

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Diabetes merupakan sebuah penyakit kronis yang disebabkan pada saat pankreas tidak dapat memproduksi cukup insulin secara efektif. Insulin merupakan sebuah hormon yang mengatur glukosa darah. Terdapat tiga tipe diabetes yang ada yaitu, diabetes tipe 1 merupakan diabetes yang terjadi karena tubuh hanya memproduksi sedikit insulin, terjadi pada remaja, atau anak-anak. Lalu ada diabetes tipe 2 yang ditandai dengan tubuh tidak merespons insulin secara maksimal, yang disebabkan karena pola makan yang buruk, gaya hidup yang kurang gerak (berolahraga) dan tingkat obesitas. Lalu ada diabetes *mellitus gestasional* (GDM) yang terjadi karena gula darah tinggi selama kehamilan.

Menurut *International Diabetes Federation*, pada tahun 2021 terdapat 537 juta orang dewasa dengan umur sekitar 20 sampai 79 tahun yang terkena atau mempunyai penyakit diabetes [1]. Pada Diabetes Atlas Edisi ke-10 tingkat perempuan terkena diabetes lebih tinggi daripada pria yang disebabkan karena faktor perbedaan hormon, penambahan berat badan dan gaya hidup selama masa pubertas [2]. Aspek lain seperti, resistensi insulin, ketika hamil tingkat resistensi insulin pada perempuan cenderung naik [3]. Perempuan penderita diabetes memiliki risiko lebih tinggi untuk keguguran, dan melahirkan bayi dengan cacat lahir [4]. Setelah menopause, penurunan hormon estrogen pada perempuan dapat menyebabkan peningkatan resistensi insulin dan gangguan metabolisme glukosa, sehingga meningkatkan risiko diabetes tipe 2. Selain itu, perempuan dengan diabetes tipe 1 dan tipe 2 biasanya memiliki indeks massa tubuh (BMI) yang lebih tinggi, dan obesitas lebih sering ditemukan pada perempuan daripada laki-laki, yang merupakan faktor risiko utama untuk diabetes tipe 2 [5].

Dalam penanganan diabetes, identifikasi dini memiliki peran penting untuk memberikan tindakan yang sesuai dan mencegah kemungkinan komplikasi yang lebih serius. Salah satu pendekatan yang dapat digunakan untuk deteksi dini diabetes adalah dengan melakukan prediksi risiko diabetes pada individu berdasarkan faktor-faktor tertentu, seperti contohnya kadar glukosa, indeks massa tubuh (BMI), usia [6]. Identifikasi dini diabetes dapat memberikan waktu yang lebih banyak bagi pasien untuk memulai untuk melakukan perubahan gaya hidup.

Meskipun tidak ada pengobatan jangka panjang untuk diabetes, diagnosis dini dapat membuat perbedaan yang signifikan dalam penanganan penyakit ini. Ketika diabetes terdeteksi cukup dini, pengobatan dapat segera dimulai, membuka peluang untuk pencegahan komplikasi dan meningkatkan kemungkinan pemulihan yang lebih baik menurut para dokter dalam dunia kedokteran [7].

Deteksi diabetes penting karena terdapat perubahan hormon seiring bertumbuhnya manusia yang memengaruhi risiko terjadinya diabetes. Dalam populasi wanita secara umum, peningkatan kadar testosteron dan penurunan globulin pengikat hormon seks menjadi faktor risiko untuk resistensi insulin dan perkembangan diabetes tipe 2 [8]. Menopause meningkatkan risiko diabetes pada perempuan karena penurunan kadar estrogen yang dapat menyebabkan resistensi insulin dan gangguan metabolisme glukosa [9]. Perempuan dengan riwayat GDM memiliki risiko lebih tinggi untuk mengembangkan diabetes tipe 2 di kemudian hari, sehingga deteksi dini berperan penting dalam upaya pencegahan diabetes jangka panjang [10].

Sebagai alat bantu awal deteksi dini sebelum dilakukan pemeriksaan diabetes oleh dokter, sistem deteksi dini berbasis *machine learning* dapat menjadi alat yang sangat berguna. Pemanfaatan *machine learning* menjadi salah satu solusi untuk melakukan prediksi secara akurat. *Machine learning* merupakan pengembangan dari kecerdasan buatan yang dapat melakukan pembelajaran secara mandiri tanpa harus diprogram berulang kali oleh manusia [11]. *Machine Learning* diperlukan terutama dalam bidang kesehatan karena kemampuannya untuk memprediksi suatu penyakit dan mengurangi kesalahan pengambilan keputusan manusia [12]. Klasifikasi diabetes yang benar dapat membantu dokter untuk melakukan perencanaan perawatan yang tepat, *machine learning* memperoleh hasil yang menjanjikan dalam diagnosis diabetes [13].

Algoritma *Logistic Regression*, sebagai salah satu algoritma dalam *machine learning*, dapat digunakan untuk memprediksi risiko penyakit kronis yang umum [14]. Algoritma *Logistic Regression* merupakan suatu algoritma yang dapat digunakan jika variabel dependen (target) merupakan variabel dikotomi (dua kategori) [15]. *Logistic Regression* adalah sebuah algoritma yang digunakan untuk mencari sebuah peluang atau probabilitas dari variabel dependen (target) yang dipengaruhi oleh variabel independen (prediktor) [16]. Probabilitas yang telah diprediksi oleh *Logistic Regression* dapat memberikan hasil yaitu nilai 0 (misalnya, "tidak") atau 1 (misalnya, "ya") [17].

Algoritma *Logistic Regression* memiliki kemampuan untuk menangani

masalah klasifikasi biner, seperti adanya atau tidak adanya diabetes. Algoritma ini efisien dalam menangani *dataset* dengan variabel independen yang bersifat kontinu maupun kategori. *Logistic Regression* mampu memberikan probabilitas prediksi yang dapat membantu dalam penilaian risiko pasien, yang mendukung upaya deteksi dini [6].

Penelitian terdahulu dengan menggunakan algoritma *Logistic Regression* dengan metode *Feature Selection*, *Max Voting*, dan *Stacking* serta menggunakan dua *dataset*, menghasilkan nilai akurasi tertinggi dengan metode *Max Voting* yaitu 77.83% pada *dataset* satu dan 93.41% pada *dataset* dua [18]. Penelitian terdahulu yang menggunakan algoritma *Support Vector Machine* dan *Naive Bayes* menunjukkan bahwa algoritma *Support Vector Machine* mendapatkan nilai akurasi 78.04%, sedangkan dengan menggunakan *Naive Bayes* menghasilkan nilai akurasi sebesar 76.98% [19]. Lalu penelitian menggunakan algoritma *Random Forest* menghasilkan akurasi sebesar 83.8% dengan seluruh atribut dengan metode *forward selection* dan 84.1% dengan menggunakan tujuh atribut dengan metode *backward elimination* [20]. Penelitian lain menggunakan algoritma tipe *Supervised Learning* dengan menggunakan metode *cross validation* dengan nilai  $k = 10$  menghasilkan nilai akurasi terbaik dengan algoritma *Logistic Regression* sebesar 76% [21]. Pada penelitian menggunakan algoritma *Naive Bayes*, *Random Forest*, dan *Decision Tree* dengan metode *k-means clustering* dengan nilai  $k = 2, 3, \text{ dan } 4$  diperoleh akurasi terbaik dengan algoritma *Random Forest* sebesar 79.57% menggunakan semua fitur pada *dataset*, lalu dengan lima fitur didapatkan akurasi terbaik sebesar 79.13% dengan algoritma *Naive Bayes*, dan dengan menggunakan tiga fitur didapatkan akurasi terbaik sebesar 77.83% dengan algoritma *Naive Bayes* [22].

Penelitian ini bertujuan untuk membuat suatu klasifikasi menggunakan algoritma *Logistic Regression* dengan tujuan membantu melakukan klasifikasi pasien yang terkena penyakit diabetes berdasarkan variabel yang terdapat pada data. Dengan menggunakan data pasien yang sudah terkumpul, diharapkan dengan menggunakan algoritma *Logistic Regression* dapat memberikan hasil yang akurat dalam mengidentifikasi pasien yang terkena diabetes.

## 1.2 Rumusan Masalah

Berdasarkan pada penjelasan latar belakang yang telah dibuat, berikut merupakan rumusan masalah pada penelitian ini, yaitu:

1. Bagaimana cara mengimplementasikan algoritma *Logistic Regression* untuk

melakukan klasifikasi pasien perempuan yang terkena penyakit diabetes?

2. Berapa nilai akurasi, *precision*, *recall*, dan *f1-score* pada algoritma *Logistic Regression* dalam melakukan klasifikasi pasien perempuan yang terkena penyakit diabetes?

### 1.3 Batasan Permasalahan

Berdasarkan penelitian yang akan dilakukan terdapat batasan masalah, yaitu:

- Data pertama berasal dari Kaggle yang berjudul "Pima Indians Diabetes Database" yang terdiri dari 768 data dan berasal dari tahun 2017 [23].
- Data kedua berasal dari IEEE Dataport yang berjudul "TYPE 2 DIABETES DATASET" yang terdiri dari 2000 data dan berasal dari tahun 2024 [24].
- Data ketiga berasal dari gabungan antara data satu dan dua.

### 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijelaskan, adapun tujuan penelitian ini, yaitu:

1. Penelitian ini bertujuan untuk melakukan implementasi algoritma *Logistic Regression* untuk melakukan klasifikasi penyakit diabetes terhadap pasien perempuan.
2. Penelitian ini bertujuan untuk menghitung nilai akurasi, *precision*, *recall*, dan *f1-score* algoritma *Logistic Regression* dalam melakukan klasifikasi penyakit diabetes terhadap pasien perempuan.

### 1.5 Manfaat Penelitian

Berikut merupakan manfaat penelitian ini, yaitu:

1. Dapat digunakan sebagai referensi pada penelitian selanjutnya terutama dalam penelitian algoritma *Logistic Regression* dalam melakukan klasifikasi pasien terhadap penyakit diabetes atau pada penyakit lainnya.

2. Pada bidang *machine learning*, sistem yang dibuat dalam penelitian ini dapat digunakan sebagai referensi untuk klasifikasi penyakit diabetes.
3. Sistem yang dibuat dapat digunakan sebagai alat bantu deteksi awal penyakit diabetes, sehingga mempermudah identifikasi dini untuk mencegah masalah kesehatan yang lebih serius.

## 1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut.

- Bab 1 PENDAHULUAN  
Pada bab 1 menjelaskan mengenai latar belakang dari permasalahan yang akan diteliti, rumusan masalah dari yang diteliti, batasan masalah, tujuan penelitian, dan manfaat penelitian.
- Bab 2 LANDASAN TEORI  
Pada bab 2 menjelaskan mengenai teori-teori yang mendukung penelitian ini. Teori tersebut berupa mengenai diabetes, *machine learning*, algoritma *Logistic Regression*, dan teori dari metode yang dipakai.
- Bab 3 METODOLOGI PENELITIAN  
Pada bab 3 menjelaskan mengenai tahap-tahap yang dilakukan dalam penelitian. Tahapan atau metode yang dipakai dalam penelitian berupa pengumpulan data, *preprocessing*, pembagian data menjadi data latih dan uji, implementasi algoritma, validasi, dan evaluasi.
- Bab 4 HASIL DAN DISKUSI  
Pada bab 4 menjelaskan mengenai hasil penelitian yang dilakukan dengan implementasi algoritma *Logistic Regression*.
- Bab 5 KESIMPULAN DAN SARAN  
Pada bab 5 menjelaskan mengenai hasil kesimpulan yang diperoleh dari hasil penelitian dan saran untuk penelitian selanjutnya. Kesimpulan ini juga menjawab dari rumusan masalah dan tujuan penelitian dilakukan penelitian ini.