

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG MASALAH

Machine Learning merupakan bagian dari kecerdasan buatan atau *Artificial Intelligence* dimana komputer dapat belajar sendiri tanpa perlu diajari secara eksplisit [1]. *Machine Learning* menerapkan perkembangan teknologi untuk peningkatan kemampuan komputer dalam pengolahan data yang besar (*Big Data*) [2]. *Machine Learning* mengandalkan pola dan langkah-langkah dari suatu algoritma dalam *Machine Learning* itu sendiri [3]. Terdapat beberapa jenis *Machine Learning* yaitu *Supervised Learning*, *Unsupervised Learning*, *Reinforcement Learning*, dan *Semi-Supervised Learning* [4]. Dalam *Machine Learning* terdapat beberapa metode yang dapat digunakan dalam pemrosesan data yaitu regresi, klustering, asosiasi, dan klasifikasi [5]. Dari beberapa metode dalam *Machine Learning* yang telah disebutkan sebelumnya, pengolahan data pada penelitian ini lebih tepat menggunakan metode klasifikasi.

Klasifikasi merupakan salah satu metode *Machine Learning* yang umumnya digunakan untuk menganalisis *dataset* yang memiliki label dan kelas [6]. Klasifikasi memiliki fungsi dalam proses *Machine Learning* untuk pengelompokan suatu data [7]. Berbagai macam algoritma klasifikasi yaitu C4.5, *Random Forest*, Jaringan Syaraf Tiruan, Genetika, *Fuzzy*, *Case-Based Reasoning*, *Naïve Bayes*, dan *K-Nearest Neighbor* [8]. Algoritma yang

digunakan dalam penelitian ini adalah *K-Nearest Neighbor*. *K-Nearest Neighbor* merupakan

algoritma dalam klasifikasi yang menggunakan metrik jarak *euclidean* untuk mengidentifikasi k tetangga terdekat dari data uji dan algoritma *K-Nearest Neighbor* banyak digunakan untuk prediksi ataupun klasifikasi penyakit [9]. Pada penelitian ini algoritma *K-Nearest Neighbor* digunakan untuk mengklasifikasi penderita anemia.

Anemia merupakan kondisi penurunan yang signifikan pada jumlah sel darah merah yang mengakibatkan penurunan besar dalam kemampuan darah membawa oksigen ke seluruh tubuh penderitanya [10]. Anemia dapat mempengaruhi seseorang diberbagai rentang usia. Namun, lebih cenderung terjadi pada wanita hamil dan remaja perempuan [11]. Pada survei kesehatan rumah tangga, diketahui sekitar 57% remaja perempuan Indonesia usia dibawah 14 tahun dan 39,5% perempuan usia diatas 15 tahun masih mengalami anemia. Anemia sering terlambat didiagnosis karena dianggap penyakit biasa oleh masyarakat [12]. Berdasarkan penjelasan diatas, perlu adanya upaya untuk mencegah terlambatnya diagnosis anemia pada masyarakat dengan melakukan klasifikasi penderita anemia.

Untuk mengklasifikasi penderita anemia diperlukan *dataset*. *Dataset* yang digunakan pada penelitian ini bernama *Dataset Anemia* yang bersumber dari *Kaggle Repository*. *Dataset Anemia* ini berisi 8.544 *record* dan mencakup 5 fitur serta 1 label dengan 2 kelas. Fiturnya terdiri dari Jenis Kelamin, Kadar *Hemoglobin*, *Mean Corpuscular Hemoglobin* (MCH), *Mean Corpuscular Hemoglobin Contracts* (MCHC), dan *Mean Corpuscular Volume* (MCV).

Labelnya yaitu *Result* dengan kelasnya yang terdiri dari menderita anemia atau tidak (diberi kode 0 untuk tidak anemia dan 1 untuk anemia).

Pada penelitian sebelumnya yang dilakukan oleh Zahirzada dkk. dalam prediksi anemia pada anak di Afghanistan menggunakan 5 algoritma klasifikasi yang salah satunya adalah *K-Nearest Neighbor* mendapatkan akurasi sebesar 69,4% [13]. Penelitian lain yang dilakukan oleh Yeruva dkk. pada klasifikasi anemia sel sabit menggunakan 6 algoritma klasifikasi yang salah satunya adalah *K-Nearest Neighbor* mendapatkan akurasi sebesar 83% [14]. Selanjutnya, pada penelitian yang dilakukan oleh Anand dkk. dalam prediksi anemia pada anak menggunakan beberapa algoritma *Machine Learning* yang salah satunya algoritma klasifikasi *K-Nearest Neighbor* dengan akurasi sebesar 60,36% [15]. Berdasarkan penelitian sebelumnya, penelitian ini dilakukan untuk mengetahui klasifikasi penderita anemia dan membandingkan akurasi model algoritma *K-Nearest Neighbor* dari penelitian sebelumnya.

Dari permasalahan yang sudah dijelaskan dapat dilihat bahwa perlu adanya klasifikasi penderita anemia agar dapat mengetahui faktor dari anemia dan perlu dilakukan peningkatan akurasi model algoritma *K-Nearest Neighbor*, maka penelitian ini diangkat dengan judul “**Penerapan Algoritma *K-Nearest Neighbor* Dalam Mengklasifikasi Penderita Anemia**”.

1.2 RUMUSAN MASALAH

Berdasarkan uraian latar belakang diatas, maka rumusan masalah dalam penelitian ini, yaitu :

1. Bagaimana penerapan algoritma *K-Nearest Neighbor* dalam mengklasifikasi penderita anemia pada *Dataset Anemia*?
2. Seberapa akurat model algoritma *K-Nearest Neighbor* dalam mengklasifikasi penderita anemia pada *Dataset Anemia*?

1.3 BATASAN MASALAH

Adapun beberapa hal yang perlu dibatasi agar penelitian ini lebih fokus dan tidak terjadi pembahasan diluar topik permasalahan. Berikut batasan masalah dalam penelitian ini:

1. *Dataset* yang digunakan pada penelitian ini bernama *Dataset Anemia* yang diperoleh dari *Kaggle Repository*, *dataset* ini memiliki jumlah data 8.545 *record* dan mencakup 5 fitur serta 1 label dengan 2 kelas.
2. Algoritma klasifikasi yang digunakan adalah algoritma *K-Nearest Neighbor*.
3. Matriks untuk mengevaluasi kinerja model algoritma klasifikasi digunakan *Confusion Matrix*.

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Adapun tujuan dari penelitian ini, sebagai berikut :

1. Menerapkan algoritma *K-Nearest Neighbor* dalam mengklasifikasi penderita anemia pada *Dataset Anemia*.
2. Mengukur akurasi model algoritma *K-Nearest Neighbor* dalam mengklasifikasi penderita anemia pada *Dataset Anemia*.

1.4.2 Manfaat Penelitian

Adapun manfaat dari penelitian yang akan dicapai :

1. Penelitian ini dapat memberikan kontribusi bagi pengembangan metode dalam mengklasifikasi penderita anemia.
2. Penelitian ini dapat memberikan rekomendasi bagi tenaga kesehatan untuk meningkatkan akurasi dalam mengklasifikasi penderita anemia.

1.5 SISTEMATIKA PENULISAN

Berikut penyajian sistematika penulisan dari penelitian ini agar mempermudah dalam memahami penulisan laporan :

BAB I : PENDAHULUAN

Pada bab ini menjelaskan tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Pada bab ini berisi penjelasan mengenai teori-teori yang digunakan untuk melandasi penelitian yang berasal dari kutipan artikel ilmiah dan buku yang berhubungan dengan topik penelitian.

BAB III : METODOLOGI PENELITIAN

Pada bab ini menjelaskan tentang kerangka kerja penelitian, kerangka *Machine Learning*, *data preprocessing*, penerapan algoritma, dan alat-alat serta bahan pendukung dalam penelitian ini.

BAB IV : HASIL ANALISIS DAN PEMBAHASAN

Bab ini menjelaskan tentang hasil dari analisis model algoritma, dan evaluasi dari hasil model algoritma pada penelitian.

BAB V : PENUTUP

Bab ini adalah penutup dari seluruh penelitian yang berisi kesimpulan dan saran dari penelitian yang dilakukan.