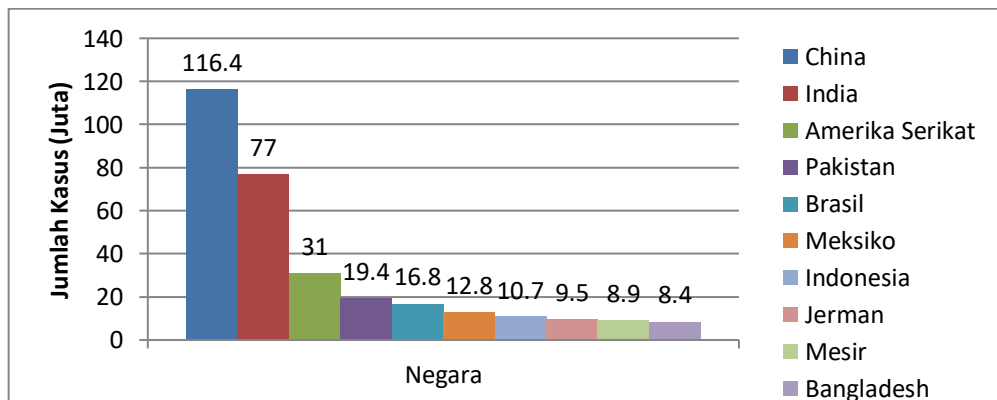


# BAB I PENDAHULUAN

## 1.1 Latar Belakang Masalah

Diabetes adalah penyakit kronis yang ditandai dengan tingginya kadar gula (glukosa) darah. Glukosa merupakan salah satu sumber energi yang dibutuhkan bagi sel-sel di dalam tubuh manusia. Kelebihan glukosa mengakibatkan menumpuknya kadar gula di dalam darah oleh karena itu glukosa yang berlebihan tidak bisa dikelola dengan baik yang mengakibatkan sel-sel tubuh mengalami kerusakan organ. Jika diabetes tidak dikontrol dengan baik, dapat timbul berbagai komplikasi yang membahayakan nyawa penderita [1].

Menurut data dari International Diabetes Federation (IDF) memperkirakan sedikitnya terdapat 463 juta orang pada usia 20-79 tahun di dunia menderita diabetes pada tahun 2019 atau setara dengan angka prevalensi sebesar 9.3% dari total penduduk pada usia yang sama. Berdasarkan jenis kelamin, IDF memperkirakan prevalensi diabetes di tahun 2019 yaitu 9% pada perempuan dan 9,65% pada laki-laki. Prevalensi diabetes diperkirakan meningkat seiring penambahan umur penduduk menjadi 19,9% atau 111,2 juta orang pada umur 65-79 tahun. angka diprediksi terus meningkat hingga mencapai 578 juta di tahun 2030 dan 700 juta di tahun 2045. Berikut 10 negara dengan jumlah penderita diabetes (juta) tertinggi tahun 2019 [2]:



Sumber : Kemenkes, 2020

China menjadi negara dengan jumlah orang dewasa penderita diabetes terbesar di dunia yaitu 116,4 juta penduduk, India menjadi negara dengan penderita diabetes tertinggi kedua di dunia yaitu sebanyak 77 juta penduduk, di posisi ketiga ada Amerika Serikat dengan jumlah penderita diabetes sebanyak 31 juta penduduk, Posisi selanjutnya ada Pakistan 19.4 juta kasus diabetes, Posisi kelima ada Brasil dengan jumlah pengidap diabetes sebanyak 16,8 juta penduduk. Meksiko menduduki posisi ke enam dengan jumlah penderita diabetes 12.8 juta penduduk, posisi ke tujuh ada Indonesia dengan jumlah penderita diabetes 10.7 juta penduduk, selanjutnya Jerman dengan jumlah penderita diabetes sebanyak 9.5 juta penduduk, Mesir ada di posisi ke sembilan dengan jumlah penderita diabetes sebanyak 8.9 juta penduduk, dan yang terakhir Bangladesh ada diperingkat 10 dengan jumlah penderita diabetes sebanyak 8.4 juta penduduk [2].

Penyebab meningkatnya jumlah penderita diabetes dikarenakan diabetes dikenal sebagai *silent killer*. Hal ini membuat penyandanginya banyak yang tidak menyadari bahwa dirinya terserang penyakit diabetes. Biasanya penderita diketahui terjangkit penyakit ini ketika sudah terjadi komplikasi [3].

Di era digital saat ini, data mining dapat diimplementasikan di bidang kesehatan. Salah satu penerapannya yang paling dikenal pada saat ini adalah penggunaan data mining untuk memprediksi penyakit, salah satunya yaitu penyakit diabetes. Prediksi terhadap penderita diabetes dapat diperoleh melalui kumpulan beberapa data pasien penderita diabetes yang tersimpan dalam basis data, kemudian diolah dengan suatu pola tertentu sehingga hasilnya dapat digunakan untuk diagnosa awal diabetes[4]. Ada beberapa penelitian terkait mengenai prediksi penyakit diabetes dengan menggunakan metode data mining diantaranya adalah jurnal yang berjudul “Penerapan Algoritma *Naïve bayes* Untuk Klasifikasi Penyakit Diabetes Mellitus”. Didapat hasil akurasi sebesar 90.20% [5]. Selanjutnya jurnal yang berjudul “*Prediction Diabetes Mellitus Using Decision Tree Models*”. Didapat hasil akurasi sebesar 72.08%[6]. Kajian-kajian tersebut belum menerapkan metode *feature selection*, dimana *feature selection* berperan memilih subset yang tepat

dari set fitur asli, karena tidak semua fitur/atribut relevan dengan masalah, bahkan beberapa dari fitur atau atribut tersebut mengganggu dan dapat mengurangi akurasi[7]. *Noisy Features* atau fitur yang tidak terpakai tersebut harus dihapus untuk meningkatkan akurasi. Zeniarja, Widia, and Sani menerapkan metode feature selection pada penelitian yang berjudul “Penerapan Algoritma *Naive Bayes* dan *Forward Selection* dalam Pengklasifikasian Status Gizi Stunting pada Puskesmas Pandanaran Semarang” Hasil akurasi klasifikasi yang diperoleh dengan algoritma *naive bayes* saja yaitu sebesar 83,33%, sedangkan untuk algoritma *naive bayes* dengan fitur *Forward Selection* mencapai akurasi sebesar 86,00%. Peningkatan hasil akurasi tampak baik ketika dilakukan penggabungan algoritma *naive bayes* dengan fitur *Forward Selection*[8].

Berdasarkan penelitian sebelumnya peneliti ingin mengembangkan dan menerapkan *feature selection* menggunakan metode *forward selection* untuk menghilangkan atribut yang tidak berpengaruh dengan masalah, sehingga dapat meningkatkan nilai akurasi dari performa model.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah bagaimana mengoptimasi algoritma *naive bayes* menggunakan *feature selection* sehingga dapat meningkatkan nilai akurasi dari performa model?

## **1.3 Tujuan Penelitian**

Tujuan dari penelitian ini yaitu mengoptimasi algoritma *naive bayes* menggunakan *feature selection* sehingga dapat meningkatkan nilai akurasi dari performa model.

#### **1.4 Manfaat Penelitian**

Manfaat dari penelitian ini yaitu diharapkan dengan mengoptimasi algoritma *naïve bayes* menggunakan *feature selection* dapat meningkatkan nilai akurasi dari performa model, sehingga dapat digunakan untuk memprediksi resiko penyakit diabetes.

#### **1.5 Batasan Masalah**

Agar tidak menyimpang dari pembahasan masalah maka peneliti membatasi ruang lingkup penelitian ini hanya pada prediksi resiko penyakit diabetes menggunakan algoritma *naïve bayes* yang akan di optimasi menggunakan *feature selection* untuk meningkatkan nilai akurasi dari performa model. Dataset yang digunakan pada penelitian ini berasal dari situs *UCI Machine Learning* dengan judul dataset *Early stage diabetes risk prediction dataset* dengan jumlah record 520 dan 17 atribut meliputi *Age, gender, polyuria, polydipsia, sudden weight loss, weakness, polyphagia, genital thrush, visual blurring, itching, Irritability, delayed healing, partial paresis, muscle stiffness, alopecia, obesity* dan *class*. Pada penelitian ini menggunakan tools Rapid Miner 9.

#### **1.6 Sistematika Penulisan**

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

#### **BAB I PENDAHULUAN**

Pendahuluan meliputi latar belakang, ruang lingkup, rumusan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

#### **BAB II TINJAUAN PUSTAKA**

Bab ini berisi tentang teori-teori yang mendukung penelitian yang dilakukan oleh penulis/peneliti. Untuk peneliti yang melakukan analisis statistik, bab ini menyajikan kerangka dan hipotesis (jika diperlukan).

### **BAB III METODOLOGI PENELITIAN**

Dalam bab ini berisi objek penelitian, alat dan bahan, metode pengumpulan data, prosedur penelitian, pengukuran variabel dan metode analisis (metode-metode pendekatan penyelesaian permasalahan yang dipakai dan metode analisis data).

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

Bab ini memuat hasil penelitian, implementasi, analisis, dan hasil dari pembahasan. Hasil dan implementasi dapat berupa gambar alat/program dan aplikasinya. Untuk penelitian lapangan, hasilnya bisa berupa data (kualitatif atau kuantitatif). Analisis dan pembahasan berupa hasil pengolahan data.

### **BAB V KESIMPULAN DAN SARAN**

Bab ini memuat simpulan dan saran dari hasil pembahasan