

BAB II

TINJAUAN PUSTAKA

2.1. Teori Dasar

Pada subbab ini peneliti akan melakukan penjabaran tentang penafsiran sejumlah teori yang menjadi pedoman dasar dalam melakukan kegiatan penelitian ini dimana definisi dan terminologi yang berhubungan dengan judul penelitian.

2.1.1. Teori Umum

Penulis memaparkan sejumlah teori umum yang berhubungan dengan terlaksananya penelitian, termasuk definisi dan istilah umum.

2.1.1.1. KDD (*Knowledge Discovery In Database*)

Knowledge Discovery in Database (KDD) didefinisikan sebagai sistem pencarian informasi pada basis data yang ada. Basis data tersebut berupa korelasi dimana informasi yang ditemukan dalam database ditransformasikan menjadi basis pengetahuan yang digunakan sebagai bahan pertimbangan dalam mengambil keputusan.

Knowledge Discovery in Database (KDD) dan data mining dalam database yang dikolaborasikan oleh pemangku kepentingan sesuai kebutuhan dalam kondisi tertentu untuk menjelaskan potensi aktivitas penambangan informasi, tetapi berada dalam database data yang cukup besar. Meskipun kedua komponen tersebut memiliki konsep yang berbeda, namun keduanya saling terkait dan juga salah satu hierarki prosedural KDD adalah data mining. Prosedur KDD berikut dijelaskan di bawah ini.

1. *Data Selection*

Langkah ini merupakan Langkah paling awal. Artinya, dilakukan dengan mengumpulkan informasi melalui proses pemilihan data yang akan digunakan sebagai sumber data yang pada akhirnya diproses pada tahap data mining.

2. *Pre-processing* atau *Cleaning*

Langkah ini dilakukan untuk menghilangkan beberapa duplikat yang ada pada data Anda, dan proses penambahan data harus disesuaikan dengan kebutuhan pencarian Anda, jadi pastikan Anda tidak memenuhi kebutuhan Anda.

3. *Transformation*

Proses transformasi data dipilih agar data tersebut sesuai untuk proses data mining. Proses ini kreatif dan sangat bergantung pada jenis atau pola informasi yang diambil dalam database.

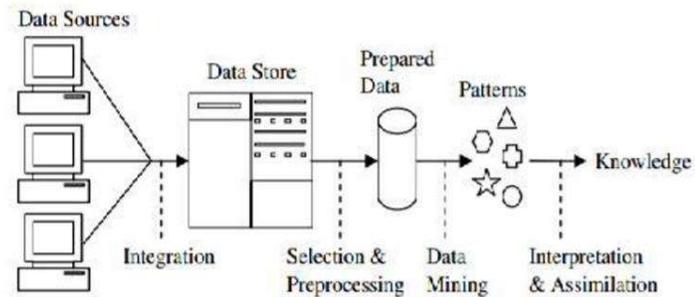
4. *Data Mining*

Data mining adalah proses menemukan pola atau informasi yang menarik dalam data yang dipilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma penambangan data sangat bervariasi. Memilih metode atau algoritma yang tepat sangat tergantung pada tujuan dan proses KDD secara keseluruhan.

5. *Interpretation*

Fase ini merupakan bagian dari proses KDD, yang terdiri dari verifikasi apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada.

Sementara itu menurut (Mardi, 2016) proses KDD atau *Knowledge Discovery in Database*, yaitu:



Gambar 2.1 Proses Dalam KDD atau *Knowledge Discovery in Database*

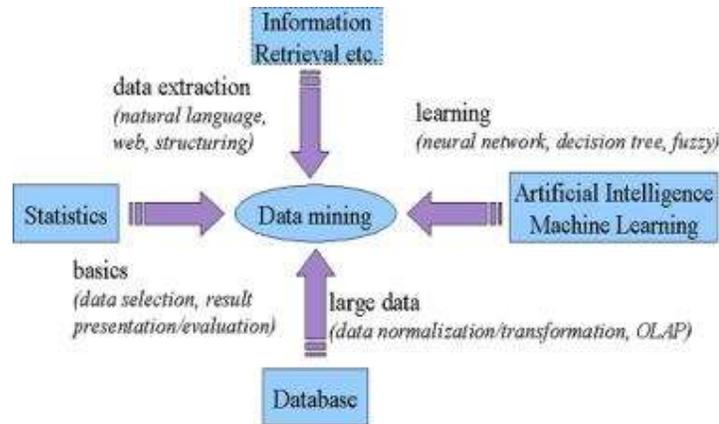
Sumber: (Mardi, 2016)

2.1.1.2. Data Mining

Data mining mengacu pada proses mengambil informasi yang tidak diketahui dari kumpulan data yang besar. Dalam proses ini, data mining berasal dari pengekstrakan informasi bermutu dengan analisa keberadaan suatu pola atau perhubungan penentuan dari jumlah data besar. Definisi lain dari data mining adalah urutan proses yang secara otomatis menganalisis dan mengekstrak pengetahuan menggunakan satu atau lebih teknik komputasi atau urutan penemuan dan atau penambahan dari sekumpulan data sebagai ilmu pengetahuan. Ini adalah prosesnya. Tanpa diketahui di tangan. Karena data mining merupakan urutan dari pemrosesan, agar data mining dapat dipecah hingga menjadi langkah. Langkah-langkah ini interaktif, di mana pengguna berpartisipasi baik secara langsung atau melalui basis pengetahuan.

Penambangan data atau pembelajaran mesin dan penambangan data biasanya digunakan untuk melakukan berbagai bentuk operasi analitik untuk

menjelaskan sains dan pengetahuan dengan cara yang jelas dan ringkas yang dapat dipahami oleh pikiran manusia (Hermawan, Sukma, & Halfis, 2019).



Gambar 2.2 Bidang Ilmu Data Mining

Sumber: (Mardi, 2016)

Dalam proses ini, data mining memiliki teknologi pengolahan data sendiri untuk membentuk sebuah model. Model diaktifkan agar dapat mengenali model lain yang tidak disimpan di lokasi penyimpanan yang sama. Dalam bisnis atau proses prediktif, banyak peneliti menggunakan kemampuan teknologi data mining untuk melakukan operasi ini. Istilah data mining berasal dari fakta bahwa data dan informasi yang diproses telah ada sejak lama, pengolahannya dilakukan dengan cara yang sama seperti penambangan, dan diperoleh hasil yang berharga dan bermanfaat (Hermawan et al., 2019).

2.1.1.3. Pengelompokan *Data Mining*

Jika pentahapan didasarkan pada tugas dan proses yang dilakukan dalam penambangan data, maka dapat dipecah menjadi beberapa tahap dan bagian.

1. *Description* atau Deskripsi

Pada titik ini, dimulai ketika analis mencoba untuk mengklasifikasikan jenis pola yang ada dalam sumber data dan informasi. Hal ini dapat dijelaskan karena penanggung jawab penghitungan kegiatan pemungutan suara dalam kampanye pemilu tidak mengetahui atau memiliki informasi yang jelas tentang bakat, keterampilan, dan kemampuan kandidat untuk mengikuti pemilu. Oleh karena itu, penjelasan masalah merupakan upaya untuk memberikan penjelasan tentang potensi kemampuan seorang calon untuk mendaftar mengikuti pemilu dengan rencana.

2. *Estimation* atau Estimasi

Selain itu, dalam fase atau proses ini, aktivitas yang dilakukan oleh estimator memiliki gaya kerja yang mirip dengan proses atau fase klasifikasi, tetapi variabel tujuan yang terkandung dalam estimasi proses atau fase yang berisi data digital dapat bervariasi. Sampel yang dihasilkan berasal dari rekaman data yang berisi nilai atau angka dari variabel yang terkandung yang digunakan sebagai dokumen untuk mendapatkan nilai dari proses peramalan. Selanjutnya, tetapkan perkiraan yang berasal dari nilai-nilai yang terkandung dalam variabel yang berasal dari nilai yang diprediksi. Hal ini dapat dijelaskan dengan nilai ($S.1$) yang termasuk dalam rata-rata indeks prestasi kumulatif atau rata-rata indeks prestasi kumulatif mahasiswa yang sedang menempuh program pendidikan tinggi. Estimasi dapat ditemukan berdasarkan hasil IPK atau Indeks Prestasi Siswa. Proses pendidikan untuk setiap semester.

3. *Prediction* atau Prediksi

Proses atau langkah peramalan mirip dengan proses klasifikasi dan estimasi, tetapi tidak termasuk peramalan apakah nilai sekarang adalah nilai masa depan. Prakiraan yang disebutkan sangat relevan dengan penelitian dan kegiatan bisnis, dan lebih tepatnya,

- a. Saat membuat prediksi untuk mengetahui potensi harga beras untuk beberapa bulan mendatang.
- b. Pada saat yang sama, ini memberikan perkiraan tingkat pengangguran potensial selama beberapa tahun ke depan.
- c. Prediksi sebagai persentase menggambarkan kemungkinan kecelakaan jika pengemudi melebihi batas kecepatan di tahun-tahun mendatang.

4. *Classification* atau Klasifikasi

Proses pemeringkatan atau kegiatan menggunakan tujuan yang berbeda dengan variabel berdasarkan kategori yang ada. Hal ini dapat dijelaskan dengan membagi pendapatan pekerja menjadi tiga kategori ketika mengelompokkan pendapatan karyawan suatu perusahaan. Artinya, karyawan kelompok pendapatan adalah pekerja dengan kategori pendapatan sangat tinggi, sedang, atau standar. Pekerja dalam kategori berpenghasilan sangat rendah Ketika klasifikasi terakhir ditafsirkan dalam kaitannya dengan penelitian dan kegiatan bisnis:

- a. Ada transaksi kredit yang curang atau tidak curang dalam proses penilaian.

- b. Saat mengantisipasi rencana kredit, klien akan melakukannya apakah mereka melakukannya dengan benar atau tidak.
 - c. saluran Dokter mendiagnosis penyakit dan penyakit yang dialami dan diperoleh pasien berdasarkan gejala yang ada.
5. *Clustering* atau Pengklusteran

Proses langkah-langkah yang termasuk dalam pengelompokan adalah suatu kegiatan atau proses yang mengelompokkan record data menjadi satu, dan record data tersebut dikelompokkan berdasarkan data pada titik atau pola yang sama. Cluster yang dimaksud memiliki hasil entri data yang sama atau mirip dengan sampel pada satu kelompok, dan kelompok lainnya tidak sama atau mirip dengan hasil entri yang ada. Satu atau lebih proses yang terlibat dalam clustering berbeda dengan proses klasifikasi atau satu atau lebih langkah karena clustering tidak memiliki variabel target. Perhitungan algoritma yang digunakan dalam proses pengelompokan menitikberatkan pada pengelompokan data yang ada berdasarkan kesamaan pola yang dimiliki data tersebut, sehingga kerja yang dilakukan pada langkah dan proses pengelompokan tersebut merupakan perkiraan. , Proses prediksi dan klasifikasi tidak dilakukan. alokasi. Nilai maksimum berasal dari atau sesuai dengan kesamaan atau kesamaan data yang dimiliki dalam kelompok. Jika kemiripan atau kemiripan data yang ada sangat rendah, maka yang terjadi adalah sebaliknya. Nilai yang dihasilkan Ini adalah nilai minimum. Hal ini dapat dijelaskan sebagai berikut.

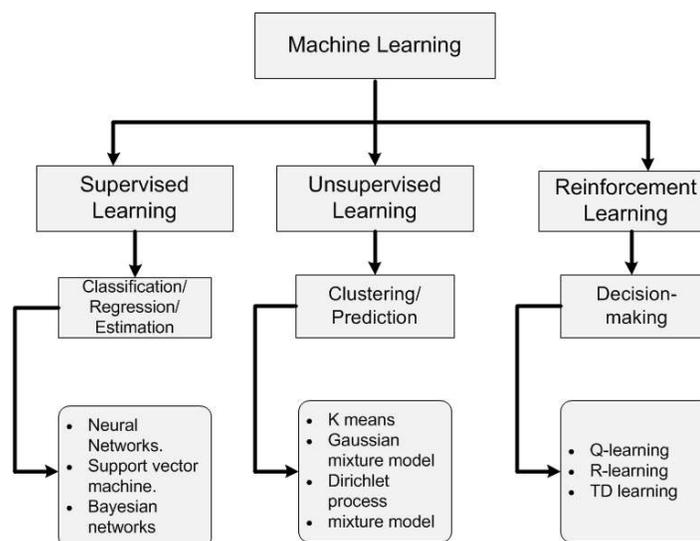
- a. Saat ini, kelompok tipe konsumen menjadi subyek kegiatan pemasaran produk yang dilakukan oleh perusahaan.
- b. Saat melakukan audit, kami memisahkan diri berdasarkan apa yang kami anggap mencurigakan dan apa yang kami anggap pasti.

6. *Association* atau Asosiasi

Asosiasi menerapkan proses pemulihan integritas atau informasi pendukung yang terkandung dalam data. Hal ini dapat dijelaskan dengan bentuk kegiatan yang dilakukan,

- a. Ketika mereka melakukan survei atau analisis terhadap jumlah pelanggan dalam suatu bisnis, mereka memberikan umpan balik positif atas layanan yang diberikan.
- b. Pelajari tentang semua jenis produk di pasar tetapi tidak pernah dibeli atau dikonsumsi.

Interpretasi visual meliputi hasil pengelompokan proses tertentu atau langkah demi tugas ketika seorang peneliti melakukan penelitian dengan menggunakan teknik data mining, seperti:



Gambar 2.3 Pengelompokan Teknik Data Mining

Sumber: (Hermawan, Sukma, & Halfis, 2019)

2.1.1.4. *Decision Tree*

Manusia masih menghadapi masalah yang berbeda dalam berbagai bidang kehidupan. Ada juga berbagai tingkat kesulitan untuk masalah ini. Untuk mengatasi masalah ini, orang telah mulai mengembangkan sistem yang membantu memecahkan masalah ini. Salah satunya adalah pohon keputusan. Metode pohon keputusan dimaksudkan sebagai kumpulan aturan untuk membagi populasi menjadi subpopulasi yang lebih kecil dari, untuk suatu tujuan. Variabel target dikelompokkan ke dalam kelompok sempit dan model pohon keputusan memiliki arah yang jelas untuk dalam menghitung potensi untuk setiap catatan versus untuk kategori tertentu. Pohon keputusan dapat diperluas dengan mengimplementasikan algoritma yang memodelkan kumpulan data yang kelasnya tidak didefinisikan dengan jelas.

Hal ini sering terjadi dalam kehidupan manusia karena berbagai jenis masalah dan tingkat kesulitan yang berbeda, dan pada akhirnya orang mengembangkan model dalam bentuk sistem untuk membantu berbagai jenis masalah, solusi yang berbeda. disebut pohon keputusan atau sering disebut pohon keputusan. Pohon keputusan adalah bentuk implementasi pemodelan dengan menerapkan berbagai langkah dan prosedur yang berkaitan dengan operasi klasifikasi dan peramalan. Pohon keputusan banyak dikenal dan digunakan oleh peneliti ketika melakukan penelitian dengan menggunakan teknik data mining, karena pohon keputusan menghasilkan alternatif pemecahan masalah dengan hasil yang bermanfaat dan kuat. Pohon keputusan berguna untuk mengubah data yang

ada menjadi model tipe aturan yang mudah dipahami manusia. Pada titik ini, pohon keputusan melalui proses awal dan akhir dengan semua bagian data yang berisi kondisi yang berbeda dan informasi yang lengkap. Memang, arsitektur pohon keputusan mirip dengan yang ada di pohon dan deskripsinya. Artinya, terlihat seperti ini:

1. Simpul bagian akar

Titik puncak dari akar berada di bagian atas pohon keputusan, sebelum struktur yang terdapat dalam struktur.

2. Simpul internal

Node internal ini merupakan cabang tambahan dari root node. Node ini memungkinkan Anda untuk keluar hingga dua pintu keluar jika Anda memiliki satu pintu masuk.

3. Simpul daun

Simpul daun adalah simpul di ujung pohon. Node daun memiliki input tetapi tidak memiliki output.

Selanjutnya, demikian penjelasan dari kelebihan dan kekurangan yang dimiliki oleh *decision tree*, yaitu:

1. Kelebihan dari *decision tree*

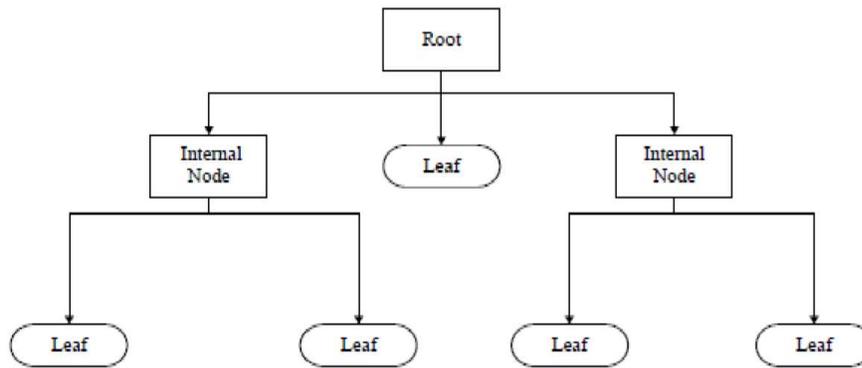
- a. Gunakan pohon keputusan saat Anda memiliki rentang data yang kompleks. Ini akan dimodifikasi agar lebih mudah digunakan sebagai kebijakan keputusan.
- b. Keuntungan menggunakan pohon keputusan adalah Anda dapat menghilangkan perhitungan yang tidak perlu untuk hanya

menggunakan kriteria yang sesuai dengan kebutuhan pengujian Anda saat melakukan prosedur pengujian.

- c. saluran Ketika pohon keputusan digunakan dalam integritas data atau proses pemilihan deskripsi, hasil dari pemilihan integritas atau akurasi data adalah faktor pembeda untuk peringkat data dengan kriteria lain.
- d. Pohon keputusan bekerja untuk menghindari masalah lain dengan menggunakan sejumlah kecil jenis kelompok atau kategori tanpa mempengaruhi kualitas hasil keputusan akhir.

2. Kekurangan dari *decision tree*

- a. Saat menggunakan pohon keputusan, jika jumlah keputusan yang akan dibuat dan jumlah kriteria yang digunakan sangat besar, waktu yang dibutuhkan untuk menghasilkan keputusan akan lama dan diperlukan sistem pendukung memori yang besar.
- b. Sebuah pohon keputusan memiliki jumlah kumulatif kesalahan yang terjadi di setiap bagian dari struktur yang terdapat dalam pohon keputusan.
- c. Sulit untuk membangun pohon keputusan yang memberikan hasil terbaik.
- d. Saat menggunakan pohon keputusan, hasil atau kualitas keputusan yang dihasilkan sangat ditentukan oleh model atau struktur pohon keputusan yang dikembangkan.



Gambar 2.4 *Model Decision Tree*

Sumber: (Takalapeta, 2018)

2.1.1.5. Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang dapat digunakan dan dibentuk menjadi pohon keputusan atau dalam bahasa Inggris disebut dengan decision tree. Decision tree adalah metode yang sangat kuat banyak diketahui untuk mengklasifikasikan dan memprediksi angka. Method decision tree data yang diubah sangat besar menjadi aturan yang diwakilkan oleh decision tree. Rule yang dimengerti bahasa alami dimulai dengan. Anda juga dapat mengungkapkan ini dalam bahasa database, seperti Structured Query Language, untuk menemukan catatan dalam katalog khusus.

Algoritma C4.5 membantu untuk melakukan klasifikasi data menggunakan angka atau kategori. Setelah melalui langkah-langkah proses klasifikasi, hasilnya menjadi seperangkat aturan model yang dapat digunakan untuk melakukan langkah-langkah proses peramalan dengan menggunakan nilai prediksi yang diperoleh dari penyimpanan data akhir dari data lama. Jenis algoritma C4.5 berasal dari algoritma ID3. Singkatnya, algoritma C4.5 adalah hasil perbaikan lebih lanjut dari algoritma ID3, yang memungkinkan algoritma C4.5 untuk memecahkan

masalah tanpa adanya data. Selesaikan masalah saat muncul. Secara umum algoritma C4.5 digunakan dalam proses membangun struktur pohon keputusan, sehingga prosedur atau deskripsi langkahnya adalah sebagai berikut:

1. Yang pertama adalah memilih integritas atau deskripsi data sebagai simpul akar dari pohon keputusan.
2. Selanjutnya, buat struktur cabang dari pohon keputusan yang sudah berisi nilai-nilai.
3. Pada langkah ketiga, masalah-masalah yang terdapat dalam struktur cabang-cabang pohon keputusan dipisahkan atau dikelompokkan bersama.
4. Langkah terakhir adalah mengulang proses dari langkah 1 ke langkah 3 sampai semua cabang dari struktur pohon keputusan memiliki kelas dan tipe yang sama.

Setelah melakukan langkah awal dari pohon keputusan dimana ketika pada saat melakukan pemilihan terhadap attribute dapat dilakukan dengan melakukan cara dicari nilai dari gain tertinggi yang ada dari total keseluruhan nilai yang ada yang dihasilkan berdasarkan perhitungan dengan rumus yang ada. Berikut rumus yang dimaksud untuk mendapatkan hasil gain tertinggi.

$$Gain (S, A) = Entropi (s) - \sum_i^n \frac{|S_i|}{|S|} \times Entropi (S_i)$$

Rumus 2.1 Perhitungan *Gain*

Keterangan dari rumus 2.1 diatas dapat dijabarkan sebagai berikut::

S berupa himpunan kasus

A berupa attribute yang digunakan.

n berupa jumlah partisi yang dimiliki oleh attribute A

$|S_i|$ berupa jumlah yang dimiliki oleh partisi ke- i

$|S|$ berupa jumlah yang dimiliki oleh partisi S

$$\text{Entropi}(s) = \sum_{i=1}^n - p_i * \log_2 p_i$$

Rumus 2.2 Perhitungan *Entropy*

Keterangan dari rumus 2.2 diatas dapat dijabarkan sebagai berikut::

S berupa himpunan kasus

A berupa attribute yang digunakan

n berupa jumlah partisi yang dimiliki oleh attribute S

p_i berupa proporsi dari S_i kepada S .

2.1.2. Teori Khusus

Penulis menjelaskan beberapa teori khusus yang berkaitan dengan pelaksanaan studi termasuk definisi dan istilah tertentu.

2.1.2.1. Profit

Menurut (Elisa, 2018), laba yang juga dikenal sebagai laba didefinisikan sebagai selisih antara pendapatan yang diterima perusahaan dari satu atau lebih objek dan biaya yang harus dikeluarkan untuk memenuhi kewajibannya, layanannya kepada para pihak. Laba atau laba adalah peningkatan nilai modal karena peristiwa masa lalu dalam pengelolaan bisnis, sebagai akibat dari transaksi moneter dalam atau luar negeri yang mempengaruhi bisnis untuk jangka waktu tertentu. Hal ini berbeda dengan laba bersih atau net profit. Ini didefinisikan sebagai

perbedaan antara kelebihan atau pendapatan/penerimaan positif dan pengeluaran/pengeluaran dan dikurangi lagi oleh bunga dan pajak.

Berdasarkan pendapat (Rika Nofitri, 2018) Sistem yang dibangun digunakan untuk memprediksi profitabilitas perusahaan. Data mentah awalnya diproses dalam format digital dan non-digital dan dikelompokkan ke dalam lapisan yang berbeda untuk memudahkan analisis data pada fase berikutnya. Data tersebut kemudian dihasilkan sebagai output sesuai dengan desain pohon keputusan. Berikut adalah beberapa hal yang harus dilakukan untuk mewujudkan profitabilitas bisnis Anda:

1. Pengadaan adalah kegiatan yang dilakukan oleh perusahaan untuk memperoleh produk yang diperlukan untuk mengelola operasi dari pengadaan hingga pengumpulan (Rotonsulu Jopie Jorie, 2015). Pembelian dalam survey ini dilakukan dalam bentuk barang yang dibeli untuk dijual kepada pembeli.
2. Omset adalah kuantitas yang terjual dari suatu produk (produk) tertentu selama periode penjualan (M. Maulana Dzikril Hakim, 2016). Penjualan dalam survei ini dinyatakan sebagai jumlah barang yang terjual per hari.
3. Biaya adalah pengorbanan atau pengeluaran yang dilakukan oleh suatu perusahaan atau individu untuk mendapatkan keuntungan yang lebih dari kegiatan yang dilakukan (Priccila Natalia, 2014). Biaya penelitian ini berupa biaya yang dikeluarkan pada satu hari untuk kegiatan operasional.
4. Persediaan adalah alat likuid berupa barang atau peralatan yang dimaksudkan untuk mendukung kegiatan investasi dan produk yang

dimaksudkan untuk dijual dan/atau ditawarkan untuk dijual sebagai bagian dari pelayanan kepada pembeli (Sri Wahyuni, 2016). Persediaan dalam survey ini adalah sisa persediaan setelah dikurangi penjualan/hari.

5. Cashback, yaitu pengembalian total nominal transaksi (Suresh Kumar, 2014). Pengembalian uang dalam penelitian ini adalah pengembalian uang yang diterima setelah membayar biaya pengiriman untuk pengiriman.
6. Paket adalah produk yang dikirim melalui pos atau dikemas oleh perusahaan pelayaran (Mumuh Mulyana, 2014). Paket untuk survei ini adalah jumlah paket yang dapat dihasilkan atau dikirim sebagai satu paket per hari per hari.

Variabel yang ditentukan untuk diambil keputusannya adalah TERCAPAI dan TIDAK TERCAPAI.

2.1.2.2. *Cross Validation*

Validasi silang, juga dikenal sebagai estimasi rotasi, adalah teknik validasi model untuk mengevaluasi bagaimana hasil analisis statistik digeneralisasikan ke kumpulan data independen. Teknik ini terutama digunakan untuk membuat prediksi model dan memperkirakan keakuratan model yang diprediksi ketika benar-benar dijalankan. Salah satu metode cross-validation adalah k-validation cross-validation. Membagi data menjadi k bagian dari dataset dengan ukuran yang sama. Gunakan validasi silang kfold untuk menghilangkan bias data. Pelatihan dan pengujian dilakukan sebanyak k kali. Pada pengujian pertama, subset S1 diperlakukan sebagai

data uji dan yang lainnya sebagai data pelatihan. Pada pengujian kedua, subset S1, S3, ... Sk adalah data pelatihan dan S2 adalah data pengujian (Bramer, 2017).

2.1.2.3. *Waikato Environment for Knowledge Analysis (WEKA)*

WEKA adalah aplikasi pembelajaran mesin yang praktis. "WEKA" adalah singkatan dari Waikato Environment for Knowledge Analysis dan berasal dari universitas yang digunakan dalam penelitian, pendidikan dan aplikasi lainnya, University of Waikato, Selandia Baru (Pujiono, Amborowati, & Suyanto, 2013). WEKA mempunyai kemampuan mengatasi banyak masalah data mining dunia nyata, mengutamakan klasifikasi berdasarkan pendekatan pembelajaran mesin. WEKA mencakup alat untuk pra-pemrosesan data, classification, regrestion, pengelompokan, rule korelasi, dan visualitation. Menguji dengan WEKA lebih fleksibel dan efisien karena pohon keputusan dibentuk mirip dengan perhitungan manual (Azwanti, 2018).



Gambar 2.5 Tampilan Utama *WEKA*

2.2. Penelitian Terdahulu

Peneliti juga menemukan beberapa hasil penelitian terdahulu yang berkaitan dengan menganalisa menggunakan algoritma C4.5 yang dirangkum yaitu:

1. Dari penelitian (Sri Lestari dan Saefudin, 2015) melakukan survei “Sistem pendukung keputusan penilaian kinerja karyawan PT. Mulya Spindo Mills Menggunakan C4.5” yang mendemonstrasikan penggunaan algoritma C4.5 telah memperoleh informasi bahwa PT. Mulya Spindo Mills dapat membangun sistem pendukung keputusan untuk menilai kinerja karyawan dengan menggunakan pendekatan algoritma C4.5 berdasarkan kriteria yang ditentukan oleh perusahaan.
2. Survei “Prediksi Profit Perusahaan Menggunakan Algoritma C4.5” yang dilakukan oleh (Elisa, 2018) jelas menunjukkan bahwa penggunaan algoritma C4.5 efektif. Dimungkinkan untuk membangun sistem pendukung keputusan untuk memprediksi profitabilitas perusahaan dengan standar yang telah ditentukan.
3. Dari “Perbandingan antara algoritma C4.5 yang dikembangkan oleh Web Intelligence dan algoritma Naive Bayes dalam menghitung bonus tahunan karyawan di PT.ABC” yang dilakukan oleh (Ibnu Alfarobi dan Taransa Agasya Tutupoly, 2019), bonus karyawan berdasarkan keputusan pohon C4.5. dan Algoritma naive Bayes dengan nilai yang sama dimana area di bawah kurva (AUC) 0,991 dan matriks akurasi atau kebingungan 99%.
4. Setelah (Rika Nofitri, 2018) melakukan penelitian “Analisis Kinerja Algoritma C4.5 dalam Prediksi Kinerja Laba”, memberikan informasi

yang menunjukkan bahwa penggunaan algoritma C4.5 terbukti dapat mempercayai keputusan yang telah dibuat. Sebuah sistem untuk membantu meramalkan profitabilitas perusahaan berdasarkan kriteria berikut yang ditentukan oleh perusahaan.

5. Berdasarkan survei “Penerapan Algoritma C4.5 Dalam Analisis Kelayakan Bonus Karyawan Tahunan (Studi Kasus: PT. Multi Pratama Nauli Medan)” yang dilakukan oleh (Syahputra, M.Kom, 2015) ditunjukkan bahwa algoritma tipe C4.5 dapat digunakan untuk membantu perusahaan besar dalam menganalisis karyawan yang memenuhi syarat untuk bonus tahunan.
6. Dari penelitian (Intan Komala Dewi Patwari dan Baiq Andriska Candra Permana, 2019) yang berjudul “Penerapan Algoritma C4.5 Pada Analisis Penentuan Kelayakan Penerima Bonus Tahunan Pegawai Bank Sinarmas Multifinance” dengan menggunakan Algoritma C4.5 diberitahu bahwa itu adalah terbukti membantu para pemimpin menentukan apakah seorang karyawan memenuhi syarat untuk mendapatkan bonus.
7. Setelah (Harina, 2020) melakukan survey “Penerapan algoritma C4.5 untuk mengetahui tingkat keuntungan perusahaan di PT Megacipta Sejati”, hasilnya menunjukkan bahwa penggunaan algoritma C4.5 dapat dibangun. Sebuah sistem pendukung keputusan untuk menentukan tingkat profitabilitas komersial suatu perusahaan sesuai dengan kriteria yang ditentukan oleh perusahaan.

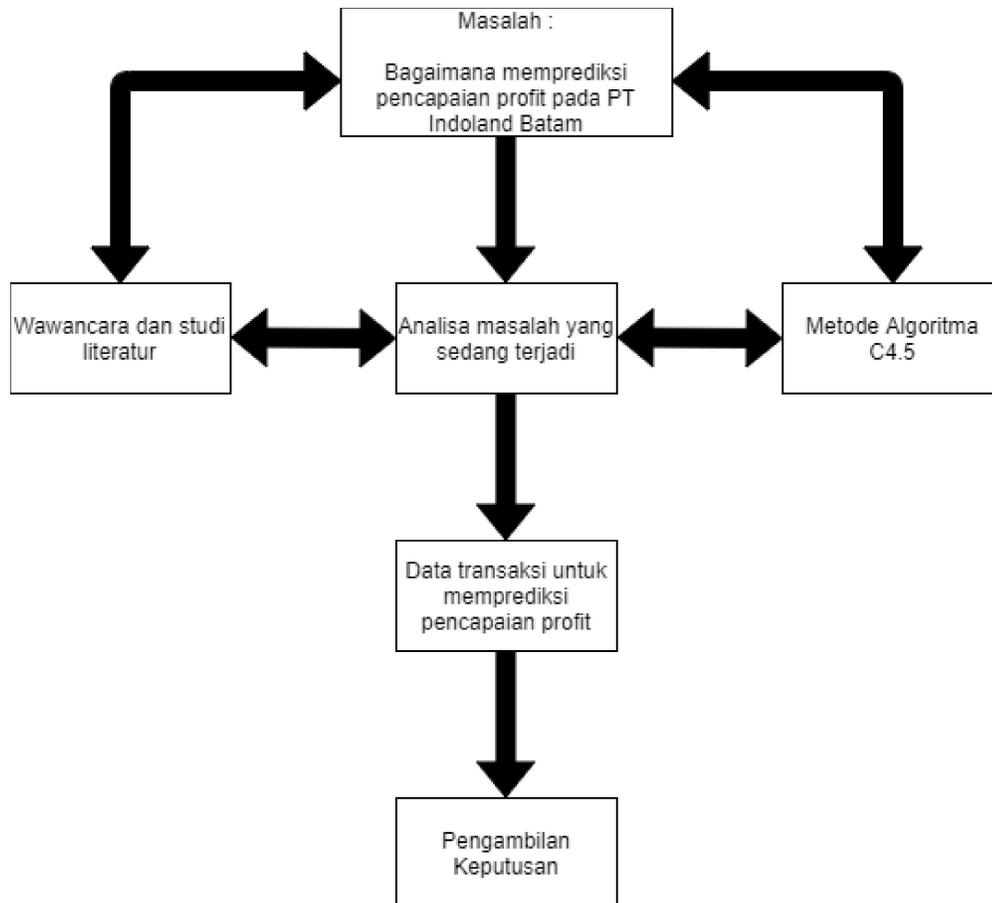
8. Sebuah studi tentang algoritma prediksi waktu classifier oleh (Rika Harman & Evan Rosiska, 2020) menunjukkan bahwa algoritma C4.5 membantu siswa menemukan dan mengidentifikasi penyebab masalah. Alasan utamanya adalah mahasiswa tidak bisa menyelesaikan semester. Belajar tepat waktu.
9. Berdasarkan penelitian (Yang and Chen, 2016) melakukan penelitian tentang "*Taiga: Performance Optimization of C4.5 Decision Tree Algorithms*", memberikan informasi tentang kemungkinan penggunaan jenis-jenis algoritma C4.5 yang terdapat dalam pengambilan keputusan. . Penyelesaian data berdasarkan kategori atau kriteria selama penambangan data dimaksimalkan, mengurangi waktu yang dibutuhkan untuk membuat keputusan akhir.
10. Sebuah studi yang dilakukan oleh (Augusto dan Char, 2014) tentang "*Performance improvement of datamining in WEKA through GPU acceleration*" memberikan informasi bahwa mungkin untuk mengadaptasi proses yang membutuhkan waktu lama untuk beradaptasi. Menggunakan GPU menunjukkan serangkaian kali. Mudah untuk mengadaptasi operasi kumulatif ke GPU. Saat memparalelkan metode perkalian matriks, metode ini dapat disesuaikan untuk memanfaatkan CPU dan GPU. Tergantung pada ukuran cetakan, metode ini dapat mempercepat proses hingga 49%.
11. Penelitian oleh (Tukino, 2019) tentang "Penerapan algoritma C4.5 untuk memprediksi keuntungan di PT. SMOE Indonesia" telah menunjukkan bahwa penggunaan algoritma C4.5 dapat membuat bantuan keputusan.

Sebuah sistem yang memprediksi keuntungan perusahaan didasarkan pada kriteria berikut. Seperti yang telah ditetapkan oleh perusahaan.

12. Dari penelitian “Penerapan Algoritma C.45 Untuk Kelayakan Penerima Uang Dan Beras” (Harman, 2018) melakukan survei tentang "Penerapan Algoritma C4.5 pada Mata Uang dan Kelayakan Penerima di Amerika Serikat" bahwa penggunaan algoritma Kelas C4.5 diharapkan dapat meningkatkan kebijakan kami. Setelah menyelesaikan studi ini, faktor kedekatan akan dipertimbangkan untuk memainkan peran yang lebih penting bagi penerima uang tunai dan beras di masa depan.

2.3. Kerangka Pemikiran

Berikut adalah gambar kerangka pemikiran yang akan diimplementasikan dalam penelitian ini:



Gambar 2.6 Kerangka Berpikir

(Sumber: Peneliti, 2021)