi. Klasifikasi yaitu tahap menemukan pola atau karakteristik yang mendeskripsikan serta membedakan kelas data atau ide (Saputro & Rosiyadi, 2022). Penelitian ini menggunakan metode klasifikasi *Random Forest*.

*Random Forest* yaitu kemajuan dari teknik *Classification and Regression Trees* (CART) yakni bagian dari teknik pohon keputusan. Perbedaan utama metode *Random Forest* dengan metode CART yaitu penggunaan metode *aggregating bootstrap (bagging)* dan penggunaan seleksi fitur random atau *random feature selection*. *Random Forest* termasuk dalam teknik *ensemble* yang terbentuk dari banyak pohon keputusan untuk mengelompokkan data ke dalam kelas yang sesuai (Siti, 2022). *Random Forest* yakni algoritma klasifikasi yang efektif karena kemampuannya dalam mengatasi *overfitting* dan kemampuannya untuk bekerja dengan cepat, terutama pada dataset yang besar. Namun, *Random Forest* juga cenderung menghasilkan *overfitting* pada dataset yang tidak seimbang, ketika kelas mayoritas jauh lebih banyak daripada kelas minoritas (Rosta et al., 2023). Dataset yang dipergunakan pada penelitian ini terdapat masalah pada data, yakni data yang tidak seimbang (*data imbalance*).

Ketidakseimbangan data (*data imbalance*) terjadi ketika suatu kelas mempunyai total data yang lebih banyak dibanding pada kelas lain. Kelas dengan total data terbanyak disebut kelas mayoritas. Hal ini disebabkan pembelajaran klasifikasi cenderung memprediksi kelas data yang besar dibandingkan dengan kelas data yang berjumlah kecil (kelas minoritas) (Zhafirah, 2023). Data dari kelas minoritas sering kali mengalami kesalahan pelabelan saat diproses oleh model pengklasifikasi. Oleh karena itu, pengambilan sampel ulang (*resampling*) data memerlukan teknik tertentu, seperti *oversampling* dan *undersampling*. Salah satu cara dalam mengatasi masalah data tak seimbang menerapkan teknik *Synthetic Minority Over-Sampling Technique* (SMOTE). SMOTE yakni cara dalam mengatasi masalah ketidakseimbangan kelas dalam kumpulan data dengan menambahkan data sintetis ke kelas minoritas. Teknik ini tidak menduplikasikan

data yang ada, melainkan membuat data baru untuk mencapai keseimbangan dalam distribusi data. Sampel dipilih secara acak dari kelas minoritas, kemudian mengidentifikasi sejumlah tetangga terdekat untuk setiap sampel tersebut. Selanjutnya, data sintetis dibuat dengan menambahkan sampel-sampel ini ke dalam kumpulan data asli (Ridwan et al., 2024).

Dikan Ismafillah et al (2023) melakukan penelitian menggunakan kombinasi model *Random Forest* dan SMOTE memberikan akurasi yang lebih baik dibanding model *Decision Tree* dan SMOTE dalam memprediksi diagnosis diabetes. Penelitian ini menghasilkan akurasi sebesar 88,9%. Penelitian lain oleh Apriliah et al (2021) bertujuan untuk merancang model yang dapat memprediksi kemungkinan terjadinya diabetes pada pasien pada taraf ketelitian yang tinggi. Pada penelitian ini, berbagai metode seperti SVM, *Naïve Bayes*, *Random Forest* dipergunakan. Hasilnya memperlihatkan bahwasannya model *Random Forest* mencapai akurasi tertinggi dengan nilai 97,88%. Ghozali et al (2023) melaksanakan penelitian mengenai klasifikasi penyakit diabetes dan menemukan bahwasannya model *Random Forest* yang dikombinasikan dengan SMOTE, dengan pembagian data sebesar 80% untuk pelatihan dan 20% untuk pengujian, menghasilkan akurasi tertinggi sebesar 98%.

Beberapa penelitian sebelumnya hanya menitikberatkan pada satu aspek dari penerapan algoritma *Random Forest* dan SMOTE. Namun, penelitian ini melakukan perbandingan langsung antara penggunaan metode *Random Forest* tanpa SMOTE dan dengan SMOTE dalam konteks klasifikasi penyakit diabetes *mellitus*. Fokus utama dari penelitian ini yaitu mengevaluasi secara langsung efektivitas teknik SMOTE dalam meningkatkan kinerja model klasifikasi, khususnya dalam menangani ketidakseimbangan kelas yang sering ditemukan dalam dataset medis. Melalui perbandingan dan analisis kedua pendekatan tersebut, tujuan utama dari riset ini yaitu untuk memberikan wawasan yang lebih mendalam tentang manfaat dan potensi penggunaan SMOTE. Penelitian ini bertujuan untuk menentukan seberapa efektif SMOTE dapat meningkatkan akurasi serta kehandalan model klasifikasi dalam mendukung proses diagnosis dan manajemen penyakit diabetes *mellitus*. Selain itu, penelitian ini menggunakan pendekatan komprehensif dalam mengintegrasikan sembilan faktor variabel yang dipilih dalam klasifikasi

diabetes tipe 1 serta 2. Hasil dari penggunaan keseluruhan kombinasi faktor ini memberikan pemahaman yang dalam tentang kaitan antara aspek-aspek tersebut pada penyakit diabetes, yang bertujuan untuk memberikan pemahaman tentang faktor risiko, diagnosis, dan manajemen terhadap penyakit diabetes, serta memberikan landasan yang lebih kuat untuk intervensi medis yang lebih efektif dalam praktik klinis.

### **1.2 Rumusan Masalah**

1. Bagaimana proses penyusunan model klasifikasi penyakit diabetes *mellitus* menggunakan algoritma *Random Forest*?
2. Faktor-faktor apa saja yang berpengaruh terhadap hasil klasifikasi penyakit diabetes *mellitus* menggunakan algoritma *Random Forest*?
3. Bagaimana perbandingan performa klasifikasi *Random Forest* tanpa SMOTE dan dengan SMOTE?

### **1.3 Batasan Masalah**

1. Data yang digunakan yakni data pasien Diabetes mellitus yang didapatkan dari Puskesmas Bungah, Kabupaten Gresik, Provinsi Jawa Timur, dengan jumlah data 512 dan 9 variabel X “usia, Gula Darah Acak, Gula Darah Puasa, Gula Darah 2 Jam PP, HbA1c, BMI, Aktivitas Fisik, Sistolik Diastolik) 2 variabel Y (DM Tipe 1 dan DM Tipe 2”
2. Penelitian ini fokus pada klasifikasi Diabetes *mellitus* tipe 1 serta tipe 2 mempergunakan algoritma *Random Forest* dan dataset yang telah disediakan.
3. Evaluasi performa model klasifikasi hanya dilaksanakan dengan *confusion matrix*.

### **1.4 Tujuan Penelitian**

1. Menjelaskan proses penyusunan model klasifikasi penyakit diabetes *mellitus* menggunakan algoritma *Random Forest*.
2. Menentukan faktor-faktor yang berpengaruh terhadap hasil klasifikasi penyakit *diabetes mellitus* menggunakan algoritma *Random Forest*

3. Membandingkan kinerja model klasifikasi *Random Forest* tanpa menggunakan metode SMOTE dan dengan SMOTE dalam mengidentifikasi penyakit diabetes *mellitus*.

## 1.5 Manfaat Penelitian

### 1. Manfaat Teoritis

Penelitian ini nantinya bisa memberi pengaruh yang signifikan dalam pemahaman penyakit diabetes *mellitus* dengan menganalisis faktor-faktor yang berpengaruh terhadap hasil klasifikasi menggunakan algoritma *Random Forest*. Penelitian ini akan meningkatkan pengetahuan tentang algoritma *Random Forest* dalam konteks kesehatan dan membantu pengembangan teori terkait penggunaannya. Evaluasi terhadap penggunaan metode SMOTE dalam penelitian ini juga akan berkontribusi pada pengembangan teori dalam pemrosesan data kesehatan. Selain itu, penelitian ini akan memberikan wawasan tentang performa model klasifikasi SMOTE *Random Forest* dalam identifikasi penyakit diabetes *mellitus*, yang akan membantu dalam pemahaman teoritis tentang efektivitas teknik klasifikasi dalam analisis kesehatan.

### 2. Manfaat Praktis

#### a. Manfaat untuk penulis

Penulis akan mendapatkan pengalaman penelitian dalam menganalisis data kesehatan dan mengimplementasikan algoritma *machine learning*, memperluas pengetahuan dan keterampilan dalam bidang ilmu data kesehatan, serta memperkaya portofolio penelitian. Selain itu, penulis juga dapat memperoleh pengakuan dan apresiasi dari komunitas ilmiah atas kontribusinya dalam bidang penelitian kesehatan.

#### b. Manfaat untuk masyarakat

Penelitian ini dapat memberikan manfaat langsung bagi masyarakat, termasuk pasien, dokter, dan pengguna layanan kesehatan lainnya. Melalui analisis yang lebih baik tentang faktor-faktor yang mempengaruhi klasifikasi penyakit *diabetes mellitus*,

dokter dapat melakukan diagnosis lebih tepat dan memberikan perawatan yang lebih efektif kepada pasien diabetes. Selain itu, penelitian ini dapat membantu mengidentifikasi kasus-kasus diabetes yang mungkin terlewatkan atau kurang terdiagnosis secara tepat. Hal ini akan mengarah pada peningkatan manajemen penyakit dan kualitas hidup pasien diabetes, serta memberikan kontribusi pada penelitian dan pengembangan dalam bidang kesehatan secara lebih luas.