

BAB I

PENDAHULUAN

1.1. Latar Belakang

Anemia merupakan penyakit yang menjadi pengaruh dalam peningkatan mortalitas, morbiditas, penurunan produktivitas kerja dan gangguan perkembangan *neurologis* sebanyak sepertiga populasi di dunia [1]. Anemia merupakan masalah kesehatan masyarakat, dikarenakan prevalensinya masih di atas 20%. Presentase sebesar 30% - 50% mengemukakan bahwa penyebab anemia pada anak-anak, remaja, dan dewasa disebabkan oleh kekurangan zat besi [2]. Anemia lebih sering terjadi pada remaja putri dibandingkan dengan remaja laki-laki dikarenakan remaja putri kehilangan zat besi (*Fe*) saat menstruasi sehingga membutuhkan lebih banyak asupan zat besi (*Fe*). Kebiasaan remaja putri yang ingin memiliki tubuh yang ideal menjadikan remaja tersebut membatasi asupan makanan harian yang beresiko lebih besar terserang anemia [1]. Tidak hanya remaja putri, ibu hamil juga menjadi salah satu kategori yang sering terpapar penyakit anemia [3].

Persentase anemia di Indonesia pada ibu hamil di tahun 2013 sebesar 37,1%, hal ini mengalami peningkatan di tahun 2018 menjadi 48,9% kelompok usia 15-24 tahun dan 25-34 tahun [4]. Prevalensi anemia yang tinggi pada ibu hamil menyebabkan gangguan pasokan nutrisi yang dibutuhkan janin pada proses perkembangannya, hal ini meningkatkan risiko ibu dan janin seperti kelahiran

prematum hingga kematian janin [3]. Hal tersebut menjadi sebuah argumen penguat bahwa penyakit anemia dapat memicu penyakit lainnya menjadi lebih buruk seperti salah satunya resiko untuk penyakit *Covid-19* [1].

Mengacu pada penelitian yang dilakukan oleh peneliti dari Cina [5], penderita penyakit anemia mempunyai resiko kematian lebih tinggi jika terpapar virus korona dikarenakan virus *Covid-19* melakukan tindakan seluler dengan menyerang sel darah merah, yang mengakibatkan sel darah merah tidak mampu mengangkut oksigen. Berdasarkan pernyataan sebelumnya, dikhawatirkan anemia semakin menjadi penyakit yang membahayakan mengingat data persentase penurunan anemia secara keseluruhan termasuk lambat dan tidak merata untuk semua kelompok usia dan jenis kelamin, anemia hanya menurun sekitar tujuh poin persentase antara tahun 1990 dan 2016, dari 40% menjadi 33% [1].

Oleh maka dari itu, anemia dapat diprediksi untuk pencegahan dan penanganan lebih awal agar tidak masuk kedalam tingkatan yang lebih berbahaya. Prediksi penyakit dalam ilmu kedokteran pada waktu yang tepat adalah masalah utama bagi para profesional untuk pencegahan dan rencana perawatan yang efektif [6]. Dikutip dari penelitian [7], saat ini *data mining* adalah metode yang dimanfaatkan untuk pembuatan aplikasi dalam ilmu kedokteran. *Data mining* adalah istilah populer yang digunakan secara bergantian dalam proyek yang terlibat dengan big data [8]. Penggunaan *data mining* pada bidang kesehatan salah satunya untuk mendukung dokter atau tenaga medis dalam memberikan perawatan yang sesuai, dengan demikian pasien diperlakukan lebih baik dengan

layanan kesehatan yang lebih murah dan terjangkau [9]. Teknik *data mining* yang sering digunakan pada studi penelitian bidang kesehatan adalah *predictive* model yang terbagi atas *classification* dan *prediction* [8].

Penelitian terdahulu mengenai komparisasi algoritma dilakukan oleh Md. Nurul Amin, Md. Ahsan Habib pada tahun 2015 dengan membandingkan algoritma *J48 Decision Tree* dan *Naïve Bayes*. Eksperimen dilakukan dua kali, sehingga mempunyai 2 hasil akurasi dari masing-masing algoritma. Rata-rata akurasi terbesar ada pada algoritma *J48 Decision Tree* sebesar 94% [10].

Penelitian selanjutnya sudah pernah dilakukan oleh Manal Abdullah pada tahun 2016 menggunakan *dataset* 41 pasien anemia. Pasien anemia diklasifikasikan berdasarkan jenisnya dan algoritma yang dibandingkan adalah algoritma *Naïve Bayes*, *J48 Decision Tree*, dan *Support Vector Machine* dengan bantuan *tools* WEKA. Hasil menunjukkan bahwa algoritma *J48 Decision Tree* mempunyai akurasi tertinggi sebesar 93,75% [11].

Komparasi algoritma *data mining* juga dilakukan oleh Manish Jaiswal, Anima Srivastava, dan Tanveer J. Siddiqui pada tahun 2019 dengan membandingkan algoritma *Random Forest*, *Naïve Bayes*, dan *Decision Tree*. Data yang digunakan merupakan data *Complete Blood Count* dengan 200 sampel data. Hasil akurasi terbesar ada pada algoritma *Naïve Bayes* sebesar 96% [6].

Penelitian mengenai prediksi anemia menggunakan algoritma *Decision Tree* dilakukan oleh Kanaak Mena dan *team* pada tahun 2019. Penelitian ini hanya

menggunakan data hemoglobin dan data demografi untuk dicari korelasinya. Akurasi yang dihasilkan sebesar 97% untuk prediksi menggunakan *hemoglobin*, dan 44% untuk hubungan antara ibu dan anak [12].

Berdasarkan acuan penelitian terdahulu, penelitian ini akan menggunakan teknik *data mining* dengan *framework* CRISP-DM dalam prediksi klasifikasi jenis-jenis anemia dengan menghasilkan *dashboard* visualisasi analisis pada tahapan *deployment* [13]. Mengacu pada penelitian [11] dan [6], penelitian ini menggunakan jumlah dataset pasien anemia yang lebih banyak dan menggunakan data tambahan berupa data demografi untuk menganalisa faktor dari masing-masing pasien anemia. Berbeda dari penelitian [12], variable penelitian tidak hanya berdasarkan 1 variabel yaitu *hemoglobin*, tetapi terdapat 4 variabel parameter lain untuk klasifikasi jenis penyakit anemia. Algoritma *supervised learning* yang akan dipakai adalah *Naïve Bayes*, *J48 Decision Tree*, dan *Random Forest* sebagai pembanding dari penelitian [10] yang hanya menggunakan dua algoritma *supervised learning*.

1.2.Rumusan Masalah

Terdapat rumusan masalah yang dapat disimpulkan berdasarkan latar belakang yang sudah dijabarkan sebelumnya. Berikut ini adalah rumusan masalah pada penelitian ini :

1. Bagaimana proses pembuatan model prediksi klasifikasi jenis penyakit anemia berdasarkan data *National Health And Nutrition Examination Survey*?

2. Bagaimana hasil perbandingan akurasi, sensitivitas, dan presisi antara algoritma *Naïve Bayes*, *J48 Decision Tree*, dan *Random Forest*?
3. Bagaimana hasil analisis visualisasi data berdasarkan data prediksi klasifikasi jenis penyakit anemia dengan data demografi?

1.3. Batasan Masalah

Terdapat batasan masalah pada penelitian kali ini, berikut adalah batasan masalah penelitian ini :

1. Penelitian ini hanya membahas mengenai prediksi klasifikasi jenis penyakit anemia menggunakan teknik *data mining*.
2. Penelitian ini hanya membandingkan hasil pengolahan data menggunakan algoritma *Naïve Bayes*, *J48 Decision Tree*, dan *Random Forest*.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk memprediksi dan mengklasifikasi jenis penyakit anemia dengan memanfaatkan teknik *data mining*. Hasil akhir yang akan dicapai adalah sebuah model algoritma dengan hasil akurasi yang paling optimal.

1.5. Manfaat Penelitian

Berikut ini adalah manfaat dari kegiatan penelitian ini :

1. Membantu tenaga medis dalam menentukan penyakit anemia melalui model pada sistem yang dibuat.
2. Menjadi sebuah referensi dalam menentukan solusi pencegahan dan pengobatan yang tepat untuk pasien penderita penyakit anemia sesuai dengan jenis-jenis penyakit anemia.
3. Menjadi sebuah informasi tambahan bagi dunia akademis sebagai referensi dan pengetahuan baru bagi pembelajaran teknik *data mining* dan data analisis.