

BAB I

PENDAHULUAN

1.1. Latar Belakang

Data mining merupakan pemrosesan suatu data menggunakan cara statistik, matematik, dll untuk mengidentifikasi suatu informasi pengetahuan potensial dan berguna yang tersimpan dalam basis data besar (Riadi., 2017). Klasifikasi adalah salah satu tugas dari data mining yang bertujuan untuk memprediksi label kategori benda yang tidak diketahui sebelumnya, dalam membedakan antara objek yang satu dengan yang lainnya berdasarkan atribut atau fitur (Parvin dkk., 2008).

Algoritma K-Nearest Neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. KNN memiliki keunggulan yaitu lebih efektif didata training yang besar, dapat menghasilkan data yang lebih akurat, dll. Naïve Bayes menurut Rani Puspita dkk, (2020) merupakan metode untuk klasifikasi text dengan kecepatan pemrosesan yang tinggi jika dalam jumlah besar. Terdapat beberapa kelebihan dari Naïve Bayes seperti efisien dalam dalam pelatihan dan penggunaannya, akurasi yang dihasilkan relatif tinggi dan metode ini hanya membutuhkan jumlah data training yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian (Saleh, 2015).

Diabetes mellitus, penyakit yang terjadi akibat dari terganggunya organ metabolisme yaitu organ pankreas tidak dapat memproduksi insulin yang cukup atau tubuh tidak dapat menggunakan secara efektif insulin tersebut (International Diabetes Federation, 2017). Penyakit diabetes dapat menyerang atau menjangkit siapa saja maupun usia berapapun, di Indonesia khususnya pada perempuan berdasarkan data dari Riskesdas pada tahun 2018 menyatakan bahwa prevalensi diabetes mellitus pada perempuan lebih tinggi dibandingkan laki-laki dengan perbandingan perempuan 1,78% dan laki-laki 1,4% serta untuk 5 tahun terakhir diabetes mellitus pada perempuan mengalami peningkatan sedangkan pada laki-laki mengalami penurunan. Diabetes mellitus merupakan penyebab kematian

terbesar nomer 3 di Indonesia setelah *stroke* dan jantung coroner. (Pusat Data dan Informasi Kementerian Kesehatan RI)

Seseorang dapat dinyatakan menderita diabetes mellitus dirasa sulit jika tidak dilakukan rekam medis atau uji laboratorium yang dapat menghasilkkan data yang valid. Untuk memudahkan proses diagnosis seseorang tersebut menderita penyakit ini, sudah banyak cara atau sistem untuk membantu memudahkan penentuan orang tersebut menderita penyakit ini atau tidak, dengan cara memperhitungkan beberapa factor atau ciri-ciri yang ada dalam penyakit diabetes mellitus tersebut.

Pada penelitian yang dilakukan oleh Januar Adi Putra, dkk, (2016) Dengan judul “Klasifikasi Pengidap Diabetes Pada Perempuan Menggunakan Penggabungan Metode *Support Vector Machine* dan K-Nearest Neighbor”. Dengan berdasarkan 768 data dan 8 atribut, hasil menunjukkan bahwa Algoritma *Support Vector Machine* mempunyai akurasi prediksi maksimum untuk diabetes sebesar 77.60%, dan K-Nearest Neighbor sebesar 91.00%. Sedangkan pada penelitian yang dilakukan oleh Fuad Nurhasan, dkk, (2018) yang berjudul “Perbandingan Algoritma C4.5, KNN, dan Naïve Bayes Untuk Penentuan Model Klasifikasi Penanggung Jawab BSI *Enterpenteure*”. Dengan 300 record data dan 12 atribut, menyimpulkan hasil dengan menggunakan metode C4.5 mempunyai nilai akurasi 73,33%, metode KNN mempunyai nilai akurasi 70% dan metode Naïve Bayes sebesar 80%.

Data set dalam penelitian ini termasuk berskala besar maka peneliti memutuskan untuk menggunakan metode Gaussian Naïve Bayes dan metode K-Nearest Neighbor. Kinerja dari kedua metode tersebut akan dibandingkan, sehingga dapat diketahui metode yang paling efektif dalam melakukan klasifikasi. Berdasarkan pada penelitian sebelumnya maka penulis ingin melakukan penelitian dengan judul “Perbandingan Algoritma K-Nearest Neighbor dan Gaussian Naïve Bayes pada klasifikasi penyakit Diabetes Melitus”.

1.2. Rumusan Masalah

Berdasarkan latar Belakang diatas, maka rumusan masalah pada penelitian adalah :

- a. Manakah algoritma yang paling tepat dalam mengklasifikasi penyakit diabetes mellitus?
- b. Berapa tingkat *accuracy*, *sensitivity*, *specifity*, *precision* dan *error rate* pada klasifikasi diabetes mellitus menggunakan algoritma K-Nearest Neighbor dan Gaussian Naïve Bayes?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah diatas maka tujuan penelitian ini adalah :

- a. Mengetahui algoritma yang paling tepat dalam klasidikasi diabetes mellitus.
- b. Mengetahui tingkat *accuracy*, *sensitivity*, *specifity*, *precision* dan *error rate* pada klasifikasi diabetes mellitus menggunakan algoritma K-Nearest Neighbor dan Gaussian Naïve Bayes

1.4. Batasan Masalah

Agar pembahasan dalam penilitian ini tidak melebar, maka terdapat beberapa batasan-batasan masalah sebagai berikut :

- a. Hanya memprediksi penyakit Diabetes Melitus pada perempuan.
- b. Menggunakan data yang bersumber dari situs <https://www.kaggle.com/salihacur/diabetes> tahun 2019.
- c. Jumlah data yang digunakan sebanyak 536 data.
- d. Menggunakan algoritma K-Nearest Neighbor dan Gaussian Naïve Bayes.
- e. Tools pada penelitian ini adalah Jupyter Notebook.

1.5. Manfaat Penelitian

Terdapat beberapa manfaat yang ingin dicapai dari penelitian ini, yaitu:

- a. Universitas

Menjadi tambahan pustaka di kampus Universitas Muhamadiyah Jember dan sebagai referensi tentang *Data Mining*.

b. Mahasiswa

Menambah wawasan mahasiswa tentang data mining khususnya algoritma K-Nearest Neighbor dan Gaussian Naïve Bayes.

c. Instansi

Instansi dapat menggunakan algoritma yang paling tepat dari K-Nearest Neighbor atau Gaussian Naïve Bayes

d. Penderita

Dapat mengetahui lebih awal tentang gejala penyakit diabetes melitus

