

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Diabetes adalah penyakit kronis yang prevalensinya terus meningkat secara global. Menurut International Diabetes Federation (IDF) [1], di tahun 2021, diperkirakan terdapat sekitar 537 juta orang dewasa yang hidup dengan mengidap penyakit ini, sementara 6,7 juta diantaranya terkonfirmasi meninggal. Prevalensi global ini telah bertambah 15,98% dibandingkan tahun 2019. Angka ini diproyeksikan akan terus mengalami peningkatan hingga mencapai 46% di tahun 2045 nanti, menjadi sekitar 784 juta orang penderita.

Indonesia sendiri menempati urutan ke-5 dunia [1], bahkan menjadi negara dengan urutan pertama di asia tenggara yang terdampak penyakit ini. Menyumbang sekitar 19.465.100 orang di tahun 2021, dengan kematian mencapai 236.711 orang. Penting untuk menyadari bahwa tingginya jumlah penderita diabetes di Indonesia saat ini merupakan isu yang serius dan menunjukkan adanya masalah kesehatan yang signifikan.

Diabetes mellitus terjadi ketika tubuh kita tidak dapat menggunakan insulin dengan efektif atau tidak memiliki cukup insulin untuk mengubah glukosa menjadi energi. Insulin merupakan hormon yang dihasilkan oleh sel-sel beta pankreas, yang memiliki peran dalam mengatur kadar gula (*glukosa*) dalam darah. Ketika terjadi gangguan, maka glukosa ini tidak dapat sepenuhnya diserap menjadi sumber energi, sehingga terjadilah penumpukan glukosa. Diabetes yang tidak terkontrol dengan baik dapat menyebabkan komplikasi jangka panjang yang serius, seperti penyakit stroke [2][3], gagal ginjal [4], kerusakan saraf [5], gangguan penglihatan [6], luka yang sulit sembuh, bahkan sampai amputasi [7].

Diabetes memiliki dampak yang signifikan pada kesehatan dan kualitas hidup individu, serta menimbulkan beban ekonomi yang besar pada sistem perawatan kesehatan. Berbagai upaya pencegahan dan penanganan diabetes telah banyak dilakukan, seperti: Edukasi masyarakat tentang gaya hidup sehat, promosi pola

makan seimbang, meningkatkan kesadaran akan pentingnya aktivitas fisik, serta akses dan penggunaan obat-obatan anti-diabetes dan juga perawatan yang tepat.

Di sisi lain, perkembangan teknologi informasi telah membuka peluang baru dalam pemanfaatan data dengan skala yang besar (*big data*), untuk memahami dan mengelola penyakit secara lebih efektif. Salah satu pendekatan yang digunakan adalah *data mining* atau penambangan data. *Data mining* adalah proses ekstraksi pengetahuan atau informasi dari sekumpulan data dengan jumlah besar dengan menggunakan metode atau algoritma tertentu. Dalam konteks diabetes, *data mining* dapat digunakan untuk menganalisis data klinis, laboratorium, genetik, dan faktor risiko lainnya untuk mengidentifikasi pola atau informasi yang dapat digunakan dalam pengelolaan diabetes.

Tahun 2020 Raghavendra S, dkk mengkomparasi lima algoritma data mining: Naïve Bayes, k-NN, Logistic Regression, Neural Network, dan Random Forest untuk memprediksi diabetes [8]. Dalam penelitian ini mereka menggunakan *dataset* Pima Indians¹. *Dataset* yang memiliki 768 *record* dan 9 atribut. Mereka bahkan melakukan pengujian untuk meningkatkan akurasi menggunakan metode Forward Selection and Backward Elimination. Sementara itu dengan *dataset* yang sama penelitian lain juga dilakukan [9][10][11][12][13], dengan menerapkan berbagai algoritma klasifikasi. Namun lain halnya penelitian yang dilakukan oleh Widya A dkk [14], mereka menguji *dataset public* yang berbeda. *Dataset* ini berasal dari Rumah Sakit Diabetes Sylhet di Bangladesh, yang berisi 520 *record* dan 17 atribut. Kemudian peneliti lain juga menerapkan *dataset* ini [15][16][17]. Meski ada juga, peneliti lain yang menguji *dataset* yang berbeda [18][19], selain dari semua yang disebutkan ini, untuk mendapatkan hasil prediksi diabetes yang variatif.

Dari semua penelitian tersebut, penggunaan *dataset* yang berbeda-beda bisa memberikan hasil yang berbeda, meski menggunakan algoritma *data mining* yang sama [20][21][22][23]. Hal ini dikarenakan karakteristik dan pola-pola data yang terdapat dalam setiap *dataset* bisa mempengaruhi hasil analisis dan prediksi yang diperoleh. Dalam penelitian ini penulis *juga* menggunakan *dataset* yang berbeda

¹ *Dataset* ini berasal dari Institut Nasional Penyakit Diabetes, Pencernaan, dan Ginjal; di Maryland, negara bagian Amerika Serikat.

dari yang telah umum digunakan ini: *Dataset* ini masih sangat jarang digunakan². *Dataset* ini penulis dapatkan dari situs kaggle [24]. Dengan demikian *dataset* ini juga merupakan *dataset public*.

Dengan mengkomparasi tiga algoritma: K-Nearest Neighbor, Random Forest, dan Deep Learning, ketiga algoritma ini telah menunjukkan tingkat akurasi yang paling baik diantara algoritma klasifikasi yang pernah diteliti sebelumnya [16][25]. Kemudian mengoptimasinya menggunakan Particle Swarm Optimization untuk meningkatkan akurasi, kita akan melihat apakah ketiga algoritma ini juga akan memberikan hasil yang baik dengan *dataset* yang penulis gunakan di sini. Penulis berharap melalui penelitian ini dapat memberikan pemahaman yang komprehensif tentang faktor risiko diabetes dan memungkinkan langkah-langkah pencegahan dan intervensi yang lebih efektif lagi.

1.2 Perumusan Masalah

Berikut adalah beberapa rumusan masalah dalam penelitian ini:

- Bagaimana prevalensi diabetes mellitus secara global dan di Indonesia, serta proyeksi perkembangan jumlah penderita di masa depan?
- Apa saja komplikasi jangka panjang yang bisa terjadi bila diabetes mellitus ini yang tidak terkontrol dengan baik? Bagaimana upaya-upaya pencegahan dan penanganannya?
- Bagaimana penelitian sebelumnya telah menggunakan *data mining* dan algoritma klasifikasi untuk memprediksi diabetes mellitus?
- Mengapa karakteristik dan pola data dalam *dataset* yang berbeda dapat mempengaruhi hasil analisis dan prediksi dalam penelitian *data mining*?
- Mengapa penulis memilih *dataset* yang jarang digunakan dan apa tujuan dari penelitian ini?
- Apa algoritma-algoritma yang akan dibandingkan dalam penelitian ini dan mengapa dipilih?

² Penulis telah melakukan *review* jurnal-jurnal terkait *dataset* diabetes yang umum digunakan dalam berbagai penelitian (disajikan dalam Tabel 2.4). Dalam rentang bulan april hingga juli tahun 2023, penulis sedikit sekali menemukan penelitian yang menggunakan *dataset* BIT Mesra ini.

- Bagaimana penggunaan Particle Swarm Optimization dapat meningkatkan akurasi algoritma-algoritma yang dipilih?

Rumusan masalah ini akan menjadi dasar penelitian penulis untuk menyelidiki dan memahami faktor-faktor risiko diabetes serta mengembangkan langkah-langkah pencegahan dan intervensi yang lebih efektif dengan menggunakan *data mining* dan algoritma klasifikasi pada *dataset public* diabetes.

1.3 Batasan Masalah

Penelitian ini akan difokuskan pada penggunaan algoritma klasifikasi *data mining* untuk menganalisis data terkait diabetes. Menggunakan tiga algoritma (k-Nearest Neighbor, Random Forest, dan Deep Learning) yang dioptimisasi dengan Particle Swarm Optimization untuk meningkatkan akurasi, dan *dataset public* yang penulis peroleh dari situs Kaggle [24] – memiliki 952 *record* dan 18 atribut. Penelitian ini diharapkan memberikan pemahaman yang lebih komprehensif tentang faktor risiko diabetes dan dapat membantu dalam langkah-langkah pencegahan dan intervensi yang lebih efektif terkait diabetes mellitus.

1.4 Tujuan Penelitian

Tujuan dilakukannya penelitian ini adalah untuk menerapkan teknik *data mining* dengan cara: Mengoptimasi tiga jenis algoritma klasifikasi (K-Nearest Neighbor, Random Forest, dan Deep Learning) menggunakan Particle Swarm Optimization untuk mencari algoritma yang dapat memberikan tingkat akurasi terbaik. Tujuan utamanya adalah untuk mengidentifikasi pola atau hubungan yang dapat digunakan untuk memprediksi dan menganalisis risiko penyakit diabetes, serta membantu dalam pengambilan keputusan terkait manajemen penyakit ini.

1.5 Manfaat Penelitian

Manfaat teoritis merupakan manfaat jangka panjang dalam pengembangan teori pembelajaran, sedangkan manfaat praktis memberikan dampak secara langsung terhadap komponen-komponen pembelajaran. Manfaat teoritis dari penelitian ini adalah:

- Pemahaman Dasar (*Basics*) tentang Diabetes: Memberikan pemahaman yang komprehensif tentang diabetes mellitus, termasuk definisi, penyebab, dampaknya pada kesehatan individu, serta upaya-upaya pencegahan dan penanganannya. Informasi ini dapat membantu para peneliti, ahli kesehatan, dan ilmuwan medis untuk memahami lebih baik tentang penyakit ini dan mengembangkan penelitian lebih lanjut.
- Pemanfaatan *Data Mining* dalam Pengelolaan Diabetes: Menggambarkan pemanfaatan *data mining* atau penambangan data dalam mengelola diabetes mellitus. Pendekatan ini memungkinkan analisis data klinis, laboratorium, genetik, dan juga faktor risiko lainnya untuk mengidentifikasi pola atau informasi yang relevan dalam pengelolaan diabetes mellitus. Manfaat ini memberikan wawasan baru tentang cara menggunakan teknologi informasi dalam pemahaman dan penanganan penyakit kronis.

Sementara manfaat praktisnya:

- Penelitian Lebih Lanjut dalam *Data Mining* dan Diabetes: Memberikan gambaran tentang penelitian sebelumnya yang menggunakan *data mining* dan algoritma klasifikasi untuk prediksi diabetes. Manfaat ini mendorong penelitian lebih lanjut dalam bidang ini, dengan menggali potensi *dataset* yang berbeda dan mencari algoritma yang lebih baik dalam memprediksi diabetes mellitus.
- Pengembangan Teknologi dalam Bidang Kesehatan: Menyoroti pentingnya perkembangan teknologi informasi dan pemanfaatan data skala besar (*big data*) dalam pengelolaan penyakit diabetes. Manfaat ini dapat mendorong pengembangan lebih lanjut dalam teknologi kesehatan untuk mendukung pengelolaan diabetes mellitus secara lebih efektif.

Melalui manfaat teoritis dan praktis ini, diharapkan dapat terjadi peningkatan kesadaran masyarakat, penelitian yang lebih mendalam, dan juga perkembangan teknologi kesehatan yang lebih baik dalam menghadapi masalah diabetes mellitus yang serius ini.

1.6 Sistematika Penulisan

Laporan penelitian tesis ini terdiri dari 5 bab dengan sistematika penulisan sebagai berikut:

- Bab 1 – Pendahuluan
Berisi penjelasan mengenai latar belakang penelitian ini, konteks masalah, dan rasionalitas pemilihan topik. Menjelaskan tujuan dari penelitian serta menguraikan kontribusi dan manfaat yang mungkin diperoleh dari hasil penelitian.
- Bab 2 – Landasan Teori
Rangkuman literatur terkait dengan topik penelitian untuk menunjukkan pemahaman peneliti terhadap kerangka konseptual dan teoritis.
- Bab 3 – Metode Penelitian
Rincian langkah-langkah dan metode yang akan diterapkan. Dimulai dari mempersiapkan *dataset*, memastikan data yang akan digunakan telah siap dan *clean*, sampai algoritma dan teknik optimasi yang digunakan.
- Bab 4 – Hasil Penelitian dan Pembahasan
Menyajikan interpretasi dan analisis hasil yang diperoleh dari penelitian sesuai dengan tujuan penelitian. Di sini kita akan melihat perbandingan akurasi, presisi, *recall*, dan visual grafik ROC-AUC dari ketiga algoritma, baik itu sebelum dan sesudah dioptimalkan.
- Bab 5 – Kesimpulan dan Saran
Simpulan dari hasil penelitian yang merangkum temuan utama, dampak hasil penelitian terhadap teori atau praktik, batasan-batasan penelitian, dan juga saran-saran untuk penelitian lanjutan atau implementasi praktis.