

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Diabetes atau penyakit gula (gula darah tinggi) adalah penyakit yang ditimbulkan karena penumpukan Glukosa dalam tubuh manusia. Glukosa merupakan sumber energi utama bagi sel tubuh, sehingga ketika terjadinya penumpukan glukosa di dalam darah karena tidak di serap sel tubuh dengan baik dapat menimbulkan berbagai gangguan pada tubuh. Penyakit diabetes merupakan penyakit yang tidak bisa dianggap remeh dan perlu ditangani dengan cepat. Jika penyakit diabetes tidak dikontrol dengan baik, maka dapat menimbulkan berbagai komplikasi yang membahayakan nyawa penderita[1].

Keterlambatan dalam diagnosis penyakit diabetes menjadi salah satu penyebab meningkatnya jumlah penderita diabetes. Tidak sedikit Pasien yang meninggal akibat komplikasi sebelum adanya diagnosis. Salah satu penyebab keterlambatan dalam penetapan diagnosis adalah dikarenakan adanya berbagai faktor yang berpengaruh terhadap pilihan yang ada atau beragamnya variabel. Diabetes menjadi penyakit yang berbahaya dengan ditandai tingginya kadar gula (glukosa) darah atau diatas nilai normal rata-rata[2].

Oleh karena itu, penting untuk mengembangkan metode yang efektif untuk mengklasifikasikan jenis penyakit. Penelitian ini menggunakan dua algoritma

yaitu *naïve bayes* dan KNN. Hal ini untuk membandingkan antara kedua algoritma tersebut yang memiliki tingkat akurasi yang terbaik. Selain itu, kedua algoritma tersebut digunakan untuk dapat menghasilkan informasi dari *dataset* diabetes yang digunakan[3].

Algoritma KNN dapat didefinisikan berdasarkan namanya dimana K melambangkan sejumlah nilai tetangga terdekat yang digunakan untuk mengidentifikasi sebuah kemiripan dengan tetangganya. Algoritma knn memiliki kelebihan karena sederhana, efektif dan telah banyak digunakan untuk klasifikasi[4].

Algoritma *naïve bayes* adalah algoritma pembelajaran mesin untuk masalah klasifikasi yang terutama digunakan untuk klasifikasi model yang melibatkan kumpulan data. Kita memprediksi kelas berdasarkan probabilitas kemunculan nilai fitur pada kelas tersebut[5].

Menurut beberapa penelitian, yang menggunakan metode klasifikasi *naïve bayes*, *k-nearest neighbor* dan *neural network* mendapatkan hasil klasifikasi dengan tingkat akurasi yang terbaik yaitu 76.92 % .Dengan menggunakan Algoritma *neural network* tingkat akurasi yang terbaik yaitu 82,97% .Dengan menggunakan algoritma *k-nearest neighbor* tingkat akurasi terbaik yaitu 93%[6].

Penelitian lainnya, dengan Hasil perbandingan terhadap diagnosa penyakit stroke, menggunakan klasifikasi metode *naïve bayes* dan *k-nearest neighbors* menghasilkan nilai *k-nearest neighbors* sebesar 94.36% dan *accuracy* terendah dimiliki oleh algoritma *naïve bayes* sebesar 90.10%[7].

Selanjutnya penelitian lainnya, terdapat persentase tingkat prediksi dari masing-masing data yang merupakan hasil prediksi terhadap input dari pengguna. Adapun *metode k-nearest neighbor* memiliki hasil akurasi sebesar 95% daripada *metode naive bayes* yang hanya memiliki nilai akurasi sebesar 93% terbukti menghasilkan akurasi terbaik untuk mengklasifikasi data[8].

Permasalahan klasifikasi sering muncul dalam berbagai bidang, salah satunya adalah diagnosis penyakit. Namun, tantangan utama dalam klasifikasi adalah menangani data yang tidak terstruktur, ketidakseimbangan data, dan tingkat akurasi model yang dipengaruhi oleh kualitas *dataset* dan pemilihan algoritma. Penelitian terdahulu telah banyak membahas pengembangan model klasifikasi dengan berbagai algoritma seperti *decision tree*, *support vector machine*, K-NN, *naïve bayes* dan *neural network*, namun masih terdapat keterbatasan pada aspek akurasi, efisiensi, dan kemampuan model dalam menangani data dengan dimensi tinggi atau data yang mengandung *noise*. Oleh karena itu, penelitian lanjutan diperlukan untuk mengatasi kendala tersebut dengan pendekatan yang lebih inovatif dan adaptif, sehingga dapat menghasilkan model yang lebih baik akurat dan stabil dan relevan untuk aplikasi dunia nyata.

Oleh karena itu pada penelitian ini yang akan diujikan perbandingan klasifikasi algoritma K-NN dan *naïve bayes* dengan parameter terdapat pada *dataset* penyakit diabetes[9].

1.2 RUMUSAN MASALAH

Berdasarkan dari uraian latar belakang di atas, maka tujuan dari penelitian ini adalah sebagai berikut :

1. Bagaimana melakukan klasifikasi *machine learning* terhadap penyakit diabetes pada penelitian terdahulu?
2. Bagaimana hasil evaluasi *machine learning* dapat mendeteksi penyakit diabetes dengan algoritma K-NN dan *naïve bayes*?

1.3 BATASAN MASALAH

1. Data yang digunakan adalah data sekunder yang dapat diakses oleh *public* melalui *kaggle: dataset of diabetes* yang memiliki data 100.000 data.
2. Penelitian akan fokus pada penerapan algoritma *k-nearest neighbors* dan *naïve bayes* dalam memprediksi diabetes.
3. Teknik evaluasi kinerja model hanya akan menggunakan metrik *accuracy*, *precision*, *recall*, dan *f1-score* yang diperoleh dari *Confusion Matrix*.
4. Atribut yang digunakan yaitu jenis *id*, *no_patien*, *gender*, *age*, *urea*, *cr*, *hba1c*, *chol*, *tg*, *hdl*, *ldl*, *vldl*, *bmi*, *class*.
5. Kelas yang akan di klasifikasi yaitu kelas *positif* diabetes dan *negative* diabetes.

1.4 TUJUAN PENELITIAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Berdasarkan rumusan masalah yang didapat, maka tujuan penelitian ini

Adalah sebagai berikut:

1. Menerapkan algoritma *k-nearest neighbor* dan *naïve bayes* dalam klasifikasi penyakit diabetes.
2. Mengukur dan menganalisis performa model prediksi yang dihasilkan berdasarkan metrik *accuracy*, *precision*, *recall*, dan *f1-score* melalui pengujian *Confusion Matrix*.

1.4.2 Manfaat Penelitian

Adapun penelitian ini memiliki manfaat sebagai berikut :

1. Bagi dunia medis: Penelitian ini dapat memberikan kontribusi dalam menyediakan alat bantu teknologi yang akurat dan efisien untuk memprediksi risiko diabetes, sehingga memudahkan dokter dalam membuat keputusan medis.
2. Bagi pengembangan teknologi: Memberikan wawasan baru tentang penggunaan *k-nearest neighbors* dan *naïve bayes* yang dapat diterapkan pada masalah klasifikasi lainnya.

3. Bagi peneliti lain: Hasil penelitian ini dapat menjadi acuan bagi penelitian lebih lanjut dalam pengembangan model prediksi menggunakan teknik *machine learning*, khususnya untuk prediksi penyakit kronis lainnya.

1.5 SISTEMATIKA PENELITIAN

Pada sistematika penulisan ini, penulis dapat menjelaskan secara umum apa saja yang dibahas dalam setiap bab pada penelitian ini. Dimana penelitian ini terdiri dari lima bab yang meliputi.

BAB I : PENDAHULUAN

Pada bab ini, penulis menjelaskan secara umum mengenai latar masalah, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian serta sistematika penelitian.

BAB II : LANDASAN TEORI

Pada bab ini, penulis mengutip sebuah gagasan dari pakar-pakar yang berhubungan dengan topik permasalahan yang berisikan definisi-definisi yang melandasi penelitian.

BAB III : METODOLOGI PENELITIAN

Pada bab ini, menjelaskan mengenai kerangka kerja penelitian, proses *machine learning* serta alat-alat pembantu dalam proses penelitian.

BAB IV : HASIL DAN ANALISIS

Pada bab ini, berisi tentang hasil data analisis penelitian serta sebuah perhitungan mengenai objek yang diteliti.

BAB V : PENUTUP

Pada bab ini, merupakan bab penutup dari seluruh penelitian yang berisikan kesimpulan dari seluruh bab yang ada dan saran yang bersangkutan mengenai penelitian ini.