

BAB I PENDAHULUAN

1.1 Latar Belakang

Ketika kadar hemoglobin dalam darah berada dibawah ambang batas normal, hal tersebut dapat di katakan Anemia [1]. Anemia menjadi momok menakutkan bagi kesehatan masyarakat global [2]. Kelompok yang paling rentan terserang anemia adalah anak-anak, remaja putri yang sedang menstruasi, wanita hamil, dan wanita nifas. Penyebaran anemia yang luas ini menandakan perlunya perhatian serius dari semua pihak. Upaya pencegahan dan penanggulangan anemia perlu dilakukan secara komprehensif untuk melindungi kesehatan masyarakat, terutama kelompok-kelompok rentan tersebut. Menurut WHO (*world health Organization*) terdapat beberapa penyebab Anemia, termasuk defisiensi nutrisi akibat pola makan yang tidak memadai atau gangguan penyerapan nutrisi, infeksi (seperti malaria, infeksi parasit, tuberkulosis, dan HIV), peradangan, penyakit kronis, kondisi ginekologi dan obstetri, serta kelainan sel darah merah yang bersifat genetik. WHO memperkirakan bahwa anemia dialami oleh 40% anak usia 6–59 bulan, 37% wanita hamil, dan 30% wanita usia 15–49 tahun di seluruh dunia [3]

Beberapa pendekatan berbasis *machine learning* telah diterapkan untuk mendeteksi dan mengklasifikasikan anemia. Penelitian [4] menggunakan teknik K-Nearest Neighbors (KNN) guna memodelkan klasifikasi anemia aplastik. Hasil eksperimen menunjukkan bahwa KNN merupakan metode yang efektif dengan akurasi mencapai 97.56%. Selain itu, hasil evaluasi menunjukkan performa yang baik dalam hal metrik lain, dengan nilai F1-Score sebesar 0.9767 dan Recall sebesar 0.9545. Hasil ini menunjukkan bahwa KNN mampu mengklasifikasikan anemia aplastik dengan baik, tidak hanya dalam hal akurasi tetapi juga dalam kemampuan model untuk mendeteksi kasus positif secara konsisten. Pada penelitian [5] menerapkan tahapan lengkap mulai dari pra-pemrosesan data, balancing data dengan SMOTE, normalisasi, hingga tuning parameter pada *random forest* untuk klasifikasi. Namun penelitian ini memiliki keterbatasan yaitu hanya fokus pada algoritma *random forest* saja sehingga tidak ada pembandingan langsung dengan

algoritma lainnya.

Pada penelitian [6] memberikan tinjauan luas terhadap berbagai jenis algoritma Decision Tree seperti ID3, C4.5, CART, dan lainnya, serta membandingkan berdasarkan kriteria teknis. Namun penelitian ini bersifat tinjauan pustaka (Review), sehingga tidak menyajikan eksperimen atau implementasi asli oleh penulis. Pada penelitian [7] menunjukkan bahwa optimasi algoritma Naïve Bayes menggunakan Genetic Algorithm mampu meningkatkan akurasi klasifikasi data bank marketing sebesar 4,57%. Namun, kombinasi dengan metode Bagging hanya memberikan peningkatan yang sangat kecil, dan bahkan ketika digunakan sendiri, metode Bagging justru menurunkan performa model.

Salah satu tantangan utama dalam klasifikasi anemia adalah ketidakseimbangan kelas pada dataset, di mana distribusi data antara kategori anemia dan non-anemia tidak merata. Hal ini dapat mempengaruhi kinerja model klasifikasi. Oleh karena itu, penanganan ketidakseimbangan kelas dilakukan melalui pendekatan resampling. Pada penelitian [8], terdapat masalah ketidakseimbangan kelas, untuk menangani hal tersebut peneliti melakukan penanganan ketidakseimbangan kelas pada dataset karena mengatasi permasalahan ketidakseimbangan kelas dapat meningkatkan kinerja metode klasifikasi dalam hal akuisisi, sensitivitas, dan spesifisitas menggunakan pendekatan level algoritmik dan pendekatan data level. Penelitian menggunakan dua metode sampling yaitu SMOTE dan Adasyn yang diterapkan pada model klasifikasi KNN, Naïve bayes, Decision tree dan *Random Forest*. Penelitian ini memberikan hasil bahwa penggunaan metode resampling memberikan peningkatan nilai akurasi.

Pada penelitan ini akan dilakukan analisis pada beberapa algoritma diantaranya *KNN*, *Naive Bayes*, *Random Forest*, dan *Decision tree* untuk menemukan metode terbaik yang dapat digunakan dalam mengklasifikasikan penyakit anemia serta menerapkan smote dan adasyn yang diharapkan dapat mengatasi ketidakseimbangan kelas. Hasil penelitian tersebut kemudian akan diimplementasikan ke dalam website.

1.2 Rumusan Masalah

Dibawah ini merupakan rumusan masalah yang dilakukan pada penelitian tentang analisis dan penerapan model *machine learning* untuk klasifikasi dataset penyakit anemia, sebagai berikut :

1. Bagaimana cara menentukan model *machine learning* yang paling sesuai untuk klasifikasi dataset penyakit anemia berdasarkan karakteristik dataset?
2. Bagaimana kinerja model *machine learning* terkait klasifikasi penyakit anemia?

1.3 Batasan Masalah

Batasan ini ditetapkan untuk menjaga fokus penelitian dan memastikan bahwa tujuan penelitian dapat dicapai secara terarah dan terukur. Penelitian ini dibatasi pada analisis dan penerapan model *machine learning* untuk klasifikasi dataset penyakit anemia. Batasan-batasan yang diterapkan adalah sebagai berikut:

1. Dataset yang digunakan dalam penelitian ini terbatas pada data penyakit anemia yang telah tersedia secara publik dan di unduh dari repository kaggle (<https://www.kaggle.com/datasets/ehababoelnaga/anemia-types-classification>) .
2. Penelitian tidak mencakup analisis terkait implementasi klinis secara langsung, tetapi berfokus pada pengembangan dan evaluasi model klasifikasi berdasarkan dataset yang digunakan.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk menganalisis dan menerapkan model *machine learning* dalam klasifikasi dataset penyakit anemia. Secara khusus, tujuan yang dicapai adalah sebagai berikut:

1. Menganalisis algoritma *machine learning* yang paling sesuai untuk klasifikasi dataset penyakit anemia berdasarkan karakteristik dataset.
2. Mengukur kinerja model *machine learning* sebagai alat bantu analisis dan pengambilan keputusan di bidang kesehatan, khususnya dalam klasifikasi penyakit anemia.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat secara teoritis dan praktis, baik dalam pengembangan ilmu pengetahuan maupun dalam aplikasi di bidang kesehatan. Adapun manfaat penelitian ini adalah sebagai berikut:

1. Manfaat Teoretis

Secara teoretis, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan ilmu pengetahuan di bidang teknologi informasi dan kesehatan, khususnya dalam penerapan algoritma *machine learning* untuk klasifikasi data medis. Penelitian ini tidak hanya bertujuan untuk menjawab permasalahan praktis, tetapi juga untuk memperkaya literatur ilmiah serta memberikan dasar konseptual bagi penelitian-penelitian selanjutnya. Adapun manfaat teoretis dari penelitian ini meliputi:

1. Menambah wawasan dan literatur ilmiah mengenai penerapan model *machine learning* dalam klasifikasi dataset penyakit, khususnya anemia.
2. Memberikan pemahaman mendalam tentang pengaruh algoritma model *machine learning*.
3. Menghasilkan referensi bagi penelitian lanjutan dalam pengembangan metode klasifikasi berbasis *machine learning* untuk aplikasi kesehatan.

2. Manfaat Praktis

Penelitian ini memberikan manfaat praktis dalam mendukung proses klasifikasi penyakit anemia melalui penerapan algoritma *machine learning* yang telah teruji. Implementasi model ke dalam aplikasi berbasis web memungkinkan tenaga kesehatan melakukan diagnosis awal secara lebih cepat, efisien, dan akurat, serta menjadi alat bantu dalam pengambilan keputusan medis.

1.6 Sistematika Penulisan

Sistematika penulisan bermanfaat sebagai gambaran garis besar untuk memudahkan dalam memahami permasalahan dan pembahasan pada penelitian yang akan di lakukan sebagai pedoman dalam penulisan skripsi.

Adapun penulisan sistematika dalam skripsi ini adalah sebagai berikut :

BAB I PENDAHULUAN

Pada BAB I dijelaskan bagaimana latar belakang dibuatnya skripsi ini. Disini juga membahas tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan juga sistematika penulisan penelitian.

BAB II TINJAUAN PUSTAKA

Pada BAB II ini memuat tentang teori – teori dan beberapa elemen unsur yang di perlukan untuk mendukung berjalannya penelitian ini.

BAB III METODE PENELITIAN

Pada BAB III ini berisi tentang objek penelitian alur penelitian alat dan bahan yang digunakan serta menjelaskan tentang metode – metode yang digunakan dalam penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Pada BAB IV memuat hasil penelitian seperti hasil analisis yang di lakukan peneliti untuk merekomendasikan algoritma terbaik yang akan digunakan dalam melakukan klasifikasi penyakit anemia, dan juga menerapkan hasil analisis tersebut kedalam sebuah website.

BAB V PENUTUP

Pada BAB V berisi kesimpulan dari penelitian dan juga saran ataupun masukan terkait berbagai hal dari penelitian yang dilakukan, mulai dari analisis sampai dengan penerapan yang berguna untuk menambah referensi pembaca.