

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Menerapkan pola hidup sehat akan sangat amat berpengaruh kepada kesehatan diri sendiri, Dikarenakan Menurut *World Health Organization*(WHO) Perilaku hidup yang tidak sehat dan minimnya aktivitas fisik aalah satu pemicu utama dalam risiko Kesehatan global. Sedangkan menurut data Riset Kesehatan Dasar (Risksdas) tahun 2018, tingkat konsumsi makanan manis (87.9%) lalu minuman manis (91.49%) untuk indonesia negara berkembang sangatlah tinggi [1].

Sebagian orang tidak mengetahui atau tidak merasakan bahwa mereka terjangkit penyakit diabetes karena gejalanya mirip dengan penyakit lain. Hal itu dikarenakan minimnya Pendidikan karenan di indonesia mayoritas hanya lulusan SMP atau bahkan SD. Penyakit ini bahkan dapat menyebabkan komplikasi. Dokter harus melakukan tes darah untuk mendiagnosis diabetes. Kerap sekali membuang air kencing, terlalu cepat merasakan kelaparan dan kehausan, menurunnya kondisi tubuh yang signifikan, kulit kering, penyembuhan luka yang selalu memakan waktu yang lama, dan bermaslah dalam penglihatan adalah beberapa gejala umum diabetes.

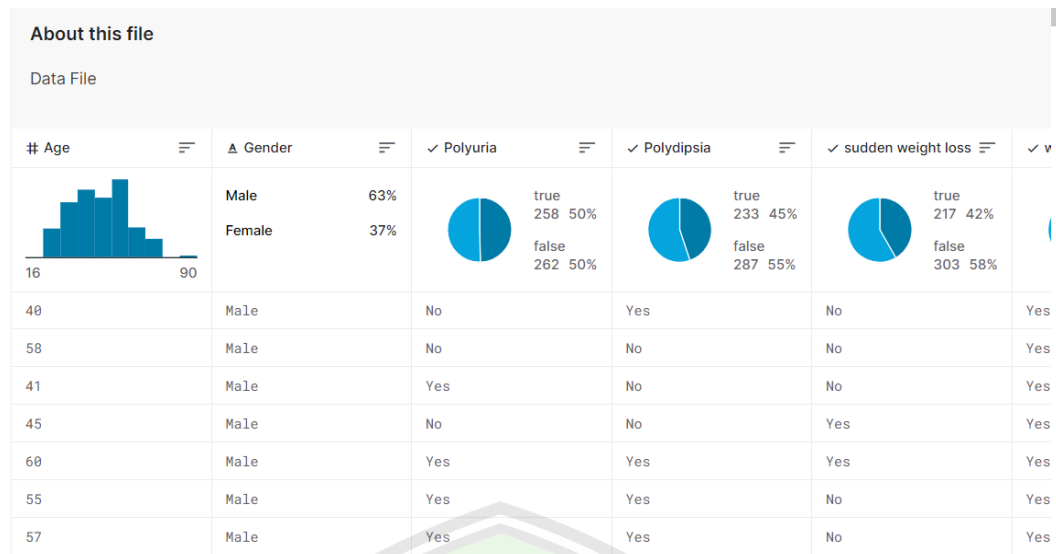
Indonesia sedang mengalami transisi epidemiologi, dengan penurunan prevalensi penyakit menular tetapi peningkatan prevalensi penyakit tidak menular (PTM) atau penyakit degeneratif. Jumlah penderita Diabetes Melitus meningkat secara signifikan di seluruh dunia, dengan diperkirakan 3,4 juta orang meninggal

akibat tingginya kadar gula darah puasa. Di negara-negara berpenghasilan rendah dan sedang, tingkat mortalitas akibat Diabetes Melitus mencapai 80% [2].

Faktor utama yang menyebabkan penyakit diabetes adalah gaya hidup atau keadaan yang tidak seimbang, tidak tertatanya pola istirahat dan makan yang tidak beraturan, dan pembagian waktu saat istirahat dan olahraga yang sangat buruk. Kemudian juga ada faktor lain, seperti tubuh tidak memiliki imunitas yang buruk sehingga tidak dapat menggunakan beraktivitas dengan baik. Studi lain menemukan bahwa penyebab risiko yang biasa terjadi pada penyakit tidak menular pada remaja adalah terlalu banyaknya mengonsumsi alkohol, tembakau, dan vape, minimnya olahraga, dan diet dengan cara yang salah sehingga menurunkan berat badan yang sangat drastis.[3].

Akibatnya, sangat dibutuhkannya teknologi yang mampu mengidentifikasi diabetes secara dini dengan tingkat analisis yang *on point* atau akurat. Dengan demikian, penyakit diabetes dapat segera ditindak lanjuti agar tidak memperbanyak penderitanya bahkan hingga Tingkat kematian yang terkait [4].

Mengingat semakin banyaknya korban yang terpapar penyakit diabetes, oleh karena itu, untuk mengatasi permasalahan di atas saya melakukan penelitian menggunakan dataset yang dapat diakses secara publik dari *Website Kaggle* dengan judul dataset *Early Stage Diabetes Risk Prediction Dataset* sebagai berikut.



Gambar 1. 1 Dataset Penelitian

Sumber *Kaggle* (2023)

Hal ini terbukti berdasarkan grafik diatas Gambar 1.1 dataset diatas menunjukkan pengidap penyakit diabetes dalam 3 tahun yang lalu pada tahun 2020 pengidap penyakit diabetes mencapai 520 Kasus dari dataset yang telah diambil dari *Kaggle*. *Kaggle* merupakan kumpulan data standar yang coba di modelkan ratusan atau ribuan individu atau tim.

Bisa disimpulkan untuk penyakit diabetes mayoritas berjenis kelamin laki-laki dalam rentang usia 40 tahun hingga 60 tahun. Kemudian daripada itu mayoritas berjenis kelamin laki-laki hingga menyetuh di angka 63% sedangkan untuk berjenis kelamin Perempuan sebesar 37% dari data survey yang saya ambil dari *Kaggle*. Mengenai analisis kinerja teknik klasifikasi data mining untuk prediksi diabetes.

Untuk mendeteksi diabetes sejak dini, dapat digunakan berbagai pendekatan untuk memprediksi penyakit. Data mining dengan prinsip klasifikasi adalah salah satu metode yang dapat digunakan.

Sebelumnya sudah dilakukan beberapa penelitian untuk memprediksi nilai akurasi tahap awal resiko penyakit diabetes menggunakan beberapa algoritma Dwi

Yuni Utami beserta rekan-rekannya menyatakan bahwa Algoritma Regresi Logistik memiliki intensitas atau akurasi tertinggi dalam meraih 75.78% dibandingkan dengan Naive Bayes dan Neural Network, masing-masing dengan 74.87% dan 69.27% [4].

Studi Penelitian yang dilakukan oleh Diniyal Amru Agatsa beserta kerabatnya, klasifikasi pengidap diabetes menggunakan metode Support Vector Machine didapatkan hasil akurasi sebesar 77.92% [5]. Karena strukturnya sederhana dan kinerjanya yang luar biasa, Algoritma *Random Forest* telah umumnya digunakan untuk melaksanakan klasifikasi dan regresi. Penelitian ini mengeksplorasi penggunaan metode *Random Forest* dalam memprediksi tahap awal risiko penyakit diabetes.

Dengan mempertimbangkan konteks permasalahan di atas, peneliti memutuskan untuk menggunakan judul “Klasifikasi Tahap Awal Risiko Penyakit Diabetes Menggunakan Algoritma *RandomForest*”.

## 1.2 Identifikasi Masalah

Berdasarkan masalah yang telah dirangkum dari penelitian di atas maka identifikasi masalahnya sebagai berikut :

1. minimnya pengetahuan masyarakat mengenai penyakit diabetes .
2. sangat dibutuhkannya alat yang bisa mengidentifikasi penyakit diabetes tahap awal dengan nilai akurasi yang tinggi dan sangat akurat.

### 1.3 Rumusan Masalah

1. Berapa nilai akurasi klasifikasi tahap awal risiko penyakit Diabetes menggunakan algoritma *Random Forest*?
2. Informasi apa yang diperoleh dari klasifikasi tahap awal risiko penyakit diabetes berdasarkan hasil klasifikasi yang telah didapat?

### 1.4 Batasan Masalah

1. Sumber data yang akan diklasifikasikan menggunakan data Tahap awal risiko penyakit diabetes yang ada di situs *Kaggle*. Total data dari 520 penulis mengambil *sample* untuk perhitungan dengan 7 *Data Training* dan 3 data *testing*.
2. Setelah itu nilai akurasi diukur menggunakan presisi, akurasi, *F1 score* dan yang terakhir *Recall*.

### 1.5 Tujuan Dan Manfaat

#### 1.5.1 Tujuan

Berdasarkan permasalahan diatas tujuan penelitian tugas akhir ini yaitu:

1. menjelaskan prediksi kemungkinan penyakit diabetes secara dini
2. mengklasifikasi penyakit diabetes tahap awal menggunakan algoritma *Random Forest*.

#### 1.5.2 Manfaat

Untuk Masyarakat banyak dapat memberitahukan informasi prediksi tahap awal risiko penyakit diabetes sehingga mudah untuk dilakukan pengecekan

terhadap penyakit diabetes dalam kehidupan masyarakat luas dan menerapkan pola hidup sehat.

## 1.6 Sistematika Penulisan

Penjelasan ini tentang subjek atau elemen dalam topik skripsi ini terbentuk dari lima bab, sebagai berikut inti dari setiap babnya:

### **BAB I PENGANTAR**

Bagian ini menjelaskan tentang pengantar yang meliputi latar belakang, identifikasi masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan dalam menyelesaikan tugas akhir berjudul Klasifikasi Tahap Awal Risiko Penyakit Diabetes Menggunakan *Random Forest*.

### **Bab II LANDASAN TEORI**

Bagian ini merangkum tentang tinjauan pustaka yang relevan dengan penelitian, pembahasan mengenai dasar-dasar teori penelitian, serta teori-teori lain yang terkait dengan topik untuk dijadikan referensi dalam studi dan pelaksanaan penelitian studi

### **BAB III METODE PENELITIAN**

Bagian ini akan membahas tahapan urutan perencanaan penelitian, subjek penelitian, dan metode yang akan digunakan dalam penelitian ini. Pada bagian ini, dijelaskan melalui bagaimana proses penelitian dijalani dan mengakuisisi data didapat serta diolah menggunakan metode yang telah diaplikasikan sebelumnya.

#### **BAB IV HASIL DAN PEMBAHASAN**

Bagian ini membicarakan tentang pengumpulan dan pengolahan data, yang mencakup beberapa sub bagian seperti pengumpulan data dan pengolahan data. Dari bagian ini, pengumpulan data dilakukan dengan tujuan untuk mendapatkan data yang sesuai dengan metode penelitian yang telah ditetapkan.

#### **BAB V PENUTUP**

Bagian ini menyajikan hasil penelitian dan rekomendasi yang diperoleh dari penelitian yang dilakukan oleh peneliti.

