BAB 1 PENDAHULUAN

1. 1 Latar Belakang

Kesehatan merupakan salah satu peran yang sangat penting bagi tubuh kita, untuk menjaga pola tubuh agar tetap sehat adalah dengan melakukan olahraga dan menjaga atur pola makanan. Gunanya menjaga kesehatan adalah agar tidak terserang penyakit. Namun sebagian orang masih menyepelekan tentang kesehatan tubuhnya sehingga mudah terserang penyakit, salah satunya adalah terkena penyakit Diabetes. Diabetes adalah suatu penyakit yang diakibatkan terlalu banyak jumlah kadar gula dalam darah dikarenakan produksi dan pelepasan insulin yang tidak normal. Salah satu penyebabnya adalah sering mengkonsumsi makanan dan minuman yang mengandung gula cukup tinggi. Penyakit tersebut tergolong penyakit yang menyebabkan kompilkasi dan kematian. Penyakit ini juga menyebabkan kebutaan, gagal ginjal, penyakit jantung, hingga stroke.

Seorang wanita dapat dikatakan memiliki diabetes apabila hasil pengujian lab menunjukkan bahwa wanita tersebut seorang penderita diabetes dan tentu harus melakukan banyak pengujian karena penyakit tersebut menyebabkan komplikasi di tubuh penderitanya (Syafar, dkk, 2021). Dalam Diabetes Atlas 2023 (International Diabetes Federation) mengatakan bahwa di usia 21 tahun hingga usia 79 tahun keatas pada tahun 2023 diperkirakan 10,5% (536,6 juta orang), meningkat menjadi 12,2% (783,2 juta) pada 2045.

Di dalam perkembangan strategi klasifikasi digunakan dalam bidang medis saat ini untuk menangani berbagai macam masalah. Dalam profesi medis, misalnya penggunaan sistem kategorisasi untuk mengkategorikan tingkat keparahan penyakit pasien dapat membantu dokter mengidentifikasi diabetes melitus dengan lebih cepat. (Meristika, 2013). Beberapa Algoritma dapat digunakan untuk perhitungan proses klasifikasi. Algoritma klasifikasi diantaranya adalah Decission Tree C4.5, Naïve Bayes, dan K-Nearest Neighbor (KNN) (Kurniawan & Ivandari, 2017). Dalam

penelitian ini peneliti akan memakai metode K-Nearest Neighbor (KNN) untuk mengetahui klasifikasi terkenanya diabetes atau tidaknya.

Teknik untuk mengklasifikasikan objek berdasarkan data latih dengan nilai jarak yang lebih dekat dengan objek disebut dengan algoritma K-Nearest Neighbor (KNN). Pendekatan pembelajaran mesin non-parametrik adalah algoritma K-Nearest Neighbor (KNN). Model yang tidak mengandaikan distribusi data dikenal sebagai model non-parametrik. Salah satu manfaat model ini adalah dapat menghasilkan garis keputusan kelas yang sangat fleksibel dan nonlinier. Seringkali sulit untuk memilih k yang ideal. Meskipun metode K-Nearest Neighbor merupakan proses yang lugas, namun beberapa ahli mempertanyakan nilai k yang dihasilkan. (Yulianto & Darwis, 2021).

Dalam mendapatkan hasil penilitian ini, peneliti telah melakukan beberapa refrensi dari penelitian sebelumnya yang berkaitan dengan klasifikasi diabetes. Menurut (Ridwan, 2020) dengan judul "Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus". Dari judul tersebut peneliti melakukan perhitungan dengan dari dataset UCI Machine Learning yaitu Early Stage Diabetes Risk tahun 2020 dan terdiri 17 attribut dengan hasil akurasi sebesar 90.20% dan nilai AUCnya yaitu 0,95. Menurut (Hana, 2020) dengan judul "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision TreeC4.5". Dari judul tersebut hasil Pengujian menghasilkan akurasi yang cukup besar yaitu 97,12%, Precision sebesar 93,02% %, dan Recall sebesar 100,00%. Menurut (Wafa, dkk, 2022) dengan judul "Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM)". Dari judul tersebut para peneliti menghasilkan akurasi yang tinggi dengan mencapai sebesar 91%.

Dari konteks pada permasalahan penelitian tersebut, penelitian ini bertujuan untuk mengklasifikasi pasien tersebut apakah pasien tersebut mengidap penyakit diabetes atau tidaknya dengan menggunakan algoritma K-Nearst Neighbor. Tujuannya adalah agar dapat mengetahui secara akurasi, presisi, dan recall dalam

penggunaan algoritma K-Nearest Neighbor. Dari permasalahan tersebut,, maka peneliti melakukan penelitian dengan judul **"PENGGUNAAN ALGORITMA K-NEAREST NEIGHBOR DENGAN STRATEGI CROSS VALIDATION DALAM MENENTUKAN PENYAKIT DIABETES PADA WANITA"**.

1. 2 Rumusan Masalah

Rumusan masalah dari penelitian ini adalah "Berapakah tingkat akurasi, presisi, dan recall dengan menggunakan metode K-Nearest Neighbor pada data penyakit diabetes khusus wanita?"

1. 3 Tujuan Penelitian

Tujuan dalam melakukan penelitian klasifikasi ini adalah untuk mengetahui tingkat akurasi, presisi, dan recall dengan menggunakan algoritma KNN pada penyakit diabetes wanita berdasarkan data medis yang tersedia.

1. 4 Manfaat Penelitian

Dalam penelitian ini mempunyai beberapa manfaat, yaitu:

1. Dapat mengetahui kinerja K-Nearest Neighbor pada klasifikasi penyakit diabetes.
2. Dapat memberikan pengetahuan dan tingkat keakuratan pada suatu pemodelan algoritma.
3. Dapat menjadi refrensi untuk penelitian selanjutnya.

1. 5 Batasan Penelitian

Agar penlitian tidak menyimpang ke yang lain, maka dibutuhkannya batasan masalah. Berikut adalah batasan masalah dalam penelitian ini:

1. Peneliti hanya memfokuskan perhitungan dengan menggunakan metode K-Nearst Neighbor.
2. Tools yang digunakan pada penelitian ini adalah RapidMiner Studio, karena RapidMiner mampu menguji serta menjalankan model tanpa memerlukan pengetahuan pemrograman yang mendalam.

3. Data yang digunakan adalah data penyakit diabetes yang berasal dari website Kaggle (https://www.kaggle.com/datasets/whenamancodes/predict-diabities/data).
4. Banyaknya data yang digunakan dalam penelitian ini adalah 768 data.
5. Terdiri dari 8 atribut yaitu: Kehamilan, Glukosa, Tekanan Darah, Ketebalan Kulit, Insulin, BMI, Keturunan Diabetes, dan Umur.
6. Nilai K yang akan digunakan dalam penelitian ini adalah K= 5, K= 7, K= 11 dan K= 13. Alasan mengambil nilai tersebut supaya tidak memiliki jarak yang sama dalam proses klasifikasi, serta nilai ganjil dapat memberikan nilai akurasi optimal dan membantu mengurangi pengaruh outlier (memiliki nilai yang sangat jauh dari nilai rata-rata) (Novianti, dkk, 2024).
7. Kfold Cross Validation yang akan digunakan adalah CV= 4, CV= 6, CV= 8, dan CV= 12. Alasan mengambil kfold tersebut karena bersifat fleksibel untuk menyeimbangkan antara akurasi, presisi dan recall.