

BAB I

PENDAHULUAN

1.1 Latar Belakang

Diabetes adalah gangguan metabolisme beberapa etiologi yang diidentifikasi oleh glukosa darah tinggi tingkat akibat kekurangan produksi insulin. Hal ini terkait dengan kerusakan jangka panjang, disfungsi dan kegagalan berbagai organ yang dapat menyebabkan penyakit serius komplikasi dan kondisi kesehatan, seperti mikrovaskular, penyakit kardiovaskular, dan serebrovaskular dan mungkin juga gangguan neuropati, nefropati, dan retinopati. Komplikasi ini menyebabkan kecacatan, penurunan kualitas hidup dan kematian dini di sebagian besar negara. Pada tahun 2013, untuk pertama kalinya Federasi Diabetes Internasional telah melaporkan tentang tingginya kadar glukosa darah dalam kehamilan dan diperkirakan 21,4 juta kelahiran hidup terpengaruh. Data ini menunjukkan bagaimana diabetes mempengaruhi setiap negara, setiap kelompok umur dan ekonomi dunia dan menentukan kebutuhan mendesak untuk diagnosis dini. Diagnosis medis adalah salah satu tugas yang paling penting[1].

Analisis data yang diambil dari pasien dan keputusan dokter spesialis adalah bagian terpenting dalam diagnosis penyakit. Seorang dokter biasanya mengambil keputusan berdasarkan pengetahuan dan pengalaman mereka. Ini melibatkan sejumlah analisis dari begitu banyak faktor yang mungkin memakan waktu berminggu-minggu atau bahkan berbulan-bulan dan mempersulit pekerjaan dokter[2]. Sayangnya, di negara berkembang kurangnya spesialis yang berpengalaman untuk penyakit tertentu, dapat menyebabkan diagnosis yang salah. Ini membuang-buang sumber daya, waktu, kualitas hidup, dan bahkan kehidupan manusia. Oleh karena itu untuk membantu dokter, kemungkinan kesalahan yang dilakukan oleh dokter yang kelelahan atau kurang berpengalaman harus diminimalisir.

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya kedalam kelas tertentu dari jumlah kelas yang tersedia. Proses ekstraksi fitur dalam klasifikasi penyakit diabetes tidak hanya menggunakan satu

pendekatan. Fitur-fitur yang digunakan dalam klasifikasi penyakit diabetes biasanya adalah gabungan dari berbagai teknik ekstraksi fitur. Selain digabungkan fitur yang dihasilkan dari berbagai teknik tersebut juga terkadang dikombinasikan dan diseleksi dengan beberapa cara sedemikian hingga diperoleh subset fitur yang dapat memberikan hasil akurasi terbaik. Proses ini sebenarnya adalah proses seleksi fitur.

Dalam penelitian sebelumnya menggunakan algoritma K-Nearest Neighbor (KNN) tentang klasifikasi penyakit kulit diperoleh akurasi sebesar 65% dan juga penelitian yang dilakukan oleh Made Bela Pramesthi Putri, Edy Santoso, dan Marji tentang penyakit kulit menggunakan algoritma Modified K-Nearest Neighbor(MKNN) diperoleh akurasi 86.67%. [3]

Dalam penelitian ini, peneliti mengusulkan metode KNN (K-Nearest Neighbor) dengan seleksi fitur Pearson Correlation untuk identifikasi penyakit yang diharapkan mampu mengetahui bagaimana implementasi dan akurasi dari metode KNN (K-Nearest Neighbor) dengan seleksi fitur Pearson Correlation untuk Identifikasi penyakit diabetes mellitus.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Diagnosa sering kali dibuat dengan berfokus pada naluri dan pengalaman dokter, bukan bersumber pada data yang dicakup dalam *database*. Hal ini menimbulkan dampak yang tidak diinginkan seperti kesalahan dalam mendiagnosis penyakit dan biaya pengobatan yang tidak perlu pada pasien, apalagi biaya pengobatan sangat mahal. Oleh karena itu diperlukan identifikasi dini penyakit diabetes mellitus menggunakan metode KNN (K-Nearest Neighbor) dan seleksi fitur Pearson Correlation.

Metode ini dipilih karena penulis ingin mengetahui apakah metode KNN dan seleksi fitur Pearson Correlation mampu jika diterapkan pada dataset penyakit diabetes karena berdasarkan penelitian-penelitian sebelumnya metode tersebut mampu memberikan akurasi yang tinggi jika diterapkan pada dataset yang berbeda.

1.2.2 Metode Usulan

Implementasi metode KNN (*K-Nearest Neighbor*) dan seleksi fitur *Pearson Correlation* untuk mengklasifikasi penyakit diabetes mellitus.

1.2.3 Pertanyaan Peneliti

1. Berapa nilai parameter K terhadap kinerja KNN, untuk menghasilkan klasifikasi penyakit diabetes mellitus?
2. Berapa hasil korelasi produksi insulin tertinggi dari metode KNN (*K-Nearest Neighbor*) dengan seleksi fitur *Pearson Correlation* untuk identifikasi penyakit jantung?.

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Peneliti

1. Meneliti pengaruh parameter K terhadap kinerja metode KNN dalam klasifikasi penyakit diabetes mellitus.
2. Meneliti hasil korelasi produksi insulin tertinggi dari metode KNN (*K-Nearest Neighbor*) dengan seleksi fitur *Pearson Correlation* untuk identifikasi penyakit jantung?.

1.3.2 Manfaat Penelitian

Membantu dan mengurangi risiko kesalahan para ahli dalam mendiagnosa penyakit diabetes .

1.4 Batasan – Batasan

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Data yang akan digunakan pada penelitian ini adalah data *Diabetes Data Set* dari situs *Kaggle Repository*
(<https://www.kaggle.com/datasets/akshaydattatraykhare/diabetes-dataset>)
2. Banyak data adalah 767 baris dengan 9 fitur
3. Menggunakan metode KNN (*Modified K-Nearest Neighbor*) dan seleksi fitur *Pearson Correlation*

1.5 Sistematika Proposal

Sistematika laporan penelitian ini terdiri dari beberapa bab, yaitu :

BAB 1 PENDAHULUAN

Pada bab ini membahas tentang latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, serta sistematika penulisan laporan.

BAB II KAJIAN PUSTAKA

Bab ini membahas landasan teori utama yang diperlukan dalam menyelesaikan penelitian dan penelitian sebelumnya yang telah digunakan untuk menyelesaikan dengan metode atau masalah yang sama.

BAB III METODE USULAN

Bab ini membahas tentang metode yang diusulkan (algoritma), arsitektur (*input*, proses, *output*), data penelitian, tahapan penelitian, serta rencana pengujian.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini membahas tentang hasil implementasi sistem dan hasil yang telah didapatkan dari proses uji coba klasifikasi penyakit diabetes mellitus berdasarkan dengan skenario pengujian,

BAB V PENUTUP

Pada bab ini membahas tentang kesimpulan dan saran berdasarkan hasil uji coba yang telah dilakukan sebagai pengembangan penelitian terkait untuk berikutnya.

REFERENSI

Berisi tentang daftar referensi yang dijadikan acuan dalam penelitian ini.