

BAB II LANDASAN TEORI

2.1 Tinjauan Pustaka

Dalam melakukan penelitian ini peneliti mengambil tinjauan pustaka sebagai salah satu alat penerapan metode penelitian. Tinjauan Pustaka ini juga bertujuan untuk membandingkan hasil penelitian yang satu dengan yang lain. Berikut Daftar Literatur penelitian dapat dilihat pada tabel 2.1 dibawah ini

Tabel 2. 1 Daftar Literatur

| No | Nama Peneliti | Tahun | Judul |
|----|-------------------------------|-------|--|
| 1 | (Muhamad, et al., 2019) | 2019 | Penerapan Algoritma <i>C4.5</i> Pada Klasifikasi Potensi Siswa Drop Out |
| 2 | (Siti Badriah1, et al., 2021) | 2021 | Klasifikasi Algoritma <i>C4.5</i> dalam Menentukan Penerima Bantuan Covid-19 (Studi Kasus: Desa di Karawang) |
| 4 | (Asroni, et al., 2018) | 2018 | Penerapan Algoritma <i>C4.5</i> untuk Klasifikasi Jenis Pekerjaan Alumni di Universitas Muhammadiyah Yogyakarta |
| 5 | (Nugroho Setio, et al., 2020) | 2020 | Klasifikasi dengan Pohon Keputusan Berbasis Algoritme <i>C4.5</i> |
| 6 | (Arifin, et al., 2018) | 2018 | Penerapan Algoritma Klasifikasi <i>C4.5</i> dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus: PT Atria Artha Persada |

| | | | |
|----|-----------------------------|------|---|
| 7 | (Massulloh, et al., 2020) | 2020 | Implementasi Algoritma <i>C4.5</i> Untuk Klasifikasi Anak Berkebutuhan Khusus Di Ibnu Sina Stimulasi Center |
| 8 | (Saputra, et al.) | 2020 | Klasifikasi Algoritma <i>C4.5</i> Dalam Penerapan Tingkat Kepuasan Siswa Terhadap Media Pembelajaran Online |
| 9 | (Ningse, et al., 2022) | 2022 | Klasifikasi Algoritma <i>C4.5</i> untuk Penentuan Penerima Program Indonesia Pintar pada MIS Al-Khoirot |
| 10 | (Azwanti, 2018) | 2018 | Nalisa Algoritma <i>C4.5</i> Untuk Memprediksi Penjualan Motor Pada Pt. Capella Dinamik Nusantara Cabang Muka Kuning |
| 11 | (Azhari, et al., 2022) | 2022 | Penerapan Algoritma <i>C4.5</i> Untuk Klasifikasi Tingkat Pemahaman Siswa Pada Mata Pelajaran Matematika |
| 12 | (Telaumbanua, et al., 2022) | 2022 | Penerapan Algoritma <i>C4.5</i> Untuk Klasifikasi Kepuasan Pelanggan Pada Jasa Layanan Pengiriman |
| 13 | (Sepharni, et al., 2022) | 2022 | Klasifikasi Penyakit Jantung Dengan Menggunakan Algoritma <i>C4.5</i> |
| 14 | (Pambudi, et al., 2018) | 2018 | Penerapan Algoritma <i>C4.5</i> Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal |

| | | | |
|----|------------------------|------|--|
| 15 | (Taufik, et al., 2021) | 2021 | Penerapan Algoritma <i>C4.5</i> Untuk Klasifikasi Keberhasilan Pengiriman Barang |
| 16 | (Islam, et al., 2022) | 2022 | Penerapan Algoritma <i>C4.5</i> dalam Klasifikasi Status Gizi Balita |

2.1.1 Literatur 1

Siswa merupakan salah satu substansi yang perlu diperhatikan dalam kaitannya dengan dunia pendidikan saat ini. Sulitnya mendapatkan siswa membuat sekolah harus mengoptimalkan sistem pembelajaran dan juga sarana prasarana untuk menarik minat calon siswa baru dan juga membuat siswa yang sudah bersekolah di sekolah tersebut tidak putus sekolah atau dropout. Salah satu faktor penyebab banyaknya jumlah siswa dropout karena kurangnya kebijakan dan tindakan dari instansi pendidikan untuk menjaga siswanya dari dropout. Tujuan Penelitian ini adalah membuat klasifikasi siswa berpotensi dropout dan tidak berpotensi dropout dengan algoritma *C4.5* sebagai acuan dalam membuat kebijakan dan tindakan untuk mengurangi jumlah siswa dropout. Hasil klasifikasi dari algoritma *C4.5* dievaluasi dan divalidasi dengan RapidMinerStudio untuk mengetahui tingkat akurasi Algoritma *C4.5* dalam membuat klasifikasi siswa berpotensi dropout.

2.1.2 Literatur 2

Covid-19 menyebabkan masyarakat Indonesia mengalami penurunan dalam pendapatan penghasilan. Dalam menanggulangi masalah penurunan pendapatan tersebut pemerintah memberikan bantuan bagi masyarakat akibat terdampak Covid-19. Berhubungan dengan adanya bantuan Covid-19 tersebut, staf pemerintah desa kesulitan dalam melakukan proses seleksi penerima bantuan dan adanya ketidaktepatan dalam penerima bantuan Covid-

19. Tujuan penelitian ini adalah membuat klasifikasi penerima bantuan sosial Covid-19 sebagai bahan acuan dalam kebijakan penentuan penerima bantuan sosial. Sehingga dilakukan penelitian untuk mempermudah dalam penentuan dan ketepatan dalam bantuan Covid-19. Algoritma klasifikasi yang digunakan yaitu algoritma *C4.5* dengan bahasa pemrograman php. Data yang digunakan terdapat 430 data dengan 5 atribut yaitu jumlah anggota keluarga, pekerjaan, penghasilan sebelum Covid-19, penghasilan setelah Covid-19 dan keputusan (data per bulan April 2020). Hasil dari penelitian ini didapat rule dari algoritma *C4.5* dengan dengan nilai akurasi yang terbaik terdapat pada perbandingan 90% data *training* dan 10% data *testing* dengan nilai akurasi sebesar 79,54%.

2.1.3 Literatur 3

Perkembangan pendidikan di Indonesia mengalami peningkatan yang sangat pesat. Salah satu hal yang menjadi tolok ukur keberhasilan mutu pendidikan di perguruan tinggi adalah jenis pekerjaan yang diperoleh lulusan setelah lulus. Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang mempengaruhi jenis metode klasifikasi pekerjaan berdasarkan algoritma alumni *C 4.5*. Metodologi penelitian ini dimulai dengan studi literatur, identifikasi proses ekstraksi data, seleksi data, pengumpulan data, pengolahan data, pengujian data, dan penarikan kesimpulan DA. Penelitian ini menggunakan beberapa fitur data beberapa dosen, tahun kelulusan, tingkat penyelesaian tahunan, dan kekuatan sebagai parameter kinerja klasifikasi. Data lulusan yang digunakan sebanyak 259, dan terdiri dari 3 Fakultas Ekonomi, Kedokteran dan Teknik angkatan tahun 2001-2013 dan lulus pada tahun 2011-2016. Hasil penelitian yang telah dilakukan adalah jika berasal dari Fakultas Ekonomi maka pada tahun 2011 dan 2012 sebagian besar pekerjaan di swasta telah lulus, jika berasal dari Fakultas Kedokteran dengan tahun 2011 dan 2012 lulus dengan kumulatif tingkat tenaga kerja antara 3 hingga 3,5 mayoritas bekerja di Sektor Swasta, 2012 dengan IPK antara 3 dan 3,5 bekerja di Sektor Swasta. Terakhir, algoritma *C4.5* cocok untuk klasifikasi jenis pekerjaan alumni.

2.1.4 Literatur 4

Data mining merupakan suatu proses menemukan pola dalam sejumlah data besar dengan tujuan untuk melakukan klasifikasi, estimasi, prediksi, asosiasi dan kluster. Salah satu metode klasifikasi yang populer adalah pohon keputusan (*decision tree*). Konsep dasar dari pohon keputusan adalah mengubah data menjadi sebuah model pohon keputusan, kemudian mengubah model pohon menjadi rule dan menyederhanakan rule. Data dalam pohon keputusan dinyatakan dalam bentuk tabel dengan atribut dan record. Terdapat beberapa cara untuk mengkonstruksikan pohon keputusan salah satunya menggunakan algoritme *C4.5*. Algoritme *C4.5* merupakan pengembangan dari *Iterative Dichotomiser (ID3)*. *ID3* adalah algoritma pembelajaran pohon keputusan (*decision tree learning*) yang paling dasar. Algoritme ini melakukan pencarian secara menyeluruh (*greedy*) pada semua kemungkinan pohon keputusan. Algoritme *ID3* membangun pohon keputusan secara *top-down* (dari atas ke bawah) yang diawali dengan suatu atribut. *Top-down* artinya pohon keputusan dibangun dari simpul akar ke daun. Algoritme ini tidak dapat mengatasi atribut dengan data kosong, oleh karena itu diperlukan algoritme *C4.5* yang dapat mengatasi atribut dengan data kosong serta dapat melakukan pemangkasan cabang yang juga tidak terdapat pada algoritma *ID3*.

2.1.5 Literatur 5

Permasalahan yang sering muncul dalam bisnis pada penjualan dengan sistem pembayaran kredit tempo adalah antara lain seperti kredit macet, order fiktif dan penipuan. Jika tidak menggunakan prosedur yang benar dalam penerimaan mitra penjualan bukan tidak mungkin masalah tersebut akan timbul dalam proses bisnis. Untuk mengembangkan bisnis tentu perlu meningkatkan penjualan dan dibarengi dengan pembayaran yang lancar. Oleh karena itu, pada proses penerimaan mitra perlu dibuatkan prosedur yang baik agar dapat meminimalisir kemungkinan masalah yang ada. Pada prosedur penerimaan mitra penjualan perlu dilakukan proses analisa untuk menentukan calon mitra penjualan tersebut diterima atau

ditolak. Dari data mitra yang ada dapat digunakan untuk dilakukan pengolahan data. Salah satu teknik pengolahan data yang dapat digunakan dalam proses tersebut adalah klasifikasi. Klasifikasi adalah teknik pengolahan data yang membagi objek menjadi beberapa kelas sesuai dengan jumlah kelas yang diinginkan. Dan menggunakan Algoritma *C4.5* dalam proses pengklasifikasi-annya. Yang dapat menentukan apakah calon mitra dapat diterima sebagai mitra atau tidak. Kemudian, Hasil dari pengklasifikasiannya divalidasi dengan *ten-fold cross validation* dengan tingkat akurasi 96,26 %, presisi 100% dan recall 71,43%.

2.1.6 Literatur 6

ABK (Anak Berkebutuhan Khusus) merupakan anak yang secara signifikan berbeda dengan anak-anak lain, mereka yang secara fisik, psikologi, kognitif atau secara sosial terhambat dalam mencapai tujuan dan potensi mereka secara maksimal seperti gangguan bicara, cacat, reterdasi mental dan gangguan emosional. Klasifikasi anak tunagrahita meliputi anak tunagrahita ringan, sedang dan berat. Tunagrahita meliputi berbagai tingkatan dari yang ringan sampai yang berat. Karena itu mereka berbeda satu dengan yang lain dalam keterlambatan perkembangan dan kondisinya sehingga berbeda dalam strategi pendidikan dan pengajaran yang dirancang dan diprogramkan. Oleh karena itu, diperlukan cara atau metode untuk memberikan informasi kepada guru dan para orang tua dalam membedakan atau mengklasifikasikan anak berkebutuhan khusus (*tunagrahita*) berdasarkan ciri-ciri yang ada pada anak. Algoritma *C4.5* merupakan salah satu metode *data mining* untuk memprediksi dan mengetahui tingkat akurasi jenis tunagrahita berdasarkan ciri-ciri yang ada pada anak. Dengan menggunakan aplikasi *RapidMiner* dan pengujian model menggunakan *Cross Validation*, evaluasi dengan *Confusion Matrix* sehingga dihasilkan akurasi dari metode tersebut. Berdasarkan hasil pengujian dengan metode *cross validation* pada aplikasi *RapidMiner* menghasilkan nilai akurasi sebesar 96.18%.

2.1.7 Lieratur 7

Penelitian ini bertujuan untuk mengetahui pengaruh tingkat kepuasan siswa terhadap media pembelajaran online di SMA YPI Swasta Dharma Budi. Media internet memiliki peranan penting untuk belajar bagi kalangan siswa. Populasi penelitian ini menerapkan Algoritma *C4.5* dalam penerapan tingkat kepuasan siswa terhadap media pembelajaran online. Jenis penelitian ini yaitu Klasifikasi dengan konsep *data mining* dengan melibatkan sebanyak 100 data siswa di SMA YPI Swasta Dharma Budi yang dikategorikan dengan : Puas dan Tidak Puas. Dalam pengambilan data menggunakan kuesioner dan variabel yang digunakan dalam penelitian ini ada empat yaitu minat belajar, cara belajar, sajian pembelajaran, media pembelajaran. Metode yang digunakan dalam penelitian ini yakni metode klasifikasi.

2.1.8 Literatur 8

Tujuan penelitian adalah untuk membantu pihak sekolah dalam menyeleksi data siswa sebagai penerima PIP (Program Indonesia Pintar) menjadi lebih objektif dan praktis serta dapat membantu dalam meningkatkan ketepatan sasaran dari penerima bantuan dana PIP tersebut. Pada penelitian ini menggunakan teknik *data mining* dengan menggunakan algoritma *C4.5*. Sumber data penelitian yang digunakan di peroleh dari observasi dan wawancara kepada pihak sekolah MIS Al-Khoirot Tambun Nabolon Pematang Siantar. Adapun variabel penelitian yang digunakan adalah pekerjaan orang tua, penghasilan orang tua, pemegang KKS (Kartu Keluarga Sejahtera), pemegang SKTM (Surat Keterangan Miskin). Pada penelitian ini alternatif yang digunakan sebagai sampel adalah data siswa MIS Al-Khoirot. Hasil dari penelitian ini didapatkan bahwa atribut yang paling dominan adalah pemegang SKTM dengan *gain* sebesar 0,833764907, selain itu penelitian ini menghasilkan 8 (delapan) *rules* dengan tingkat akurasi sebesar 98,00%. Berdasarkan hal ini dapat disimpulkan bahwa algoritma *C4.5* dapat digunakan untuk klasifikasi Penentuan Penerima Program Indonesia Pintar pada MIS Al-Khoirot.

2.1.9 Literatur 9

Motor menjadi pilihan setiap orang sebagai andalan transportasi karena dari harga yang terjangkau serta dapat digunakan dalam jangka waktu yang lama. Di Indonesia sendiri, motor yang paling banyak diminati adalah motor merk Honda. Tercatat sebagai pemimpin pasar, dengan kontribusi 72,28 persen dan total penjualan sebesar 2.362.047 unit pada tahun 2016 (menurut Asosiasi Industri Sepeda Motor Indonesia). Untuk wilayah kota Batam perusahaan yang menjual motor Honda salah satunya adalah PT. Capella Dinamik Nusantara Cabang Muka Kuning. Wilayah penjualan sangat mempengaruhi terhadap penjualan motor, selain ke *showroom* langsung, dibeberapa tempat ramai seperti *mall* bisa menjadi promosi dan penjualan motor karena mudah dijangkau oleh masyarakat. Jumlah data yang sangat banyak akan sangat sulit untuk dianalisa. Analisa diperlukan untuk melihat pola dari data penjualan sehingga dapat menghasilkan prediksi penjualan motor yang nantinya akan berguna untuk pendistribusian motor dibeberapa wilayah. Dari data konsumen yang begitu banyak, maka dilakukan *data mining* dengan menggunakan algoritma *C4.5*. Hasil dari kegiatan mining ini diharapkan dapat memberikan sebuah pohon keputusan untuk melihat pola prediksi perilaku konsumen membeli motor.

2.1.10 Literatur 10

Mata pelajaran matematika merupakan salah satu pelajaran yang penting bagi siswa yang mempunyai peran cukup besar dalam dunia pendidikan. Matematika merupakan ilmu yang perlu dipahami setiap orang, terutama siswa yang masih berada pada jenjang pendidikan. Kemampuan pemahaman matematika dapat mencapai tujuan pembelajarannya apabila dapat memahami konsep dengan baik. Penelitian ini bertujuan untuk mengklasifikasikan konsep pemahaman siswa pada mata pelajaran matematika siswa SMA Swasta Tamansiswa Tapian Dolok. Dalam dunia pendidikan matematika sering dianggap sebagai mata pelajaran yang sulit

untuk dipahami karena matematika selalu berhubungan dengan angka, rumus, dan hitung-menghitung. Banyaknya rumus dan konsep yang harus dipahami, membuat siswa kesulitan dalam menyelesaikan persoalan pada materi yang ada. Sumber data diperoleh dari hasil kuesioner pemahaman matematika siswa SMA Swasta Tamansiswa Tapan Dolok. Metode yang digunakan dalam penelitian ini yaitu *data mining* dengan Algoritma *C4.5* dan dibantu dengan *software RapidMiner*. *Data mining* merupakan proses yang digunakan untuk mengidentifikasi informasi yang bermanfaat dari berbagai *database* berjumlah besar. Algoritma *C4.5* merupakan algoritma yang digunakan untuk membentuk *decision tree* berdasarkan *training data*. Atribut yang digunakan yaitu minat belajar siswa, komunikasi, cara belajar siswa, suasana pembelajaran, media pembelajaran, dan cara mengajar guru. Hasil Perhitungan nilai *entropy* dan *gain* diperoleh 18 *rules* keputusan pemahaman siswa pada mata pelajaran matematika dengan 9 *rules* berstatus paham dan 9 *rules* berstatus tidak paham. Pemodelan klasifikasi dengan Algoritma *C4.5* pada Rapidminer diperoleh akurasi sebesar 95.19%.

2.1.11 Literatur 11

Pelayanan yang berkualitas akan memberikan kepuasan dan meningkatkan loyalitas bagi para pelanggan, bagi perusahaan kualitas layanan menjadi salah satu indikator untuk dapat bersaing dan mempertahankan eksistensi perusahaan khususnya pada jasa layanan pengiriman barang yang semakin berkembang. Klasifikasi kepuasan pelanggan atas pelayanan yang diberikan oleh jasa layanan pengiriman menggunakan algoritma *C4.5* dengan konsep dasar dari pohon keputusan yaitu mengubah data menjadi sebuah model pohon keputusan, menjadi rule dan menyederhanakan rule. Penelitian ini bertujuan untuk membentuk model klasifikasi pohon keputusan dengan menganalisa kepuasan pelanggan PT. JNE Agen Arcadia. Hasil dari penelitian ini diketahui bahwa indikator Bukti Fisik memiliki nilai *gain* tertinggi yaitu 298 yang berarti pelanggan merasa puas terhadap indikator tersebut, sedangkan Empati dan Ketanggepan

menjadi indikator kelemahan pelayanan. Hasil penelitian ini dapat dijadikan referensi oleh PT.JNE Agen Arcadia untuk melakukan evaluasi terhadap pelayanan khususnya indikator empati dan ketanggapan agar dapat meningkatkan loyalitas pelanggan.

2.1.12 Literatur 12

Penyakit jantung merupakan salah satu penyakit yang berbahaya. Penyakit jantung dapat membahayakan nyawa penderitanya jika ada keterlambatan dalam penanganannya. Permasalahan ini disebabkan sulitnya deteksi awal pada penderita penyakit jantung karena penderita selalu mengabaikan gejala awal yang timbul. Selain itu biaya yang diperlukan dalam pemeriksaan penyakit jantung tidaklah murah sebab diperlukan pemeriksaan yang dilakukan dokter spesialis serta uji laboratorium. Sistem prediksi merupakan salah satu opsi yang dapat digunakan untuk melakukan deteksi dini pada penderita penyakit jantung dengan biaya yang lebih murah dalam penggunaannya, hal ini disebabkan biaya yang digunakan dalam pemeriksaan dokter spesialis dan tes laboratorium bisa dihilangkan dan digantikan oleh sistem prediksi. Penelitian ini bertujuan untuk membuat sistem prediksi dengan menggunakan algoritma *C4.5* dimana algoritma ini membuat prediksi berdasarkan data historis pasien yang akan diperiksa. Hasil yang didapat dari penggunaan algoritma *C4.5* untuk melakukan prediksi mendapatkan akurasi 79% sehingga diharapkan bisa menjadi sumber informasi untuk penelitian selanjutnya tentang prediksi dengan menggunakan algoritma *C4.5*.

2.1.13 Literatur 13

Pendidikan dalam kehidupan suatu negara memegang peranan yang sangat penting untuk menjamin kelangsungan hidup negara dan bangsa. Statistik menunjukkan bahwa tingkat pendidikan Portugal berada di peringkat terbawah yang disebabkan banyak siswa yang putus sekolah. Faktor eksternal berpengaruh pada kegagalan siswa dalam menyelesaikan bidang studi khususnya bidang studi matematika. Algoritma *C4.5* merupakan salah satu metode *data*

mining untuk memprediksi kemampuan siswa dalam menyelesaikan bidang studi dilihat dari faktor eksternal siswa. Algoritma *C4.5* digunakan untuk mengetahui tingkat akurasi prediksi kemampuan siswa sekolah menengah. Parameter pemilihan fitur adalah faktor-faktor yang mempengaruhi kemampuan siswa sekolah menengah dalam bidang studi matematika. Hasil pengujian dan analisis menunjukkan bahwa Algoritma *Decision tree C4.5* akurat diterapkan untuk prediksi nilai akhir siswa sekolah menengah dengan tingkat akurasi 60%.

2.1.14 Literatur 14

Tingkat keberhasilan pengiriman barang merupakan faktor yang sangat penting dalam dunia bisnis ecommerce dengan jasa logistik pengiriman menjadi rantai penghubung terakhir dengan *Customer*. Hal ini sangat mempengaruhi kepuasan *Customer* yang berharap barang yang mereka beli dapat terkirim tepat waktu dan dalam keadaan baik pada saat diterima. Salah satu teknik dari *data mining* yang dapat digunakan untuk membantu prediksi adalah menggunakan teknik klasifikasi. Klasifikasi dapat dilakukan dengan *decision tree* yaitu dengan algoritma *C4.5*. Tujuan dari penelitian ini adalah membuat klasifikasi serta menerapkan klasifikasi *data mining* berdasarkan pohon keputusan dan *rules* yang di hasilkan. Selanjutnya hasil klasifikasi data dievaluasi dengan menggunakan *confusion matrix* dan kurva ROC pada aplikasi *rapid miner* untuk mengetahui tingkat hasil akurasi. Pada penelitian ini menghasilkan nilai akurasi sebesar 93% dan menghasilkan nilai AUC (*Area Under Curve*) sebesar 0.739 dengan nilai akurasi klasifikasi cukup (*Fair classification*). Dari hasil penelitian ini dapat menjadi rekomendasi bagian distribusi dalam memilih jasa logistik dalam pengiriman barang berdasarkan kategori barang, alamat tujuan agar resiko kegagalan dalam pengiriman dapat dikurangi.

2.1.15 Literatur 15

Posyandu di Dawuan Barat menentukan status gizi anak dengan melihat grafik pertumbuhan di KIA dan menghitung *z-score* secara manual kemudian mencocokkan hasilnya dengan tabel kategori dan ambang batas, hal ini memakan waktu lama dan berisiko tidak akurat. Rumusan masalah dalam penelitian ini adalah bagaimana melakukan klasifikasi *data mining* untuk mengetahui status gizi balita dan bagaimana evaluasi hasil dari model klasifikasi tersebut. Penelitian ini menggunakan algoritma *C4.5* dengan metodologi CRISP-DM untuk mengklasifikasikan status gizi balita dan menggunakan matriks konfusi untuk mengetahui nilai akurasi, *precision*, *recall* dan *f1-score* dari model klasifikasi tersebut kemudian model tersebut diimplementasikan ke dalam sebuah aplikasi. Hasil evaluasi terhadap 3 skenario model pada penelitian ini menyatakan bahwa skenario 1 menghasilkan performa terbaik diantara model lainnya dengan akurasi 90% dan nilai presisi, *recall* dan *f1-score* 87%.

2.2 Data Mining

Menurut Prasetyo, Eko (2014:1) mengartikan *data mining* sebagai berikut: “*Data mining* adalah campuran dari statistik, kecerdasan buatan, dan riset basis data yang masih berkembang”. Proses *data mining* secara sistematis, ada tiga langkah utama yaitu:

1. Eksplorasi/pemrosesan awal data

Eksplorasi/pemrosesan awal data terdiri dari ‘pembersihan’ data, normalisasi data, transformasi data, penanganan data yang salah, reduksi dimensi, pemilihan subset fitur, dan sebagainya.

2. Membangun model dan melakukan validasi terhadapnya

Membangun model dan melakukan validasi terhadapnya berarti melakukan analisis berbagai model dengan kinerja prediksi yang terbaik. Dalam langkah ini digunakan

metode-metode seperti, klasifikasi, regresi, analisis cluster, deteksi anomali juga masuk dalam langkah eksplorasi.

3. Penerapan

Penerapan berarti menerapkan model pada data yang baru untuk menghasilkan perkiraan/prediksi masalah yang diinvestigasi (Prasetyo, 2014).

2.2.1 Jenis-Jenis Metode dalam *Data Mining*

Ada beberapa metode yang dapat digunakan dalam *data mining*, yang berguna untuk mengolah data agar dapat dianalisis sesuai kebutuhan. Adapun jenis-jenisnya adalah sebagai berikut:

1. *Predictive modelling* yang merupakan pengolahan *data mining* dengan melakukan prediksi/peramalan. Tujuan metode ini untuk membangun model prediksi suatu nilai yang mempunyai ciri-ciri tertentu.
2. *Association* (Asosiasi) merupakan teknik dalam *data mining* yang mempelajari hubungan antar data.
3. *Clustering* (Klastering) atau pengelompokan merupakan teknik untuk mengelompokkan data ke dalam suatu kelompok tertentu.
4. *Classification* merupakan teknik mengklasifikasikan data. Perbedaannya dengan metode *clustering* terletak pada data, dimana pada *clustering variabel* dependen tidak ada, sedangkan pada *classification* diharuskan ada variabel dependen.

2.2.2 Proses *Data Mining*

Secara sistematis ada tiga langkah utama dalam *data mining*:

1. Eksplorasi/pemrosesan awal data

Eksplorasi/pemrosesan awal data terdiri dari *cleansing* data untuk menghindari redundansi data dan menghilangkan data terdapat *record* yang kosong, menormalkan

data guna memberikan jangkauan antar data, transformasi data, penanganan data yang salah, reduksi dimensi, pemilihan subset fitur, dan sebagainya.

2. Membangun model dan melakukan validasi terhadapnya

Memodelkan dan memvalidasi terhadap kinerja dan tingkat akurasi dari algoritma prediksi yang digunakan untuk mendapatkan hasil terbaik. Dalam langkah ini digunakan algoritma-algoritma seperti, klasifikasi, regresi, analisis *cluster*, deteksi anomali juga masuk dalam langkah eksplorasi.

3. Penerapan

Penerapan berarti menerapkan model pada data yang baru untuk menghasilkan prediksi masalah yang akan terjadi pada kejadian dimasa mendatang (Prasetyo, 2014).

2.3 Klasifikasi

Klasifikasi merupakan proses untuk menemukan fungsi dan model yang dapat membedakan atau menjelaskan konsep atau kelas data dengan tujuan memperkirakan kelas yang tidak diketahui dari suatu objek. Dalam proses pengklasifikasian biasa terdapat dua proses yang harus dilakukan, yaitu:

1. Proses *Training*

Pada proses ini akan digunakan data *training set* atau data sampel yang telah diketahui label – label atau atribut dari data sampel tersebut untuk membangun model dengan menggunakan 80% dari keseluruhan *dataset*.

2. Proses *Testing*

Pada proses *testing* ini dilakukan untuk mengetahui keakuratan model yang telah dibuat pada proses *training* maka dibangun data yang disebut dengan data *testing* untuk mengklasifikasi label – labelnya. Klasifikasi merupakan penempatan objek – objek kesalah satu dari beberapa kategori yang telah ditetapkan sebelumnya.

Klasifikasi sekarang ini telah banyak digunakan dalam berbagai aplikasi, sebagai contoh pendeteksian pesan email, spam berdasarkan *header* dan isi atau mengklasifikasikan galaksi berdasarkan bentuk – bentuknya. Pada proses klasifikasi data yang di-*input*-kan adalah data *record* atau data sampel. Pada setiap *record* dikenal sebagai *instance* atau contoh yang ditentukan oleh sebuah *tuple* (x, y). Dimana x adalah himpunan atribut dan y adalah atribut tertentu yang menyatakan sebagai label *class* (Nugroho & Subanar, 2013).

2.4. Algoritma C4.5

Algoritma C4.5 adalah salah satu algoritma yang digunakan dalam membangun suatu keputusan yang ditampilkan dengan pohon keputusan (*Decision tree*). Algoritma C4.5 merupakan pengembangan dari ID 3 yang dapat mengatasi *missing value*, mengatasi *continue data*, dan *pruning*. Secara umum alur proses algoritma C4.5 untuk membangun pohon keputusan dalam *data mining* adalah:

- 1) Pilih atribut sebagai simpul akar
- 2) Buat cabang untuk tiap-tiap nilai.
- 3) Bagi kasus dalam cabang
- 4) Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama (Yunus et al., 2014).

Pemilihan atribut sebagai simpul, baik simpul akar (*root*) atau simpul internal diperoleh berdasarkan nilai *gain* tertinggi, Perolehan informasi dari atribut A relatif terhadap output data S. Perolehan informasi didapat dari *output data* atau *variabel dependent* S yang dikelompokkan berdasarkan atribut A, dinotasikan dengan *gain* (S, A). Sebelum mencari *gain* tertinggi, langkah pertama yang perlu dilakukan adalah mencari nilai *entropy* untuk setiap atribut yang ada untuk memperkirakan jumlah bit yang dibutuhkan untuk mengekstrak suatu kelas positif

atau negatif dari sejumlah data acak pada ruang sampel (S). Berikut persamaan yang digunakan untuk mencari nilai *entropy* (Prasetyo, 2014):

$$Entropy (S) = \sum_{i=1}^c -p_i \log_2 p_i \dots\dots\dots 2.1$$

Keterangan:

- S : himpunan data set
- C : jumlah kelas
- p_i : Probabilitas frekuensi kelas ke-i dalam *Dataset*

$$Gain (A) = Entrophy (S) - \sum_{i=1}^k \frac{|S_i|}{|S|} x Entrophy (S_i) \dots\dots 2.2$$

Keterangan:

- S : Himpunan data set
- A : Atribut
- $|S_i|$: Jumlah sampel untuk nilai i
- $|S|$: Jumlah seluruh sampel data

2.5 Confusion Matrix

Confusion matrix adalah algoritma yang dapat menghitung akurasi, *precision*, dan *recall*. Akurasi adalah hasil dari membagi semua nilai yang diprediksi dengan benar dengan jumlah total data. Nilai akurasi tertinggi adalah 100 dan nilai akurasi terendah adalah 0%. (Sigit, 2018).

$$Akurasi = \frac{TP+TN}{TP+TN+FN+FP} x 100\% \dots\dots\dots (2.3)$$

Precision didapat dengan menghitung jumlah keseluruhan nilai prediksi positif yang benar dibagi dengan jumlah keseluruhan prediksi kelas yang benar. Nilai terbaik *precision* adalah 100% sementara yang terburuk 0% (Sigit, 2018).

$$Precision = \frac{TP}{TP+FP} \times 100\% \dots\dots\dots (2.4)$$

Sedangkan *true positive rate* atau biasa disebut dengan *recall* adalah jumlah prediksi benar dibagi dengan keseluruhan jumlah kelas yang salah. Untuk nilai terbaik *recall* adalah 100% sementara yang terburuk adalah 0% (Sigit, 2018).

$$Recall = \frac{TP}{TP+FN} \times 100\% \dots\dots\dots (2.5)$$

Secara definisi, *F1-Score* adalah *harmonic mean* dari *precision* dan *recall*, yang secara matematik dapat ditulis begini:

$$\frac{1}{f1} = \frac{1}{2} \left(\frac{1}{precision} + \frac{1}{recall} \right) \dots\dots\dots (2.6)$$

Nilai terbaik *F1-Score* adalah 1 dan nilai terburuknya adalah 0. Secara representasi, jika *F1-Score* punya skor yang baik mengindikasikan bahwa model klasifikasi kita punya *precision* dan *recall* yang baik (Sigit, 2018).

2.6 Pengertian Bank

Bank termasuk industri jasa karena produksinya memberikan pelayanan jasa kepada masyarakat. Adapun pengertian bank menurut undang-undang RI nomor 10 tahun 1998 bank adalah “badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkannya kepada masyarakat dalam bentuk kredit dan atau bentuk- bentuk lainnya dalam rangka meningkatkan taraf hidup rakyat banyak” (Kasmir, 2007).

Definisi lain mengenai bank yang dikemukakan oleh Suyatna (2003) bank adalah suatu badan yang tugas utamanya sebagai perantara untuk menyalurkan penawaran dan permintaan kredit pada waktu yang ditentukan. Sedangkan menurut Sinungan (2003) bank adalah suatu

lembaga keuangan yang usaha pokoknya memberikan kredit dan jasa-jasa dalam lalu lintas pembayaran dan peredaran uang. Menurut Kasmir (2003) bank adalah suatu badan usaha yang bergerak dalam bidang keuangan dan juga sebagai tempat untuk pembayaran dan setoran.

2.7 Pengertian Nasabah

Nasabah adalah pelanggan (*customer*) yaitu individu atau perusahaan yang mendapatkan manfaat atau produk dan jasa dari sebuah perusahaan perbankan, meliputi kegiatan pembelian, penyewaan serta layanan jasa. (Hayati, Mislah Nasution, 2015) Nasabah menurut Pasal 1 ayat (17) UU No. 10 tahun 1998 adalah “Pihak yang menggunakan jasa bank.” Nasabah mempunyai peran penting dalam industri perbankan, dimana dana yang disimpan nasabah di bank merupakan dana yang terpenting dalam operasional bank untuk menjalankan usahanya. Adapun pengertian nasabah menurut para ahli, sebagai berikut.

Menurut Kasmir “Nasabah merupakan konsumen yang membeli atau menggunakan produk yang dijual atau ditawarkan oleh bank.” (Koeswara, Sonny, 2015) Menurut Saladin nasabah adalah “Konsumen-konsumen sebagai penyedia dana.” (Yupitri, Evi dan Raina Linda Sari, 2012) Menurut Tjiptono pengertian nasabah adalah “Setiap orang yang membeli dan menggunakan produk atau jasa Perusahaan.” (Perwitasari, Dwi Wiryaningtyas, 2016) Menurut Komaruddin nasabah adalah “Seseorang atau suatu perusahaan yang mempunyai rekening koran, deposito atau tabungan serupa lainnya pada sebuah bank.” (Perwitasari, Dwi Wiryaningtyas, 2016). Dari pengertian tersebut dapat ditarik kesimpulan bahwa nasabah adalah seseorang atau badan usaha (korporasi) yang mempunyai rekening simpanan dan pinjaman dan melakukan transaksi simpanan dan pinjaman tersebut pada sebuah bank. (Perwitasari, Dwi Wiryaningtyas, 2016)

2.8 Perpindahan Nasabah

Perpindahan konsumen merupakan suatu kondisi dimana pelanggan berpindah dari satu penyedia jasa ke penyedia lain. Menurut Ganesh, Arnold dan Reynold, perpindahan konsumen

adalah perilaku konsumen yang mencerminkan pergantian dari merek produk yang biasa di konsumsi dengan merek lain. Munculnya perilaku perpindahan konsumen dipengaruhi oleh dua faktor yaitu faktor ekstrinsik dan faktor intrinsik.

a. Faktor ekstrinsik yaitu faktor-faktor yang menjadi penentu perpindahan konsumen yang berasal dari luar diri konsumen (Rahman, Ardhista, 2013). Faktor-faktor tersebut antara lain:

1. *Sales promotion* merupakan suatu cara untuk mempengaruhi konsumen agar dengan suka langsung membeli barang dengan merek tertentu, sehingga dengan adanya *sales promotion* yang baik akan mudah mempengaruhi konsumen dalam berpindah.
2. Kualitas atau mutu adalah ukuran kemampuan suatu merek untuk melaksanakan fungsi-fungsinya. Kualitas merupakan ukuran menyeluruh yang mencerminkan nilai suatu produk berkenaan dengan soal keawetan, keandalan, keseksamaan, kemudahan operasi reparasi dan sebagai atribut yang bernilai tinggi lainnya (Sumarwan, Ujang, 2011)

b. Faktor instrinsik yaitu faktor-faktor yang mempengaruhi perilaku perpindahan konsumen yang berasal dari dalam diri konsumen tersebut. Faktor-faktor tersebut antara lain:

1. *Variety seeking* adalah perilaku konsumen yang berusaha mencari keberagaman merek di luar kebiasaannya karena tingkat keterlibatan beberapa produk rendah. Perilaku ini sering terjadi pada beberapa produk, dimana tingkat keterlibatan produk dikatakan rendah, apabila dalam proses pembelian produk konsumen tidak melibatkan banyak faktor dan informasi yang harus ikut dipertimbangkan.
2. Loyalitas, loyalitas konsumen terbentuk karena konsumen merasa puas dengan suatu produk atau jasa tertentu. Oleh sebab itu, konsumen yang memiliki loyalitas yang rendah memiliki kemungkinan untuk melakukan perpindahan. (Rahman, Ardhista, 2013)

2.9 Perilaku Konsumen

Perilaku konsumen sebagai tindakan yang langsung terlibat dalam mendapatkan, mengkonsumsi, dan menghabiskan produk dan jasa, termasuk proses keputusan yang mendahului dan mengikuti tindakan ini. Perilaku konsumen adalah semua kegiatan, tindakan, serta proses psikologi yang mendorong tindakan tersebut pada saat sebelum membeli, ketika membeli, menggunakan, menghabiskan produk, dan jasa setelah melakukan hal-hal atau kegiatan mengevaluasi. (Sumarwan, Ujang, 2011)

2.9.1 Faktor yang Mempengaruhi Perilaku Konsumen

Faktor yang dapat mempengaruhi perilaku konsumen biasanya adalah faktor kepribadian konsumen. Perkembangan kepribadian individu dipengaruhi oleh berbagai faktor di antaranya faktor hereditas dan lingkungan. Di samping itu, meskipun kepribadian seseorang itu relatif konstan, kenyataannya sering ditemukan perubahan kepribadian. Perubahan itu terjadi dipengaruhi oleh faktor gangguan fisik dan lingkungan. (Yusuf, Syamsu dan Juntika Nurihsan, 2007)

