

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Anemia merupakan salah satu kondisi medis yang umum terjadi, ditandai dengan berkurangnya kemampuan darah dalam mengangkut oksigen ke seluruh jaringan tubuh. Kondisi ini umumnya disebabkan oleh rendahnya jumlah sel darah merah atau adanya kelainan pada hemoglobin [1]. Eritrosit memiliki peran penting dalam proses transportasi oksigen ke jaringan tubuh serta pembuangan karbon dioksida hasil metabolisme [2]. Anemia didiagnosis apabila konsentrasi hemoglobin berada di bawah ambang normal, yaitu <13 g/dL pada laki-laki, <12 g/dL pada perempuan tidak hamil, dan <11 g/dL pada perempuan hamil [3].

Penyebab anemia sangat beragam, antara lain kehilangan darah dalam jumlah besar, penyakit kronis, kelainan genetik, serta defisiensi nutrisi seperti kekurangan zat besi, vitamin B12, dan asam folat [4]. Secara klinis, anemia diklasifikasikan berdasarkan tingkat keparahannya menurut kriteria *World Health Organization* (WHO), yaitu anemia ringan, sedang, dan berat. Klasifikasi ini ditentukan berdasarkan konsentrasi hemoglobin dalam darah, di mana anemia ringan ditandai dengan sedikit penurunan dari ambang batas normal, anemia sedang menunjukkan penurunan yang lebih lanjut, sedangkan anemia berat ditandai dengan kadar hemoglobin yang sangat rendah sehingga berpotensi menimbulkan dampak serius terhadap kesehatan [5].

Menurut *World Health Organization* (WHO), anemia masih menjadi salah satu masalah kesehatan global dengan jumlah penderita yang sangat besar. Diperkirakan sekitar 1,62 miliar orang di seluruh dunia menderita anemia, termasuk 43% anak balita dan sekitar 300.000 bayi yang lahir setiap tahunnya [6][7]. Kondisi ini menunjukkan bahwa anemia tidak hanya berdampak pada kelompok usia tertentu, tetapi juga menyerang hampir semua lapisan masyarakat, mulai dari bayi, anak-anak, remaja, hingga orang dewasa. Pada tahun 2019, prevalensi global anemia tercatat mencapai 22,8% [8], sehingga dapat disimpulkan bahwa hampir seperempat populasi dunia masih menghadapi permasalahan ini.

Di Indonesia, anemia merupakan isu kesehatan masyarakat yang serius. Prevalensi pada remaja masih cukup tinggi, yakni 27,2% pada anak perempuan dan 20,3% pada anak laki-laki berusia 15–24 tahun [9], dan angka anemia di daerah pedesaan lebih tinggi dibandingkan dengan di daerah perkotaan [10]. Hampir satu dari empat remaja perempuan mengalami anemia, terutama akibat kebutuhan zat besi yang lebih tinggi saat menstruasi, status gizi yang kurang optimal, serta pola makan dan gaya hidup yang tidak mendukung. Kondisi ini menegaskan bahwa anemia pada remaja perempuan masih menjadi masalah kesehatan masyarakat yang serius [11].

Diagnosis anemia tidak mudah dilakukan, karena gejala klinis seperti kelelahan, sesak napas, dan pucat pada telapak tangan atau konjungtiva memiliki sensitivitas rendah dan spesifisitas sedang, sehingga kurang memadai jika dijadikan dasar utama diagnosis [12]. Sebagai contoh, membedakan anemia defisiensi besi dengan thalassemia membutuhkan pemeriksaan lanjutan seperti analisis DNA atau

pengukuran kadar feritin, yang tidak selalu tersedia karena memerlukan peralatan khusus, keahlian teknis, serta biaya dan waktu yang besar [13]. Di negara berpenghasilan rendah, keterbatasan fasilitas laboratorium dan layanan kesehatan dasar turut menghambat pelaksanaan skrining anemia secara menyeluruh [14].

Anemia umumnya didiagnosis melalui pemeriksaan laboratorium seperti hitung darah lengkap, feritin serum, dan elektroforesis hemoglobin [15]. Namun, metode ini masih menghadapi keterbatasan, antara lain kurangnya fasilitas laboratorium dan tenaga medis terlatih, potensi kesalahan dalam interpretasi hasil, serta keterlambatan diagnosis akibat beban kerja tinggi dan distribusi sampel [16][17]. Kondisi tersebut dapat menghambat penanganan pasien secara tepat waktu, sehingga diperlukan metode alternatif yang lebih cepat, akurat, dan mudah diakses, salah satunya dengan memanfaatkan teknologi berbasis kecerdasan buatan.

Perkembangan teknologi informasi, khususnya di bidang *machine learning*, telah mendorong hadirnya sistem cerdas yang mampu membantu proses diagnosis penyakit secara lebih cepat, akurat, dan efisien, terutama pada wilayah dengan keterbatasan sumber daya medis [18][5][19]. Berbagai penelitian telah menunjukkan bahwa *machine learning* berhasil diterapkan dalam klasifikasi penyakit dengan hasil yang menjanjikan, baik dalam meningkatkan ketepatan diagnosis maupun mendukung pengambilan keputusan klinis [20].

Namun demikian, kinerja model sangat dipengaruhi oleh kualitas data serta relevansi fitur yang digunakan dalam proses pelatihan [21]. Keberadaan fitur yang tidak relevan atau bersifat redundan dapat meningkatkan kompleksitas komputasi sekaligus menurunkan akurasi prediksi [22][23]. Dengan demikian, pemilihan fitur

(*feature selection*) merupakan tahap penting dalam meningkatkan kinerja model pembelajaran mesin, karena memungkinkan algoritma berfokus pada atribut yang relevan sehingga efisiensi komputasi dan akurasi analisis data medis dapat tercapai [24].

Penelitian mengenai penerapan *machine learning* dalam klasifikasi penyakit telah banyak dilakukan, namun studi yang secara khusus membandingkan metode pemilihan fitur untuk mengoptimalkan klasifikasi anemia masih relatif terbatas [25]. Penelitian oleh Tuba et al. (2021) menggunakan algoritma Naïve Bayes untuk klasifikasi anemia dengan tingkat akurasi sebesar 74%, tetapi tanpa melibatkan tahap pemilihan fitur sehingga menimbulkan beban komputasi yang lebih tinggi [5]. Selanjutnya, Manasvi et al. (2023) menerapkan algoritma *K-Nearest Neighbor* untuk klasifikasi anemia dan hanya memperoleh akurasi 63,15%. Penelitian tersebut juga membandingkan beberapa model klasifikasi, namun tidak menyertakan proses pemilihan fitur sehingga berpotensi mengurangi peluang optimasi kinerja [26]. Sementara itu, Justice et al. (2024) menggunakan algoritma *Support Vector Machine* (SVM) dengan akurasi 89,45% untuk prediksi anemia, tetapi tetap tidak melibatkan metode pemilihan fitur, sehingga membatasi potensi peningkatan performa model [19]. Secara umum, algoritma Naïve Bayes, *K-Nearest Neighbor*, dan *Support Vector Machine* banyak digunakan dalam diagnosis medis karena memiliki tingkat interpretabilitas yang baik, relatif sederhana, serta telah terbukti efektif dalam menyelesaikan permasalahan klasifikasi pada data kesehatan [27].

Meskipun *machine learning* telah banyak digunakan dalam klasifikasi penyakit, penelitian komparatif yang secara khusus mengevaluasi metode pemilihan fitur ANOVA dan *Chi-Square* untuk klasifikasi anemia masih terbatas. Sebagian besar penelitian sebelumnya cenderung mengabaikan tahap pemilihan fitur atau hanya menggunakan satu metode tertentu, sehingga peluang untuk mengoptimalkan kinerja model menjadi kurang maksimal [5][26][19]. Padahal, dataset klinis yang digunakan dalam klasifikasi anemia umumnya terdiri atas kombinasi variabel numerik dan kategorikal yang memerlukan strategi pemilihan fitur yang tepat agar atribut paling relevan dapat dipertahankan. Ketidadaan strategi yang sesuai seringkali menyebabkan performa model tidak optimal serta membatasi kemampuan generalisasi [28]. Pemilihan fitur yang kurang tepat juga dapat menimbulkan bias pada proses pembelajaran model dan meningkatkan risiko *overfitting*, terutama pada dataset kesehatan berukuran kecil hingga menengah [29].

Penelitian oleh Shanthi (2024) menunjukkan bahwa penerapan teknik pemilihan fitur mampu meningkatkan akurasi prediksi sekaligus efisiensi komputasi pada dataset kesehatan berdimensi tinggi [30]. Temuan tersebut menegaskan pentingnya menjaga relevansi fitur agar kinerja model dapat dioptimalkan, khususnya dalam analisis data medis di mana keberadaan atribut yang tidak relevan atau redundan dapat menutupi pola penting. Berdasarkan hal tersebut, penelitian ini berupaya mengisi kesenjangan dengan melakukan analisis komparatif antara metode ANOVA dan *Chi-Square* pada tiga algoritma klasifikasi yang banyak digunakan, yaitu *Naïve Bayes*, *K-Nearest Neighbor*, dan *Support Vector Machine*. Melalui perbandingan ini, penelitian diharapkan dapat

menentukan pendekatan yang paling efektif untuk meningkatkan akurasi klasifikasi anemia sekaligus mendukung pengembangan sistem diagnosis yang lebih efisien dan andal [31][32].

ANOVA umumnya digunakan untuk menganalisis pengaruh statistik pada fitur numerik dengan cara mengevaluasi varians antar kelas, sedangkan *Chi-Square* digunakan untuk menilai fitur kategoris berdasarkan tingkat independensinya terhadap label kelas [33][34]. Kedua metode ini memiliki keunggulan dari sisi efisiensi komputasi serta sesuai untuk diterapkan pada dataset dengan tipe campuran. Pada penelitian ini, data yang digunakan berasal dari *Kaggle* dengan karakteristik berupa atribut numerik dan kategoris, sehingga penerapan strategi pemilihan fitur yang tepat menjadi penting untuk meminimalkan redundansi informasi sekaligus meningkatkan efektivitas model. Dataset tersebut terdiri atas sekitar 15.300 *instance* dengan 25 fitur, meliputi atribut seperti usia, kadar hemoglobin, volume sel darah rata-rata, serta keberadaan gejala klinis.

Penelitian ini berfokus pada klasifikasi anemia menggunakan Algoritma Naive Bayes, *K-Nearest Neighbor* (KNN), dan *Support Vector machine* (SVM). Naive Bayes unggul dalam kecepatan dan menangani banyak fitur dengan asumsi independensi variabel. Sementara itu *K-Nearest Neighbor* (KNN) dikenal sederhana, mudah diterapkan, dan tidak memerlukan pelatihan kompleks, memungkinkan klasifikasi cepat. *Support Vector Machine* (SVM) efektif untuk data berdimensi tinggi dengan *hyperplane* optimal, terutama pada data yang tidak terpisahkan secara linear. Selain itu, seleksi fitur *Analysis of Variance* (ANOVA) meningkatkan kinerja model dengan memilih fitur paling berpengaruh dan

mengurangi dimensi data tanpa kehilangan informasi penting. *Chi-Square* juga digunakan sebagai metode seleksi fitur yang dapat mengukur hubungan antara fitur dengan kelas target, terutama pada data kategorikal. Metode ini membantu menghilangkan fitur yang kurang relevan, sehingga meningkatkan efisiensi dan akurasi model klasifikasi. Dengan mempertimbangkan kelebihan-kelebihan tersebut, penelitian ini berjudul **“Perbandingan Metode Seleksi Fitur ANOVA dan *Chi-Square* Untuk Peningkatan Performa *Machine Learning* Dalam Klasifikasi Anemia”** yang bertujuan untuk mengevaluasi dan membandingkan akurasi ketiga algoritma dalam memprediksi penyakit anemia.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang diatas, maka rumusan masalah pada penelitian ini adalah :

1. Bagaimana *machine learning* dapat diterapkan dalam memprediksi penyakit anemia?
2. Bagaimana performa Algoritma Naive Bayes, *K-Nearest Neighbor* (KNN), dan *Support Vector Machine* (SVM) dalam memprediksi penyakit anemia?
3. Bagaimana membangun model klasifikasi yang mampu memberikan informasi akurat dalam memprediksi penyakit anemia?

1.3 BATASAN MASALAH

Agar nantinya permasalahan sesuai dengan rumusan masalah dan tidak meluas, permasalahan perlu dibatasi sebagai berikut :

1. Menggunakan metode Algoritma Naïve Bayes, *K-Nearest Neighbor* (KNN), dan *Support Vector Machine* (SVM) untuk memprediksi penyakit anemia.

2. Penelitian ini menggunakan seleksi fitur *Analysis of Variance* (ANOVA) dan *Chi-Square* untuk meningkatkan performa akurasi.
3. Penelitian ini menggunakan dataset dari penyakit Anemia dengan total keseluruhan data sebanyak 15.300 baris data dan 25 atribut.

1.4 TUJUAN PENELITIAN

Tujuan dari penelitian ini adalah sebagai berikut :

1. Mengevaluasi kinerja Algoritma Naïve Bayes, *K-Nearest Neighbor* (KNN), dan *Support Vector Machine* (SVM) untuk memprediksi penyakit anemia.
2. Mendapatkan akurasi yang baik untuk klasifikasi prediksi penyakit anemia dengan menggunakan Algoritma Naïve Bayes, *K-Nearest Neighbor* (KNN), dan *Support Vector Machine* (SVM).
3. Menghasilkan model klasifikasi yang dapat memberikan informasi akurat dalam memprediksi penyakit anemia dengan akurasi yang tinggi.

1.5 MANFAAT PENELITIAN

Adapun manfaat dari penelitian ini adalah sebagai berikut :

1. Dapat mengetahui status penyakit anemia dengan melihat tingkat akurasi yang tinggi.
2. Dapat mengetahui tingkat akurasi dari ketiga metode yaitu Algoritma Naïve Bayes, *K-Nearest Neighbor* (KNN), dan *Support Vector Machine* (SVM) dengan melihat perbandingan hasil akurasi yang diperoleh.
3. Hasil penelitian ini dapat menjadi rujukan untuk penelitian serupa yang akan dilakukan di masa depan.

1.6 SISTEMATIKA PENULISAN

Sistematika penulisan ini menggambarkan secara umum mengenai apa yang akan penulis bahas dalam setiap bab dari laporan ini. dimana sistematika penulisan ini terdiri dari (5) bab meliputi :

BAB I : PENDAHULUAN

Bab ini memaparkan latar belakang penelitian mengenai tingginya prevalensi anemia dan kendala dalam diagnosis yang akurat, serta pentingnya penerapan *machine learning* untuk mendukung klasifikasi anemia secara cepat, tepat, dan efisien. Dalam bab ini dirumuskan **masalah penelitian** yang akan dijawab, ditetapkan **batasan masalah** agar ruang lingkup penelitian menjadi jelas dan terfokus, serta dijelaskan **tujuan penelitian** dan manfaatnya, baik secara teoritis maupun praktis, khususnya dalam mendukung pengambilan keputusan klinis. Bab ini juga menyajikan **sistematika penulisan**, yang memberikan gambaran susunan isi laporan penelitian secara keseluruhan.

BAB II : LANDASAN TEORI

Bab ini berisi telaah teori dan konsep yang menjadi landasan penelitian. Pembahasan diawali dengan pengertian *Artificial Intelligence* (AI) beserta perkembangannya dalam berbagai bidang. Selanjutnya dijelaskan konsep *machine learning* sebagai salah satu

cabang AI, termasuk peranannya dalam bidang kesehatan untuk mendukung proses diagnosis medis. Bab ini juga membahas algoritma yang digunakan dalam penelitian, yaitu *Naïve Bayes*, *K-Nearest Neighbor*, dan *Support Vector Machine*. Selain itu, diuraikan pula metode seleksi fitur ANOVA dan *Chi-Square* yang diterapkan untuk meningkatkan akurasi serta efisiensi model klasifikasi.

BAB III : METODOLOGI PENELITIAN

Bab ini menjelaskan jenis penelitian, sumber data yang digunakan, serta tahapan *preprocessing* seperti pembersihan, normalisasi, dan penanganan *missing value*. Selanjutnya dipaparkan proses seleksi fitur dengan metode ANOVA dan *Chi-Square* untuk memilih atribut yang relevan. Penerapan algoritma *Naïve Bayes*, *K-Nearest Neighbor*, dan *Support Vector Machine* juga dijelaskan beserta parameter yang digunakan. Bagian akhir membahas metode evaluasi performa menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*.

BAB IV : ANALISIS DAN HASIL

Bab ini menyajikan hasil implementasi klasifikasi penyakit anemia dengan berbagai skenario. Bagian awal menjelaskan dataset dan tahap *preprocessing*, kemudian diikuti penerapan algoritma *Naïve Bayes*, *K-Nearest*

Neighbor, dan *Support Vector Machine*. Proses seleksi fitur dengan ANOVA dan *Chi-Square* dipaparkan untuk menunjukkan atribut paling relevan. Hasil klasifikasi, baik dengan maupun tanpa seleksi fitur, disajikan dalam tabel dan grafik, kemudian dibandingkan berdasarkan metrik evaluasi.

BAB IV : PENUTUP

Bab ini berisi simpulan penelitian mengenai pengaruh seleksi fitur ANOVA dan *Chi-Square* terhadap peningkatan performa algoritma *machine learning* dalam klasifikasi penyakit anemia. Selain itu, diberikan saran untuk penelitian selanjutnya terkait pengembangan metode, perluasan dataset, dan penerapan dalam sistem pendukung keputusan di bidang kesehatan.