

BAB II

TINJAUAN PUSTAKA

2.1. Implementasi

Menurut Kamus Besar Bahasa Indonesia (KBBI) pengertian implementasi ialah kegiatan pelaksanaan rencana yang disusun dengan cermat dan rinci. Implementasi dapat didefinisikan sebagai proses mengubah rencana menjadi kenyataan.

2.2. Tolak Ukur

Tolak ukur artinya sesuatu yang digunakan sebagai dasar untuk mengukur atau menilai, dan sebagainya. Atau tolak ukur sebagai patokan untuk pengambilan keputusan.

2.3. Bank Syariah

Bank syariah adalah bank yang beroperasi dengan tidak mengandalkan pada bunga. Bank Islam atau biasa disebut dengan bank tanpa bunga adalah lembaga keuangan/perbankan yang operasional dan produknya dikembangkan berdasarkan pada al-qur'an dan hadist. Dengan kata lain bank Islam adalah lembaga keuangan yang usaha pokoknya memberikan pembiayaan dan jasa-jasa lainnya dalam lalu lintas pembayaran serta peredaran uang yang pengoperasiannya disesuaikan dengan prinsip-prinsip syariat Islam. Dalam kegiatan penyaluran dana bank syariah melakukan investasi dan pembiayaan, disebut investasi karena prinsip yang digunakan adalah prinsip penanaman dana atau penyertaan, dan keuntungan akan diperoleh bergantung pada kinerja usaha yang menjadi obyek penyertaan tersebut sesuai dengan nisbah bagi hasil yang diperjanjikan sebelumnya, disebut pembiayaan karena bank syariah menyediakan dana guna membiayai kebutuhan nasabah yang memerlukannya dan layak memperolehnya (Ilyas 2019).

2.4. Kredit

Pengertian kredit adalah kemampuan untuk melaksanakan pembelian atau pinjaman dengan janji pembayaran akan dilaksanakan pada jangka waktu yang telah disepakati.

2.5. Data Mining

Data mining ialah istilah yang digunakan untuk menggambarkan penemuan pengetahuan dari dalam basis data. Data mining adalah teknik yang memanfaatkan fakta, aritmatika, kecerdasan buatan, dan *machine learning* untuk memperoleh pengetahuan tentang teknik untuk mengekstraksi dan memilih informasi yang berguna dan pengetahuan terkait dari *database* besar. Pengetahuan ini dapat digunakan sebagai acuan untuk mengambil keputusan dalam bisnis perusahaan (Hidayati, Nugroho dan Nurjoko 2021). Faktor-faktor yang menjadi alasan data mining dilakukan ialah :

1. Data saat ini telah mencapai jangkauan dan ukuran yang besar.

Teknik data mining menghasilkan informasi dengan cara mendasari pengambilan keputusan sehingga fakta dan tingkat akurasi sangat signifikan. Data yang besar dan semakin banyak dapat meningkatkan hasil validasi.

2. Prosedur data *warehouse* telah dilakukan.

Sumber data dalam data mining merupakan kumpulan data dari berbagai perangkat/bidang. Dalam menjaga konsistensi data untuk memberikan hasil yang tepat direkomendasikan untuk menggunakan data yang telah melewati tahap *warehouse* untuk menegakkan integritas dan perspektif catatan yang tepat.

3. Komputasi data yang terjangkau.

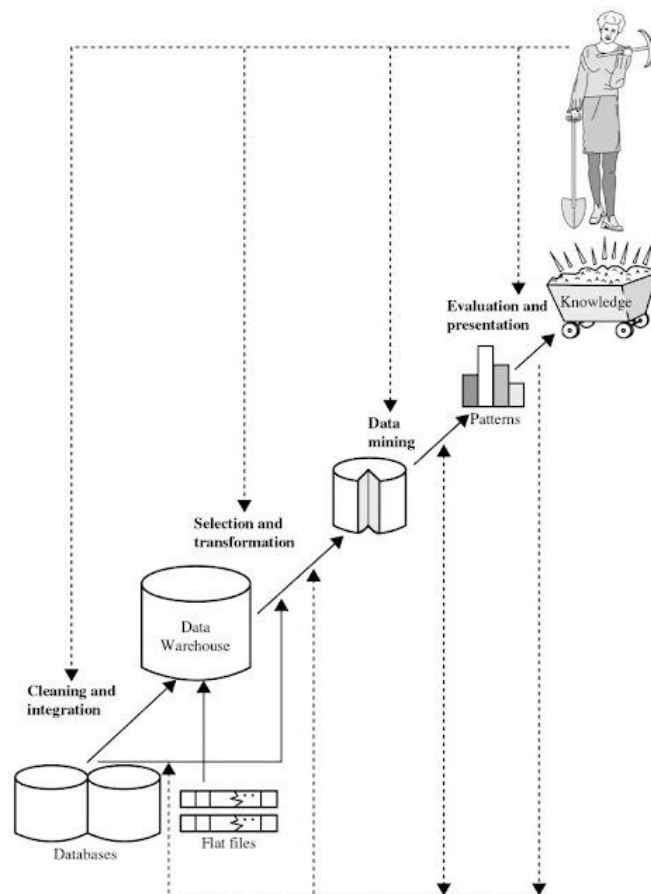
Proses data mining membutuhkan akses yang cukup banyak. Proses komputasi membutuhkan sumber daya yang sangat besar. Dengan perkembangan teknologi yang berkembang saat ini, sangat memungkinkan untuk dilakukan fasilitas data mining.

4. Ada persaingan bisnis yang ketat.

Persaingan yang ketat membuat organisasi dituntut untuk berinovasi dengan tujuan untuk bersaing dalam teknologi internasional. Situasi ini memungkinkan data mining dilakukan untuk menemukan tren dan fakta dari produk yang dimiliki oleh setiap organisasi (Ginantra dan Dkk 2021).

2.6. Tahapan Data Mining

Beberapa tahapan proses data mining yang diilustrasikan pada gambar 2.1.



Gambar 2.1 Tahapan Data Mining

Tahapan data mining sebagai berikut :

1. Pembersihan data

Pada umumnya data yang diperoleh, baik dari *database* suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti

data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil data mining nantinya. *Garbage in garbage out* (hanya sampah yang akan dihasilkan bila yang dimasukkan juga sampah) merupakan istilah yang sering dipakai untuk menggambarkan tahap ini. Pembersihan data juga akan mempengaruhi performansi dari sistem data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

2. Integrasi data

Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dsb. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada. Dalam integrasi data ini juga perlu dilakukan transformasi dan pembersihan data karena seringkali data dari dua database berbeda tidak sama cara penulisannya atau bahkan data yang ada di satu database ternyata tidak ada di database lainnya.

3. Transformasi data

Beberapa teknik data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa teknik standar seperti analisis asosiasi dan klustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut binning. Disini juga dilakukan pemilihan data yang diperlukan oleh teknik data mining yang dipakai. Transformasi dan pemilihan data ini juga menentukan kualitas dari hasil data mining nantinya karena ada beberapa karakteristik dari teknik-teknik data mining tertentu yang tergantung pada tahapan ini.

4. Aplikasi data Teknik Mining

Aplikasi teknik data mining sendiri hanya merupakan salah satu bagian dari proses data mining. Ada beberapa teknik data mining yang sudah umum dipakai. Akan dibahas lebih jauh mengenai teknik-teknik yang ada di seksi berikutnya. Perlu diperhatikan bahwa ada kalanya teknik-teknik data mining umum yang tersedia di pasar tidak mencukupi untuk melaksanakan data mining di bidang tertentu atau untuk data tertentu. Sebagai contoh akhir-akhir ini dikembangkan berbagai teknik data mining baru untuk penerapan di bidang bioinformatika seperti analisa hasil *microarray* untuk mengidentifikasi DNA dan fungsi-fungsinya.

5. Evaluasi pola yang di temukan

Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti : menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba teknik data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

6. Presentasi pola yang ditemukan

Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining (Sudarsono, et al. 2021).

2.7. Decision Tree (Pohon Keputusan)

Salah satu teknik klasifikasi yang paling umum digunakan dalam proses data mining adalah *decision tree* (pohon keputusan). Pohon keputusan adalah metode klasifikasi yang menggunakan struktur pohon dimana setiap simpul

mewakili atribut, cabang mewakili nilai atribut, dan daun mewakili kelas. *Node* teratas dari pohon keputusan disebut sebagai *root*. Pohon keputusan adalah metode yang sangat kuat dan terkenal untuk klasifikasi dan prediksi. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang mewakili aturan. Aturan mudah dipahami dalam bahasa alami. Itu juga dapat diekspresikan dalam bahasa *database* seperti *structured query language* untuk menemukan catatan dari kategori tertentu. Pohon keputusan juga berguna untuk mengeksplorasi data dan menemukan hubungan tersembunyi antara sekumpulan kandidat variabel *input* dan variabel target (Partogi, Pasaribu dan Sutrisno 2021).

2.8. Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 (*Iterative Dichotomiser 3*) yang dikembangkan oleh J. Ross Quinlan. Idenya adalah membuat pohon dengan percabangan awal merupakan atribut yang paling signifikan, dan seterusnya dibagi menjadi percabangan-percabangan sampai aturan terpenuhi. Pohon keputusan itu sendiri diartikan sebagai suatu cara untuk membagi sekumpulan data menjadi himpunan-himpunan yang lebih kecil dengan menerapkan serangkaian *rule* atau aturan keputusan. Algoritma IDS ini pada dasarnya hanya mengulang langkah pemartisian sampai diperoleh keadaan dimana semua sampel pada sebuah simpul tergolong ke dalam kelas yang sama. Setiap jalur dari akar menuju daun akan merepresentasikan aturan keputusan (*decision rule*) yang akan digunakan sebagai prediktor kelas data berikutnya. Pohon dibangun dengan cara membagi data secara rekursif sehingga tiap bagian terdiri dari data yang berasal dari kelas yang sama. Bentuk pemecahan (*split*) yang digunakan untuk membagi data tergantung dari jenis atribut yang digunakan dalam *split*. Algoritma C4.5 dapat menangani data numerik baik yang bersifat diskrit maupun kontinyu. Jika suatu set data mempunyai beberapa nilai pengamatan *missing value*, jika jumlah pengamatan terbatas maka atribut *missing value* ini dapat diganti dengan nilai rata-rata dari variabel yang bersangkutan. Pada algoritma ini, pemilihan atribut mana yang

akan menempati suatu simpul dilakukan dengan melakukan perhitungan entropi informasi (*information entropy*) dan mencari nilai yang paling minimum. Pemilihan atribut pada algoritma ini berdasarkan asumsi bahwa kompleksitas yang dimiliki oleh pohon keputusan sangat berkaitan erat dengan jumlah informasi yang diberikan oleh nilai-nilai atributnya. Dengan kata lain, atribut dipilih berdasarkan perolehan informasi terbesar (*highest information gain*) dalam menghasilkan subpohon (*subtree*) (Junaedi, Nuswantari dan Yasin 2019).

Algoritma C4.5 didasarkan pada pembuatan pohon keputusan yang dipangkas dalam dua tahap dengan menambahkan ID3 ke atribut kontinu, nilai atribut, dan pemrosesan informasi untuk menghasilkan pohon keputusan. Dalam menggunakan algoritma C4.5 kita dapat menemukan tingkat perolehan informasi *gain ratio*. Untuk membuat pohon keputusan dipilih nilai atribut yang memiliki *gain ratio* tertinggi. Rumus perhitungan algoritma C4.5 adalah sebagai berikut :

- a. Sebelumnya, nilai indeks entropi dihitung menggunakan rumus:

$$Entropy(i) = -\sum_{j=1}^m f(i, j) \cdot \log_2[f(i, j)]$$

Gambar 2.2 Rumus *Entropy*

- b. Nilai gain dengan rumus:

$$gain = -\sum_{i=1}^p \dots \cdot IE(i)$$

Gambar 2.3 Rumus *Gain*

- c. Untuk menghitung gain ratio perlu diketahui suatu term baru yang disebut Split Information dengan rumus:

$$SplitInformation = -\sum_{t=1}^c \frac{s_t}{s} \log_2 \frac{s_t}{s}$$

Gambar 2.4 Rumus *Split Info*

d. Selanjutnya menghitung gain ratio

$$\text{Gainratio}(S, A) = \frac{\text{Gain}(S,A)}{\text{SplitInformation}(S,A)}$$

Gambar 2.5 Rumus *Gain Ratio*

e. Ulangi langkah a-d sampai semua record telah terpecah. Proses pemisahan pohon keputusan berakhir ketika:

1. Semua tupel dalam catatan simpul m adalah kelas yang sama.
2. Atribut dalam dataset tidak dapat dibagi lagi.
3. Cabang kosong tidak memiliki catatan.

(Purwanto dan Nugroho 2023)

2.9. Rapidminer

Rapidminer adalah aplikasi data mining berbasis open-source yang terkemuka dan ternama. Didalamnya terdapat aplikasi yang berdiri sendiri untuk analisis data dan sebagai mesin data mining seperti untuk loading data, transformasi data, pemodelan data, dan metode visualisasi data. Rapidminer awalnya bernama *Yet Another Learning Environment* atau disingkat YALE. Nama perusahaan berubah menjadi Rapidminer pada tahun 2007 (Nofitri dan Irawati 2019).

2.10. Penelitian Terdahulu

Dalam melakukan penelitian ini terdapat penelitian-penelitian terdahulu yang dijadikan sebagai bahan referensi bagi peneliti. Berikut penelitian terdahulu yang peneliti gunakan sebagai referensi, dan penelitian terdahulu tersebut dapat dilihat dalam bentuk tabel 2.1.

Tabel 2.1 Penelitian Terdahulu

No	Peneliti	Judul	Metode	Kesimpulan
1	Yusfina Susanti Ripka Igo, Abdul Aziz, dan Moh. Ahsan	Klasifikasi Kelayakan Pemberian Kredit Nasabah Bank XYZ Menggunakan Metode Algoritma C4.5 Dan Naive Bayes	Algoritma C4.5 dan Naive Bayes	Metode algoritma C4.5 lebih cocok untuk pengklasifikasian data kredit nasabah dengan hasil akurasi sebesar 67,70% , sedangkan hasil akurasi menggunakan metode naive bayes sebesar 66,67% .
2	Dyah Wulandari, Nur Lutfiyana, dan Heny Sumarno	Metode Algoritma <i>Decision Tree</i> Untuk Analisis Kelayakan Kredit Nasabah Pada BSM KCP Kemang Pratama	Algoritma C4.5	Mendapatkan sebuah pola yaitu dari faktor plafon dan jangka waktu yang sangat berpengaruh sekali dalam menentukan kelayakan nasabah menerima kredit dengan menghasilkan nilai akurasi sebesar 80% sehingga perusahaan dapat meminimalisir resiko yang akan terjadi.
3	Teguh Budi Santoso dan Dela Sekardiana	Penerapan Algoritma C4.5 Untuk Penentuan Kelayakan Pemberian Kredit (Studi Kasus : Koperia – Koperasi Warga Komplek Gandaria)	Algoritma C4.5	Algoritma C4.5 terbukti akurat dalam penentuan pengajuan kelayakan kredit, ini dibuktikan dengan nilai akurasi mencapai 93% .

Penelitian terdahulu diatas digunakan karena merujuk pada penelitian ini. Pada penelitian terdahulu untuk penentuan kelayakan pemberian kredit pada nasabah metode yang digunakan ialah algoritma C4.5. Algoritma C4.5 menghasilkan akurasi yang baik untuk pengklasifikasian data kredit nasabah, sehingga perusahaan dapat meminimalisir resiko kredit macet. Maka dari itu pada penelitian ini peneliti menggunakan algoritma C4.5 sebagai tolak ukur pemberian kredit.