

BAB II

TINJAUAN PUSTAKA

2.1. Anemia

Anemia adalah kelainan darah yang didefinisikan sebagai penurunan *hemoglobin* tingkat, juga dinyatakan sebagai penurunan jumlah sel darah merah (*Red Blood Cells*) dan tidak sempurna nya darah untuk membawa oksigen ke seluruh tubuh secara efisien [12]. Penyakit anemia adalah kondisi hematologis, sederhana tes darah yang disebut hitung darah lengkap sangat penting untuk diagnosa. Gejala umum anemia adalah perasaan lemah atau lelah [2]. Anemia mempunyai jenis-jenis berdasarkan klasifikasi atribut yang terkait [11].

2.2. Jenis-jenis Anemia

Anemia mempunyai lima jenis yang berbeda dari segi penyebab dan penanggulangannya. Berikut adalah penjabaran jenis-jenis anemia menurut penelitian terdahulu [11] yaitu :

2.2.1. *Anemia Aplastik*

Jenis anemia ini disebabkan karena kegagalan tulang sumsum. Hal ini menyebabkan ketidak mampuan sumsum tulang dalam mengganti sel darah baru. Jenis ini bisa disebabkan oleh faktor-faktor yang mana menghancurkan sel induk atau karena perubahan sumsum tulang. Ini mengarah pada pengurangan ketiga jenis sel darah (*Pancytopenia*) sel darah

merah (anemia), darah putih sel (*leukopenia*), dan trombosit (*trombositopenia*). Istilah *aplastic* mengacu pada impotensi sel induk menghasilkan sel darah matang. Ini jarang terjadi tetapi darah serius gangguan yang menyebabkan masalah kesehatan jantung dan paru-paru. Gejala utama dari jenis ini adalah perdarahan dan infeksi.

2.2.2. Anemia Kronis

Anemia kronis adalah bentuk anemia yang diamati secara kronis infeksi, aktivasi kekebalan kronis, dan keganasan. Itu peningkatan tinggi dalam produksi *Interleukin-6* mengaktifkan produksi *hepcidin* yang melepaskan dari hati yang berkurang protein pembawa zat besi disebut *ferroportin*. Ini mengurangi akses zat besi ke sirkulasi. Mekanisme seperti berkurangnya *erythropoiesis* juga dapat berperan. Anemia kronis peradangan adalah istilah yang disukai karena tidak semua kronis penyakit terkait dengan bentuk anemia ini.

2.2.3. Iron Deficiency Anemia

Kekurangan zat besi disebut keadaan di mana jumlah zat besi yang ada dalam darah adalah menurun drastis. Anemia jenis ini biasa terjadi di antara mereka remaja dan wanita sebelum masa manopause. Kehilangan darah selama menstruasi, kehilangan darah kronis dalam kasus tukak lambung, *hernia* perdarahan *gastrointestinal hiatal*, kekurangan zat besi dalam makanan, ketidakmampuan untuk menyerap zat besi karena gangguan usus, sering donor darah dapat menyebabkan anemia defisiensi besi.

2.2.4. *Thalassemia*

Thalassemia merupakan kelainan darah bawaan pada kondisi tubuh yang membuat bentuk *hemoglobin* menjadi tidak normal. *Hemoglobin* mempunyai arti sebagai molekul protein dalam sel darah merah yang membawa oksigen. Gangguan ini mengakibatkan kerusakan pada sel darah merah yang berlebihan, sehingga menyebabkan penyakit anemia.

2.2.5. *Anemia of Renal Disease*

Anemia umumnya terjadi pada orang dengan penyakit ginjal kronis (CKD) yang penyimpangan fungsi ginjal yang permanen dan parsial. Anemia mulai berkembang pada tahap awal CKD, ketika seseorang memiliki 20 hingga 50 persen fungsi ginjal normal. Anemia cenderung memburuk saat CKD berkembang. Kebanyakan orang yang total kehilangan fungsi ginjal, atau gagal ginjal, menderita anemia.

2.3. *Data mining*

Data mining merupakan metode untuk menggali dan mengolah data mentah menjadi informasi pengetahuan yang bermakna, penggunaan metode ini biasanya ada dalam kombinasi bidang-bidang seperti intelijen bisnis, pemasaran, kesehatan, bioteknologi, multimedia dan pencarian internet [14]. *Data mining* mengacu pada proses mengekstraksi pengetahuan yang berasal dari proses *Knowledge Data Discovery* (KDD) di mana algoritma diterapkan untuk mengekstraksi pengetahuan [15]. Seiring dengan pengembangan metode, *data mining* dapat didefinisikan

sebagai proses penemuan pengetahuan [16]. Definisi *data mining* yang lebih luas adalah sebagai ekstraksi otomatis informasi berguna yang sebelumnya tidak diketahui dari set data atau database dengan menggunakan algoritma dan teknik pencarian canggih untuk menemukan korelasi dan pola [17]. Oleh karena itu, tugas *data mining* dapat diringkas sebagai tugas deskripsi dan prediksi dalam menemukan pola dan asosiasi yang dapat ditafsirkan manusia, setelah mempertimbangkan seluruh data dan membuat model prediksi yang berusaha untuk meramalkan beberapa respon yang menarik [16].

2.3.1. Teknik *Data mining*

Teknik pada *data mining* dibantu dengan pengembangan model algoritma yang dapat berupa prediksi dan deskriptif. Model prediksi digunakan contohnya untuk membuat diagnosa penyakit tertentu, teknik pada model prediksi antara lain adalah *classification*, *regression*, dan *time series analysis*. Model deskriptif digunakan untuk mengidentifikasi pola dalam data, teknik pada model deskriptif antara lain adalah *clustering*, *association rule*, dan *visualization*. Berikut ini adalah penjelasan teknik *data mining* menurut *data science journal* [18] :

A. *Classification*

Classification adalah memetakan atau mengklasifikasikan *item* data ke dalam salah satu dari beberapa kelas yang telah ditentukan. Seperangkat aturan klasifikasi dihasilkan dari model klasifikasi, berdasarkan fitur data dalam set pelatihan, yang dapat berupa digunakan untuk mengklasifikasikan

data masa depan dan mengembangkan pemahaman yang lebih baik dari masing-masing kelas dalam *database*.

B. Regression

Regression adalah metode untuk memetakan *data target* menggunakan beberapa jenis fungsi yang diketahui. Ini berkaitan dengan estimasi nilai output berdasarkan nilai input.

C. Time Series Analysis

Time series analysis adalah nilai dari atribut yang diperiksa selama periode waktu biasanya pada waktu yang sama *interval*. Teknik ini dapat digunakan untuk memprediksi nilai masa depan atau untuk menentukan kesamaan antara *interval* waktu yang berbeda.

D. Visualization Techniques

Teknik visualisasi adalah metode yang berguna untuk menemukan pola dalam kumpulan data medis. Setelah *subset* yang menarik diperoleh, kami dapat menggunakan teknik penambangan data lainnya untuk menemukan *subset* ini pengetahuan lebih lanjut.

E. Association Rule

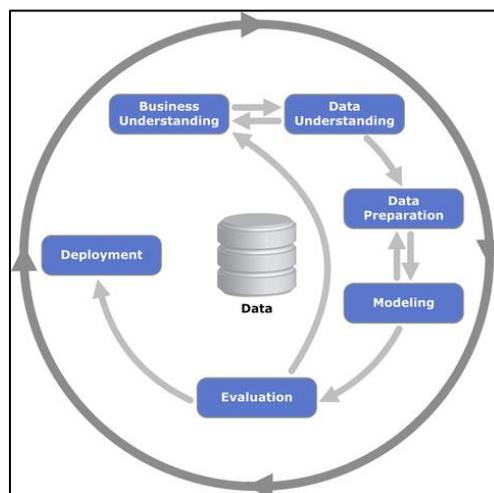
Association rule adalah penemuan asosiasi antar objek. Asosiasi mempunyai aturan tertentu seperti contohnya pada seseorang dapat menemukan bahwa satu set gejala sering terjadi bersamaan dengan serangkaian gejala lainnya.

F. Clustering

Clustering adalah identifikasi kelas atau kelompok untuk satu set objek yang tidak diklasifikasi berdasarkan atributnya. Ini adalah proses penemuan pengetahuan untuk menemukan kelompok kasus yang saling terkait dan perilaku statistik yang dihasilkan mereka mematuhi kelompok. Setelah cluster diputuskan, objek diberi label yang sesuai *cluster*, dan fitur umum dari objek dalam sebuah *cluster* diringkas untuk membentuk deskripsi kelas.

2.3.2. CRISP-DM

Cross-Industry Standard Process untuk *data mining* atau CRISP-DM yang dikembangkan berdasarkan analisis dari beberapa industri seperti Daimler Chrysler, NCR dan SPSS. CRISP-DM dapat diartikan sebagai standarisasi proses *data mining* dalam strategi pemecahan masalah secara umum dari bisnis atau unit penelitian. CRISP-DM merupakan *framework data mining* memiliki siklus hidup yang terbagi dalam enam fase [19]. Enam fase ini bersifat seperti siklus berulang dari *understanding business* sampai *deployment* seperti pada gambar 2.1.



Gambar 2.1. Tahapan Proses CRISP-DM [19]

Pada tabel 2.1 deskripsi *phase* CRISP-DM menjelaskan gagasan utama, tugas dan keluaran dari fase-fase ini secara singkat, berdasarkan panduan pengguna CRISP-DM [19].

Tabel 2.1. Deskripsi *Phase* CRISP-DM

<i>Phase</i>	<i>Description</i>
<i>Business Understanding</i>	Penentuan tujuan <i>data mining</i> merupakan salah satu aspek terpenting dalam fase ini. Pertama jenis <i>data mining</i> harus dijelaskan dan kriteria keberhasilan <i>data mining</i> . Rencana proyek wajib harus dibuat.
<i>Data Understanding</i>	Mengumpulkan data dari sumber data, mengeksplorasi dan mendeskripsikannya serta memeriksa kualitas data. Untuk membuatnya lebih konkret, panduan pengguna mendeskripsikan tugas deskripsi data dengan menggunakan analisis statistik dan menentukan atribut dan kolasinya.
<i>Data Preparation</i>	Tahapan menyiapkan dan membersihkan data. Bergantung pada model yang digunakan (didefinisikan di fase pertama) atribut turunan harus dibangun. Metode preparation setiap model akan berbeda-beda.
<i>Modeling</i>	Tahap pemodelan data terdiri dari pemilihan teknik pemodelan, pembuatan kasus uji dan model bergantung pada masalah bisnis dan data. Untuk membangun model, parameter khusus harus ditetapkan. Untuk menilai model, sebaiknya mengevaluasi model terhadap kriteria evaluasi dan memilih yang terbaik.

<i>Phase</i>	<i>Description</i>
<i>Evaluation</i>	Tahap evaluasi, hasil diperiksa terhadap tujuan bisnis yang ditetapkan. Oleh karena itu, hasil harus diinterpretasikan dan tindakan selanjutnya harus didefinisikan serta proses tersebut harus ditinjau ulang secara umum.
<i>Deployment</i>	Fase penerapan dijelaskan secara umum dalam panduan pengguna. Ini bisa berupa laporan akhir atau komponen perangkat lunak. Panduan pengguna menjelaskan bahwa fase penerapan terdiri dari perencanaan penerapan, pemantauan, dan pemeliharaan.

2.4. Naïve Bayes

Naïve Bayes mengimplementasikan pembeda *Naïve Bayes* yang probabilistik. *Naïve Bayes* menggunakan distribusi normal untuk memodelkan atribut numerik. *Naïve Bayes* dapat menggunakan penduga kerapatan *kernel*, yang mengembangkan kinerja jika asumsi normalitasnya jika benar. *Naïve Bayes Updateable* adalah versi inkremental yang memproses satu permintaan pada satu waktu. Versi ini dapat menggunakan estimator *kernel* tetapi tidak diskritisasi [8].

2.5. J48 Decision Tree

Algoritma *J48 Decision Tree* disebut sebagai implementasi C4.5 yang dioptimalkan atau versi yang ditingkatkan dari C4.5. *Output* yang diberikan

oleh *J48 Decision Tree* adalah pohon Keputusan. Pohon keputusan sama dengan struktur pohon yang memiliki simpul berbeda, seperti simpul akar, simpul tengah dan simpul daun. Setiap *node* di pohon berisi keputusan dan keputusan itu mengarah ke hasil. *Decision tree* membagi ruang input dari suatu set data ke dalam area yang saling eksklusif, di mana setiap area memiliki *label*, nilai atau tindakan untuk menggambarkan atau menguraikan poin datanya. Kriteria pemisahan digunakan dalam pohon keputusan untuk menghitung atribut mana yang terbaik untuk membagi pohon bagian dari data pelatihan yang mencapai *node* tertentu [10].

2.6. Random Forest

Random Forest adalah teknik pembelajaran ensemble yang dikembangkan oleh Breiman pada tahun 2001 berdasarkan versi asli yang diperkenalkan oleh Bell Labs pada tahun 1995. RFC adalah pengklasifikasi ansambel yang menggabungkan pohon klasifikasi yang lebih lemah yang tidak berkorelasi [10]. Pengklasifikasi RF memiliki dua manfaat utama: akurasi yang relatif tinggi dan kecepatan pemrosesan. Kekuatan dan korelasi antara pohon klasifikasi individu adalah signifikan dan akan mempengaruhi kinerja ansambel RF. Pengklasifikasi RF terdiri dari kumpulan pengklasifikasi mirip pohon yang melatih beberapa pengklasifikasi, di mana setiap pohon memberikan satu suara untuk tugas sehingga menggabungkan hasil melalui proses pemungutan suara untuk menemukan kelas yang paling populer [8]. Perhitungan performa algoritma *data mining* dapat menggunakan hasil akurasi, sensitivitas, dan presisi [20].

2.7. Akurasi

Akurasi adalah angka persentase prediksi yang benar untuk data sebuah pengujian pada klasifikasi *data mining*. Probabilitas untuk memprediksi kelas dengan benar dari *instance* yang tidak berlabel dan dapat diperkirakan dengan berbagai cara. Perhitungan akurasi dengan mudah dengan membagi jumlah prediksi yang benar dengan jumlah total prediksi [20]. Akurasi dapat disimpulkan seperti pada rumus 2.1.

$$Accuracy = \frac{\text{correct prediction}}{\text{all predictions}} \quad [20]$$

Rumus 2.1 Akurasi

2.8. Sensitivitas

Sensitivitas adalah angka tingkat keberhasilan system dalam menemukan kembali sebuah informasi. Sensitivitas atau biasa disebut dengan *recall* mengacu pada persentase total hasil yang relevan yang diklasifikasikan dengan benar oleh algoritma yang dibentuk. Proporsi kasus nyata positif yang benar diprediksi positif. Sensitivitas mengukur cakupan dari kasus positif nyata dengan +P (Prediksi Positif) [20]. Sensitivitas dapat disimpulkan seperti pada rumus 2.2.

$$Recall = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \quad [20]$$

Rumus 2.2. Sensitivitas

2.9. Presisi

Presisi adalah angka yang menunjukkan proporsi prediksi kasus-kasus positif yang menyatakan benar positif. Presisi atau biasa disebut dengan *confidence* merupakan bagian dari contoh yang relevan (positif benar) di antara semua contoh yang diperkirakan termasuk dalam kelas tertentu [20]. Presisi dapat disimpulkan seperti pada rumus 2.3.

$$Precision = \frac{true\ positives}{true\ positives + false\ positives} \quad [20]$$

Rumus 2.3. Presisi

2.10. RapidMiner

RapidMiner adalah perangkat lunak sumber terbuka yang menyediakan lingkungan yang baik untuk proses penambangan data. *RapidMiner* memiliki fasilitas *drag-and-drop* yang digunakan untuk membangun aliran data dan memfasilitasi berbagai *format file*. Tugas-tugas regresi, klasifikasi dan pengelompokan dapat dilakukan dengan mudah dengan algoritma pembelajaran yang berbeda. *RapidMiner* mendukung sejumlah besar algoritma klasifikasi dan regresi, pohon keputusan, aturan asosiasi, algoritma pengelompokan, dan banyak fitur yang tersedia untuk *preprocessing* data, normalisasi, *filtering*, dan analisis data. Itu dapat mengimpor data dari *database* tradisional dan standar yang berbeda [21].

2.11. Penelitian Terdahulu

Berikut adalah tabel 2.2 perbandingan penelitian terdahulu yang berisikan rangkuman dari bahasan mengenai *classification* pada prediksi penyakit anemia dan teknik *data mining* :

Tabel 2.2. Perbandingan Penelitian Terdahulu

Judul Jurnal / Vol / No	Penulis / Tahun	Source	Algoritma Data mining / Tools	Hasil dan Simpulan
Anemia types prediction based on <i>data mining</i> classification algorithms / vol.1 / no.9	Manal Abdullah, Salma Al-Asmari / 2017	Research Gate	- <i>Naive Bayes</i> - <i>J48 Decision Tree</i> - SVM	Penelitian menggunakan 41 data, dengan pembagian data sebesar 60%, <i>naive bayes</i> mempunyai akurasi 68.75%, <i>J48 Decision Tree</i> 93.75%, SVM 93.75%, menggunakan tools WEKA.
Using classification techniques for statistical analysis of Anemia / vol.94 / no.10	Kanak Meena, Devendra, Vaidehi, Aiman / 2019	Science Direct	- <i>Decision Tree</i>	Akurasi 97% hanya dengan identifikasi hemoglobin, dan 44% hubungan antara ibu dan anak, menggunakan tools R.
Machine Learning Algorithms for Anemia Disease Prediction / vol. 524 / no.13	Manish Jaiswal, Anima Srivastava and Tanveer J. Siddiqui / 2019	Springer Professional	- <i>Naive Bayes</i> - <i>Decision Tree</i> - <i>Random Forest</i>	Penelitian menggunakan 200 data pasien. Hasil akurasi dari <i>Naive Bayes</i> sebesar 96%, <i>Random Forest</i> sebesar 95%, <i>Decision Tree</i> sebesar 95,5%, menggunakan tools WEKA.

Judul Jurnal / Vol / No	Penulis / Tahun	Source	Algoritma Data mining / Tools	Hasil dan Simpulan
Comparison of Different Classification Techniques Using WEKA for Hematological Data / vol. 4 / no. 3	Md. Nurul Amin, Md. Ahsan Habib / 2016	American Journal of Engineering Research (AJER)	- <i>J48</i> <i>Decision Tree</i> - <i>Naïve Bayes</i>	Hasil akurasi dari <i>J48 Decision Tree</i> sebesar 94%, <i>Naïve Bayes</i> sebesar 70%, menggunakan tools WEKA.
Perancangan dan Pembuatan Visualisasi Data Dana Penelitian Internal dan Hibah Dikti LPPM Universitas Multimedia Nusantara / vol.10 / no. 1	Wenni Indriani Loka, Friska Natalia	Ultima InfoSys: Jurnal Ilmu Sistem Informasi	-	Hasil penelitian berupa 5 <i>dashboard</i> visualisasi dengan menggunakan metode CRISP-DM, data yang pakai tahun 2012-2017.