

BAB II

TINJAUAN PUSTAKA

2.1 Teori

Dalam subbab ini kita membahas teori, menafsirkannya sebagai sekumpulan bagian atau variabel yang saling terkait, definisi dan proposal yang bertujuan untuk menjelaskan fenomena alam dengan membangun hubungan antara variabel yang dipelajari dan memberikan penjelasan fenomena secara sistematis.

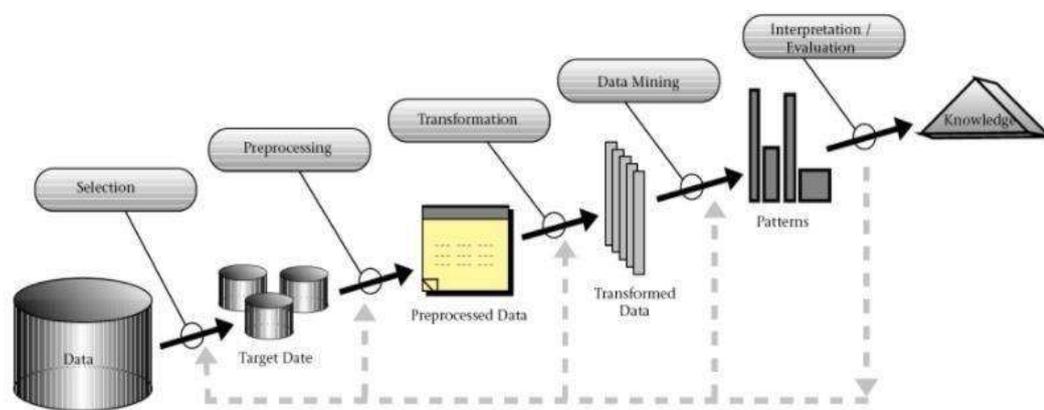
2.1.1 Teori Umum

Sub Bab ini akan membahas teori utama di balik karya penelitian ini, termasuk ide-ide utama.

2.1.1.1 *Data Mining*

Penambangan data adalah sebuah proses mencari pola-pola yang ada dan data dari sejumlah besar data yang disimpan dalam basis data atau sumber data lainnya. perbedaan; Sangat berguna untuk mengekstrak wawasan yang berguna seperti struktur dan cacat. Penambangan data banyak digunakan dalam waktu beberapa tahun terakhir karena ketersediaan informasi yang didapat dalam jumlah besar dalam bentuk elektronik dan kebutuhan untuk mengubah informasi ini menjadi informasi dan pengetahuan yang berguna dalam sistem besar. Aplikasi ini dapat ditemukan di berbagai bidang seperti kecerdasan buatan, pembelajaran mesin, analisis pasar, statistik, sistem basis data, manajemen bisnis, dan dukungan

keputusan. Penambangan data adalah proses menganalisis dan mengekstraksi pengetahuan secara otomatis menggunakan satu atau lebih teknik pembelajaran mesin. Tujuan dari data mining adalah untuk menemukan pola global dalam data yang ada. Untuk mengekstraksi pengetahuan dari konsensus yang ada, perlu menggunakan metode ilmiah yang dikenal sebagai *Knowledge Discovery in Database (KDD)* dalam database. (Sugara et al. 2018).



Gambar 2.1 Tahapan proses KDD

Sumber: Sugara 2018

Proses analisis data (KDD) terdiri dari 5 (lima) langkah, yang dapat diuraikan berikut ini :

1. *Selection*

Kesatu, dapat memilih data yang akan diolah dari beberapa sumber data dengan membersihkan data yang bertentangan atau nilai yang hilang.

2. *Preprocessing*

Kedua, mengintegrasikan semua sumber informasi yang dikumpulkan sebelum melanjutkan ke proses selanjutnya.

3. *Transformation*

Ketiga, ubah data menjadi format yang lebih cocok untuk penambangan data.

4. *Data Mining*

Gunakan pemrosesan data dan sistem untuk membangun model dari data yang dikumpulkan.

5. *Interpretation/ Evaluation*

Kelima, interpretasikan dan evaluasi pola yang diterima sehingga mengidentifikasi atau mewakili pola yang ingin terima.

Saat menggunakan KDD, teknik analisis data adalah mengekstraksi pola dari data. Skema dapat dibuat dalam operasi basis data. Secara umum, ada dua jenis metode penambangan data: metode analisis data, yang menggambarkan karakteristik utama dari data yang ada, dan metode penambangan data deskriptif, yang berupaya membuat prediksi berdasarkan data yang ada. Ekstraksi informasi dapat dilakukan dalam bentuk data kuantitatif, tekstual atau multimedia.

Aplikasi penambangan data dapat menggunakan parameter yang berbeda untuk menganalisis data. Ini termasuk korelasi (bagaimana satu faktor berhubungan dengan yang lain), konvergensi atau analisis jalur (bagaimana satu faktor mengarah ke yang lain), diferensiasi (menemukan desain baru dengan tujuan tertentu), dan integrasi (pengelompokan satu atau serupa)..

Dalam penambangan data, ekstraksi data dapat mencakup beberapa langkah dasar, yang dijelaskan di bawah ini:

1. *Problem definition*

Langkah pertama adalah menentukan tujuan. Bergantung pada tujuannya, seperangkat alat yang tepat dapat diterapkan pada data untuk memodelkan perilaku.

2. *Data exploration*

Jika kualitas data tidak cukup tinggi untuk model yang akurat, buat rekomendasi untuk data mendatang. Pada tahap ini teknik pengumpulan dan penyimpanan dapat diterapkan. Untuk keperluan analisis, semua data harus dikumpulkan agar dapat diolah secara konsisten.

3. *Data preparation*

Tujuan dari proses ini adalah untuk membersihkan dan memodifikasi untuk memperhitungkan nilai yang hilang dan salah, dan memperlakukan semua fakta yang diketahui secara konsisten untuk analisis yang lebih kuat.

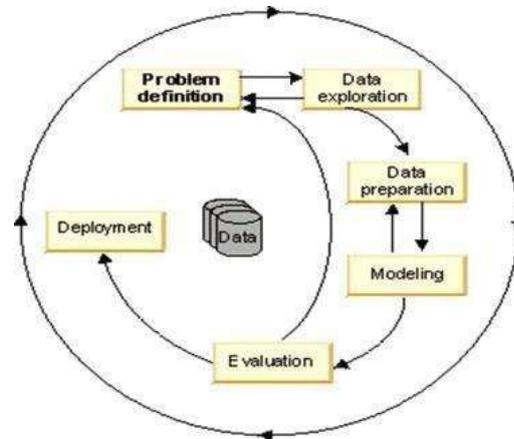
4. *Modeling*

Berdasarkan data dan hasil yang diinginkan, pilih algoritma penambahan data atau kombinasi algoritma untuk analisis. Algoritme ini berkisar dari teknik klasik seperti statistik, lingkungan, dan pengelompokan hingga teknik generasi berikutnya seperti pohon keputusan, jaringan, dan algoritme berbasis aturan. Pemilihan algoritma tertentu bergantung pada tujuan spesifik yang ingin dicapai dan kualitas data yang dianalisis.

5. *Evaluation and Deployment*

Berdasarkan hasil algoritma data mining, analisis dilakukan untuk menentukan kesimpulan utama dari analisis dan mengembangkan

seperangkat rekomendasi untuk dipertimbangkan.



Gambar 2.1 Representasi proses data mining

Sumber: Sugara 2018

2.1.1.2 Algoritma C4.5

Algoritma C4.5 adalah program yang menerima sinyal dan menghasilkan algoritma keputusan. Pohon keputusan ini divalidasi terhadap data yang belum diuji untuk pelatihan. Program membuat aturan klasifikasi menggunakan pohon keputusan untuk kumpulan data tertentu. Sistem C4.5 merupakan perpanjangan dari sistem ID3 asli yang dikembangkan oleh Quinlan. C4.5 adalah salah satu kurva pembelajaran. Algoritma C4.5 menggunakan konsep entropi informasi untuk membangun pohon keputusan dari set pelatihan, mirip dengan algoritma ID3. C4.5 Klasifikasi statistik. (Sugara et al. 2018)

Algoritma dapat memecahkan masalah dalam sistem dengan membuat keputusan berdasarkan langkah-langkah berikut:

1. Pilih perangkat sebagai *root*.
2. Buat cabang untuk setiap rekaman aset.
3. Bagilah benda-benda menjadi cabang-cabang,

4. Ulangi proses untuk setiap cabang sampai semua kasus di cabang terselesaikan dengan baik.

Nilai atribut dihitung dengan memilih atribut sebagai baris akar. Nilai keuntungan maksimum digunakan sebagai dasar dari pohon keputusan. Rumus ini digunakan untuk menghitung nilai keuntungan:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Rumus 1. Rumus *gain*

Keterangan:

S= Himpunan Kasus

A= Atribut

n= Jumlah partisi atribut A

|Si| = Jumlah kasus pada partisi ke-i

|S| = Jumlah kasus dalam S

Sementara itu, perhitungan nilai entropi dapat dilihat pada persamaan 2 berikut ini.

$$Entropy(S) = - \sum_{i=1}^n p_i * \log_2 p_i$$

Rumus 2. Perhitungan Nilai *Entropy*

Keterangan:

S = Himpunan Kasus A = Fitur

n = Jumlah partisi S

Pi = Proporsi dari Si terhadap S

Setelah perhitungan selesai, pohon keputusan dibuat, yang dapat diubah menjadi aturan dengan aturan *IF-THEN-ELSE*. Aturan-aturan tersebut dijelaskan

sebagai berikut:

1. *IF-THEN Rule*

Aturan Induktif: Proses menurunkan aturan "jika maka" yang berguna dari data berdasarkan signifikansi statistik. Sistem berbasis aturan membuat seperangkat aturan jika-maka. Pengetahuan sudah merupakan kumpulan bentuk di mana strategi penyimpanan dapat dirancang. Semua data yang digunakan untuk analisis harus dikumpulkan sehingga dapat diproses secara terus menerus.

2. *IF* kondisi *THEN* Kesimpulan

Aturan ini memiliki dua bagian. Perintah (klausa *IF*) mendefinisikan satu atau lebih kondisi di sekitar nilai prediksi yang berisi prediksi tentang nilai properti target berdasarkan atribut perintah (klausa *THEN*). Evaluasi yang akurat dari nilai atribut objektif meningkatkan pengambilan keputusan. Aturan umum untuk estimasi *IF-THEN* dalam data mining. Mereka memberikan abstraksi ekstrim dari fungsi pencarian. Metode penalaran berbasis aturan dapat membuat prediksi dengan kasus yang dikecualikan.

2.1.1.3 Pohon Keputusan (*decision tree*)

Pohon keputusan adalah representasi pengetahuan yang diatur oleh node dan cabang di pohon, sehingga setiap node non-teks menggunakan atributnya. Label domain dari node internal berisi nilai atribut dari node tersebut. Setiap node disebut kelas (nilai target karakteristik). Model berbasis pohon, termasuk klasifikasi dan regresi pohon, adalah aplikasi umum model induktif. Model pohon keputusan paling cocok untuk pemrosesan data. Mereka lebih murah untuk dibangun, lebih

mudah diinterpretasikan, lebih mudah diintegrasikan dengan sistem basis data, dan memiliki akurasi implementasi yang sama atau lebih baik. Ada banyak algoritme pohon keputusan seperti algoritme HUNTS (ini adalah salah satu algoritme paling awal), CART, ID3, C4.5 (versi terbaru dari algoritme ID3), SLIQ, SPRINT. (Sugara et al. 2018)

Dalam pembelajaran pohon keputusan, ID3 (*Iterative Dichotomizer 3*) adalah algoritma yang dikembangkan oleh Ross Quinlan untuk mengekstraksi pengetahuan keputusan dari dataset. ID3 sering digunakan dalam pembelajaran mesin dan pemrosesan bahasa alami. Metode pengurangan termasuk membangun pohon yang menunjukkan proses klasifikasi.. Setelah pohon dibangun, itu diterapkan ke setiap grup dalam database dan membuat taksonomi untuk grup itu. Masalah yang dihadapi oleh sebagian besar algoritma pohon keputusan adalah sebagai berikut:

1. Memilih atribut pemisahan
2. Memesan atribut pemisahan
3. Jumlah perpecahan yang harus diambil
4. Saldo struktur pohon dan pemangkasan
5. Menghentikan kriteria

2.1.2 Teori Khusus

Dalam sub bagian ini, kami membahas prinsip-prinsip khusus yang menarik terkait dengan pelaksanaan penelitian ini.

2.1.2.1 Rumah Layak Huni

Untuk menentukan apa yang dimaksud dengan habitat layak huni, peneliti

menerapkan prinsip-prinsip Hukum Perpajakan No. 1 tahun 2011 Bagian 1 tentang *real estat*, yang mendefinisikan *real estat* sebagai rencana penting, pengelolaan perumahan, pengelolaan perumahan, pemeliharaan dan perbaikan. Pencegahan dan perbaikan permukiman kumuh dan kualitas hidup di permukiman kumuh, kebijakan pertanahan, pembiayaan dan sistem pembayaran serta peran masyarakat. UUD No. 1 Pasal 19 Tahun 2011 perumahan dan pengenalan perumahan dalam rangka memenuhi kebutuhan perumahan, salah satu kebutuhan dasar manusia, untuk memajukan kesejahteraan dan pemerataan masyarakat, sebagaimana disebutkan dalam pasal 1, pemerintah ini, yang dikelola oleh suatu daerah. Pemerintah dan/atau seluruh rakyat harus menikmati hidup dalam lingkungan yang sehat, aman, serasi, dan tertib serta merupakan hak setiap warga negara.

2.1.2.2 Indekos

Indekos adalah tempat tinggal sementara yang disewakan kepada wisatawan atau orang yang tinggal jauh dari tempat tujuan biasanya, terutama kepada pekerja atau pekerja yang jauh dari tempat kerjanya. Seiring pertumbuhan populasi, mereka yang ingin tinggal di rumah sering menghadapi tantangan. Banyak kendala internal adalah banyaknya pemangku kepentingan. Di luar itu, pencari kerja di pesantren kebanyakan adalah orang yang baru pindah rumah, berasal dari berbagai daerah dan belum mengenal daerah Batam, sehingga perlu diketahui secara lengkap tentang pesantren tersebut. (Rianto 2019).

2.1.2.3 Kelayakan

Studi kelayakan atau feasibility study ini adalah proses analisis mendalam terhadap suatu proyek atau perusahaan atau bisnis untuk menentukan apakah bisnis dapat dilanjutkan.

Kelayakan berarti analisis mendalam untuk menentukan apakah manfaat bisnis lebih besar daripada biayanya. Dengan kata lain, kapasitas dapat diartikan bahwa usaha yang dikelola akan memberikan keuntungan finansial dan non finansial sesuai dengan tujuan yang dituju. Ditafsirkan dengan benar di sini, bermanfaat tidak hanya bagi perusahaan yang mengelolanya, tetapi juga bagi investor, pemberi pinjaman, pemerintah, dan masyarakat secara keseluruhan. (R Nурmalina, T Sarianti 2018).

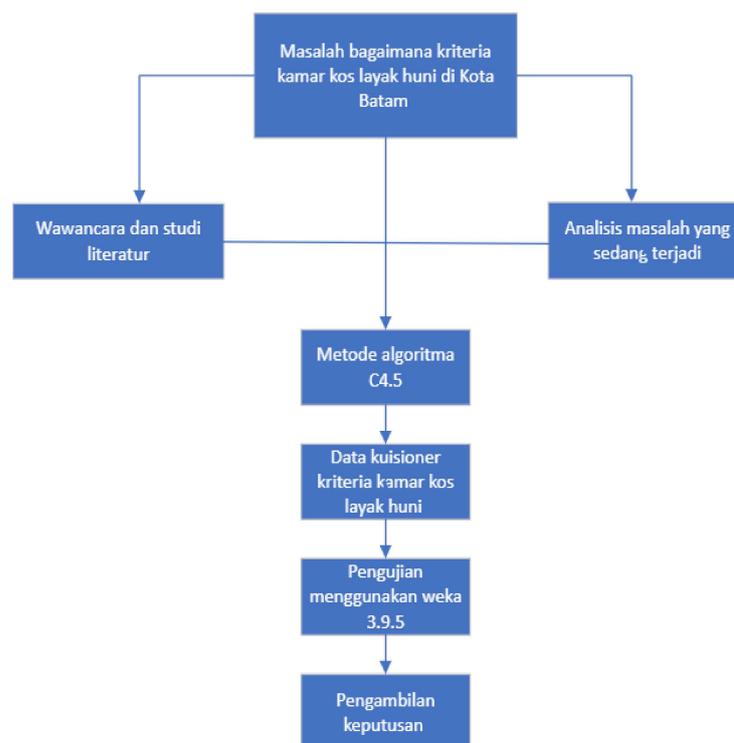
2.1.2.4 Software Weka

Weka (*Waikato Environment for Knowledge Analysis*) Lingkungan Analisis Pengetahuan Waikato dikembangkan oleh Universitas Waikato di Selandia Baru untuk penelitian, pendidikan, dan berbagai aplikasi. Weka mampu memecahkan masalah penambangan data dunia nyata, terutama masalah klasifikasi, yang merupakan inti dari teknik pembelajaran mesin. Perangkat lunak ini ditulis menggunakan pendekatan berorientasi objek dalam hierarki kelas Java dan berjalan di hampir semua platform. Weka mudah digunakan dan diimplementasikan pada beberapa tingkatan. Ada implementasi algoritme pembelajaran modern yang dapat diterapkan ke database dari baris perintah. WEKA meliputi preprocessing data, klasifikasi, regresi, pengelompokan, aturan asosiasi, dan visualisasi. Pengguna dapat memproses data terlebih dahulu, memasukkannya ke dalam program pembelajaran, dan menganalisis serta melakukan klasifikasi yang dihasilkan tanpa

menulis kode program apa pun. Inferensi adalah salah satu dari banyak teknik data mining yang dapat digunakan di pohon keputusan (Cobit and Utami 2019)

2.2 Kerangka Pemikiran

Model kerangka pemikiran berdasarkan penelitian yang dilakukan, yaitu sebagai berikut:



Gambar 2.2 Kerangka pemikiran

Sumber: (Penulis, 2022)

2.3 Hipotesis Penelitian

Berikut ini adalah hipotesis dari hasil penelitian berdasarkan dugaan penulis penelitian ini:

1. Teknik data mining dengan perhitungan algoritma C4,5 dapat menganalisa kriteria kamar kos layak huni di Kota Batam.
2. Hasil dari penelitian ini dapat mengetahui kriteria kamar kos yang diminati oleh karyawan di Kota Batam.

2.4 Penelitian Terdahulu

Berikut ini adalah informasi yang berupa referensi dari penelitian terdahulu yang berkaitan tentang penelitian ini yang di rangkum sebagai berikut:

1. Dengan judul “Penerapan *data mining* menggunakan algoritma c4.5 dalam menentukan kelayakan penerima bantuan bedah rumah pada desa tiga dolok” yang dilakukan oleh (V.M. et al. 2021) penelitian ini menunjukkan bahwa menentukan penerima bantuan bedah rumah pada desa tiga dolok dengan menggunakan algoritma c4.5 akan bermanfaat sekali dalam proses menentukan penerima bantuan bedah rumah.
2. Dengan judul “Penerapan data mining untuk prediksi penjualan Readymix menggunakan metode algoritma c4.5 pada PT.Remicon widyaprima” yang dilakukan oleh (Harman 2020) penelitian ini menunjukkan bahwa Mencari perkiraan penjualan yang telah disiapkan dihitung menurut algoritma c4.5 menggunakan metode data mining, menggunakan data penjualan di PT. Remicon Vidiaprim sebagai referensi dan panduan yang memungkinkan untuk pohon keputusan.

3. Dengan judul “Penerapan algoritma c4.5 untuk deteksi dini gangguan autisme pada anak” yang dilakukan oleh (Sugara et al. 2018) penelitian ini menunjukkan deteksi dini autisme pada anak dapat diprediksi dengan menggunakan metode data mining yang menggunakan algoritma C 4.5 untuk memprediksi autisme pada anak berdasarkan gejala yang dialami anak tersebut.
4. Dengan judul “Perbandingan algoritma C4.5 dan *Naïve Bayes* untuk menentukan kelayakan penerima bantuan program keluarga harapan” (Fitriani 2020) penelitian ini menunjukkan bahwa algoritma C4.5 merupakan metode yang cukup baik dalam memprediksi kelayakan warga dalam menerima bantuan Program Keluarga Harapan.
5. Dengan judul “Membangun model algoritma C4.5 dalam menentukan kelayakan rumah ibadah berdasarkan indicator epidemiologi pada masa pandemic covid-19” (Amaliyyah 2021) penelitian ini menunjukkan bahwa berdasarkan dataset perkembangan kasus covid-19 perkelurahan, maka terbentuklah model C4.5 dalam bentuk rule dan pohon keputusan yang dapat digunakan untuk menentukan kluster terkait Kelayakan rumah ibadah.
6. Dengan judul “Aplikasi berbasis web untuk klasifikasi kelayakan pemberian kredit pada koperasi Bumdes Mitra Baru menggunakan algoritma C4.5 (Ulfa 2020) penelitian ini menunjukkan bahwa akurasi model pohon keputusan yang di hasilkan algoritma C4.5 sangat akurat.
7. Dengan judul “Komparasi algoritma C4.5 dan *naïve bayes* untuk

menganalisa kelayakan pengajuan kredit” (Khasanah 2019) penelitian ini menunjukkan bahwa untuk penentuan kelayakan kredit apakah calon nasabah akan menjadi nasabah yang pembayarannya lancar atau bermasalah dengan menggunakan algoritma C4.5 dan naïve bayes.

8. Dengan judul “Penerapan algoritma C4.5 untuk penentuan kelayakan pemberian kredit” (Santoso and Sekardiana 2019) penelitian ini menunjukkan bahwa dengan hasil evaluasi penelitian bahwa algoritma C4.5 mendapatkan nilai akurasi data training serta data testing menggunakan algoritma C4.5.
9. Dengan judul “Penerapan algoritma C4.5 untuk kelayakan penerimaan uang dan beras” (Rika Harman 2018) penelitian ini menunjukkan bahwa dihasilkannya keputusan yaitu berupa perhitungan jumlah kepala-kepala keluarga yang menerima bantuan uang dan beras dan yang tidak menerima bantuan uang dan beras yang mana dalam proses perhitungan node cukup satu kali perhitungan saja.
10. Dengan judul “Analisis Algoritma C4.5 untuk menentukan factor pembelian sepeda bekas pada took sepeda Batam” (Internasional and Ananda 2021) penelitian ini menunjukkan bahwa dihasilkannya keputusan yaitu berupa sejumlah opsi aktivasi dapat dibuat untuk mendemonstrasikan hal ini, termasuk: a). Jika roda sepeda bengkok, sepeda tidak akan dibawa; b) Jika roda sepeda tidak bengkok atau normal dan setang sepeda berayun, sepeda tidak akan dibeli; dan c). Jika roda sepeda bengkok atau patah, dan setang sepeda tidak bergerak atau terasa

nyaman, maka sepeda akan dibeli.