

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Demam *Tifoid* atau sering disebut tifus merupakan penyakit yang disebabkan oleh infeksi bakteri yang menyerang sistem pencernaan manusia. Hal ini ditandai dengan gejala demam yang berlangsung lebih dari seminggu, disertai gangguan saluran cerna, dengan atau tanpa gangguan kesadaran (TH, 2008). Penyakit ini disebabkan oleh bakteri *Salmonella Typhi* yang masuk ke dalam tubuh manusia melalui makanan atau minuman yang terkontaminasi (dr. Widoyono, 2011). Tifus merupakan penyakit menular yang tercantum dalam Undang-Undang Epidemi No. 6 Tahun 1962 dan dapat penyakit ini dapat menyerang banyak orang sehingga dapat menimbulkan wabah (Artanti, 2013). Oleh karena itu, diagnosis dan penanganan yang tepat sangat penting untuk mencegah komplikasi yang serius.

Pencegahan serta penanganan terhadap penderita demam *Tifoid* memerlukan kecepatan dan ketepatan untuk menghindari komplikasi atau keakutan. Misalnya gejala demam panas tinggi yang merupakan gejala pada demam *Tifoid* juga dimiliki oleh gejala pada penyakit lain. Demam panas tinggi juga dapat diketahui pada penyakit Covid-19, Demam Dengue, Malaria, dan Difteri. Diagnosa yang tepat sangat penting dalam hal ini. Untuk mengatasi hal ini, salah satu cara yang dapat digunakan adalah memanfaatkan metode klasifikasi. Klasifikasi adalah proses yang digunakan untuk menemukan model (atau fitur) dengan menggambarkan dan membedakan antara kelas data atau konsep (Han et al., 2012).

Teknik klasifikasi memiliki beberapa metode yang dapat digunakan, salah satunya adalah *K-Nearest Neighbor*. Metode *K-Nearest Neighbor (K-NN)* merupakan metode klasifikasi klasik yang dapat dikatakan juga dengan metode sederhana dan sering disebut merupakan metode *lazy*. Metode ini merupakan metode yang berbasis *instance*. *K-NN* bekerja dengan cara mengklasifikasikan objek berdasarkan jarak antara objek tersebut dengan objek lainnya. Metode *K-NN* menggunakan teknik menghitung mayoritas tetangga terdekat untuk memprediksi kelas baru (El Houby et al., 2017).

Metode *K-Nearest Neighbor* banyak digunakan dalam penelitian di bidang kesehatan, yaitu pada penelitian dengan judul *Case Based Reasoning* Diagnosis Penyakit Difteri dengan Algoritma *K-Nearest Neighbor*. Penelitian ini menitikberatkan pada kasus difteri di Jawa Timur pada tahun 2016. Pada penelitian ini dibangun sistem pakar menggunakan metode *Case-Based Reasoning (CBR)* dan *K-Nearest Neighbor (K-NN)*. Studi difteri ini menggunakan metode algoritma *K-Nearest Neighbor (K-NN)* yang melakukan perhitungan kesamaan antara kasus lama dan baru. Studi penyakit difteri ini telah disempurnakan dengan menggunakan *Case-Based Reasoning (CBR)*, sehingga menghasilkan hasil diagnosis yang lebih akurat. Hasil dari penelitian ini adalah bentuk diagnosis difteri berdasarkan gejala yang disajikan dengan akurasi sebesar 95,17% (Fatoni & Noviandha, 2018). Penelitian lain dengan judul *Klasifikasi Diagnosa Penyakit Demam Berdarah Dengue Pada Anak Menggunakan Metode K-Nearest Neighbor* Studi Kasus Rumah Sakit Pku Muhammadiyah Ujung Pangkah Gresik pada tahun 2021 ini menyebutkan bahwa berdasarkan implementasi pemodelan metode *K-Nearest Neighbor* pada 150 data latih dan 75 data uji diperoleh hasil pengukuran tingkat akurasi dari *Confusion Matrix* dengan nilai akurasi tertinggi mencapai 94,22 % pada $K=3$, $K=5$, $K=7$, $K=4$, $K=6$, dan $K=8$, adapun nilai K tersebut selanjutnya dapat digunakan sebagai landasan untuk tahap implementasi atau K tersebut menjadi model yang digunakan (Amilia & Prasetyo, 2021). Berdasarkan rangkaian latar belakang yang telah ditulis di atas, penulis selaku peneliti akan mengajukan penelitian dengan judul *Pemodelan Machine learning Pada Data Pasien Diagnosis Demam Tifoid Menggunakan Metode K-Nearest Neighbor*.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah ditulis di atas, berikut merupakan rumusan masalah yang menjadi pokok penelitian:

1. Bagaimana parameter terbaik yang dihasilkan dari pemodelan metode *K-Nearest Neighbor* pada klasifikasi data pasien Demam *Tifoid*?
2. Bagaimanakah hasil pengujian dari pemodelan yang dibangun oleh metode *K-Nearest Neighbor* pada data pasien Demam *Tifoid*?

1.3. Tujuan

Berdasarkan rumusan masalah yang dibangun, berikut tujuan yang ingin dicapai oleh penulis dalam penelitian ini:

1. Mengetahui nilai parameter *K-Nearest Neighbor* terbaik yang dihasilkan pemodelan pada data pasien Demam *Tifoid*.
2. Mengetahui nilai akurasi, presisi, dan *recall* dari hasil pengujian metode *K-Nearest Neighbor* pada data pasien Demam *Tifoid*.

1.4. Manfaat

Penulis berharap penelitian yang dibangun ini memiliki manfaat terhadap halayak. Berikut manfaat dari penelitian yang dibangun ini:

1. Pengembangan *Machine learning* pada bidang kesehatan, khususnya pada penderita demam *Tifoid*.
2. Penelitian ini dapat menjadi acuan untuk penelitian selanjutnya.
3. Penelitian ini dapat menjadi acuan atau landasan bagi pihak kesehatan dalam menangani penderita demam *Tifoid*.

1.5. Batasan Masalah

Luasnya penelitian penggunaan *Machine learning* dalam bidang kesehatan dapat menimbulkan bias pada penelitian ini. Mengatasi hal tersebut, peneliti memberi batasan-batasan pada penelitian yang sedang disusun yaitu:

1. Data yang digunakan dalam penelitian ini bersumber dari puskesmas Ampelgading Desa Tirtomarto, Kec. Ampelgading, Kabupaten Malang, Jawa Timur dengan jumlah data 275 pasien pada tahun 2022.
2. Metode klasifikasi yang digunakan adalah *K-Nearest Neighbor* dengan nilai $k = 3, 5, 7, 9, 11, 13, \text{ dan } 15$.
3. *Metrics Distance* yang digunakan dalam pengukuran jarak pada metode *K-Nearest Neighbor* adalah *Euclidean Distance*.
4. Pemodelan menggunakan *K Fold Cross Validation* dengan nilai $k = 10$.

5. Pengukuran dilakukan pada tingkat akurasi, presisi, dan *recall*.
6. Penulis dalam analisis dan implementasi metode yang digunakan dibantu dengan Python sebagai bahasa pemrograman, Jupyter Notebook sebagai IDE dalam penulisan kode pemrograman, dan Scikit-learn sebagai library yang membantu pengimplementasian metode *K-Nearest Neighbor*.
7. Untuk mengatasi *imbalanced* data atau ketidakseimbangan data menggunakan metode *Synthetic Minority Over-sampling Technique (SMOTE)*.

