

BAB II

TINJAUAN PUSTAKA

2.1 Teori

Pada subbab ini akan dilakukan penjelasan mengenai maksud dan definisi yang berhubungan dengan judul dari penelitian ini dengan melakukan pembahasan mengenai beberapa logika dasar atau argumentasi dasar yang dijadikan fundamen pemahaman bagi peneliti.

2.1.1 Teori Umum

Pada subbab ini akan dilakukan penjelasan mengenai beberapa teori yang menjadi fundamen pemahaman bagi peneliti yang dirasa diperlukan guna melengkapi atau memperkuat argumentasi hasil kesimpulan yang hendak dicapai melalui penelitian ini yang bersifat umum dan luas.

2.1.1.1 KDD (*Knowledge Discovery In Database*)

KDD ialah singkatan dari *Knowledge Discovery In Database* yang memiliki definisi yakni satu dari banyaknya kaidah memperoleh informasi atau rekognisi menggunakan evidensi-evidensi yang ditelaah dari database atau evidensi yang tersimpan pada suatu tempat. Informasi-informasi yang telah diperoleh selanjutnya dijadikan dasar pengetahuan guna mengambil putusan-

putusan terkait dengan permasalahan. Adapun di bawah ini ialah langkah-langkah atau tahap-tahap pada *Knowledge Discovery In Database*, yakni :

1. *Data Selection*

Tahapan yang terlebih dahulu harus dilakukan ialah menghimpun segenap embaran atau evidensi yang dieksekusi dengan melakukan penyeleksian-penyeleksian tertentu sehingga terbentuk sebuah sumber data yang bernilai kredibel kemudian masuk kepada pengolahan *data mining*..

2. *Pre-processing* atau *Cleaning*

Tahapan selanjutnya ialah langkah antisipasi yang berguna meniadakan adanya evidensi berganda melalui pemeriksaan-pemeriksaan tertentu merujuk pada kepentingan-kepentingan yang mungkin muncul ketika melakukan proses penganalisisan data karena adanya evidensi-evidensi yang mungkin bertambah seiring dengan dilakukannya sebuah penelitian.

3. *Transformation*

Kemudian dilanjutkan dengan tahapan merubah serta menyesuaikan pola-pola penyimpanan basis data.

4. *Data Mining*

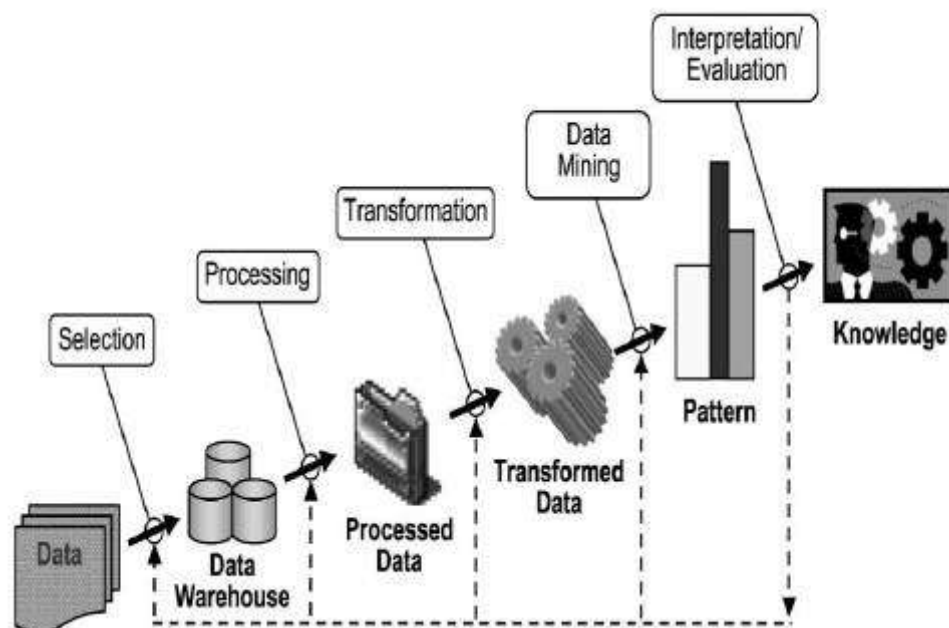
Tahapan yang menjadi sangat penting dari serangkaian tahapan yang telah dilakukan sebelumnya yakni menentukan daya upaya pada proses pencaharian kemudian terbentuknya sebuah kaidah-kaidah tertentu yang memanifestasikan embaran-embaran kemudian menuntun ke arah

pengambilan putusan-putusan sesuai dengan pengidentifikasian masalah yang hendak dipecahkan.

5. *Interpretation*

Tahapan terakhir dari semua tahapan yang ada yakni pemberian kesan atau taksiran dari grafik atau gambaran naik turunnya sebuah fenomena yang telah dianalisis dari *data mining* sehingga mudah dinalar berdasarkan pemahaman-pemahaman sederhana terlebih lagi evidensi yang sesuai dengan tujuan penelitian (Sukma, 2019).

Adapun proses KDD atau *Knowledge Discovery in Database* mampu digambarkan seperti dibawah ini (Sukma, 2019) :

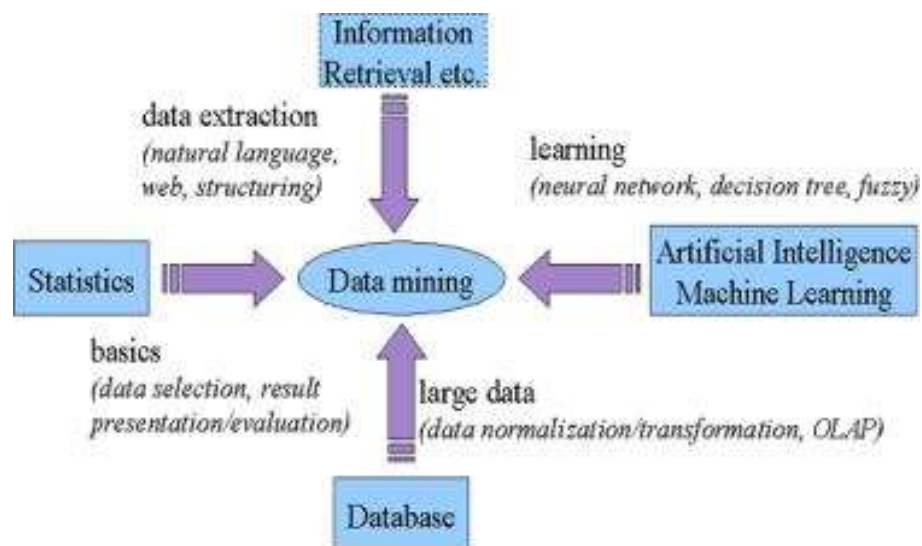


Gambar 2.1 Proses *Knowledge Discovery in Database*

Sumber: (Sukma, 2019)

2.1.1.2 Data Mining

Proses yang disebut juga penemuan informasi yang awalnya tidak dimengerti yang merupakan penemuan yang berasal dari sebuah perkumpulan informasi yang besar dapat dikatakan sebagai data mining yang mengacu pada suatu bidang ilmu yang penting dan dapat menyajikan kenyataan suatu skema yang memiliki keterkaitan suatu perkumpulan informasi yang luas. Pengertian selain yang telah dijelaskan diatas dimana suatu perangkaian sistematis yang mempergunakan beberapa teknik studi bidang komputerisasi dalam analisa dan penyajian informasi secara otomatis yang dapat dipahami juga sebagai suang perangkaian sistematis yang dipergunakan untuk mendapatkan informasi yang masih kurang agar informasi yang kurang tersebut dengan menggunakan bidang ilmu tidak perlu dicari secara tradisional (Cynthia & Ismanto, 2018).



Gambar 2.2 Bidang Ilmu *Data Mining*

Sumber: (Mardi, 2019).

2.1.1.3 Pengelompokan *Data Mining*

Data Mining sendiri terdiri dari beberapa hierarki-hierarki yang mampu dijabarkan sehingga memperjelas langkah-langkah satu demi satu yang harus dijalani sehingga menemukan interpretasi yang baik dalam pengambilan keputusan (Wiwi Verina, 2020), yakni:

1. *Description* atau Deskripsi

Langkah awal yang menjadi pintu gerbang dari semua hal yang akan dilakukan ialah menganalisis melalui proses mendeskripsikan ciri-ciri tertentu dari evidensi-evidensi yang terdapat dalam database. Proses ini dapat dianalogikan seperti seorang yang bertugas mengumpulkan suara pada sebuah proses *election* atau eleksi tanpa terpengaruh oleh informasi-informasi perihal kebaikan dan keburukan atau hal-hal yang mampu mengganggu elektabilitas dari orang yang akan terpilih. Dari analogi yang dipaparkan memiliki arti bahwa deskripsi yang akan dilakukan oleh peneliti ialah menyajikan dan memaparkan informasi-informasi terkait evidensi terkait ke dalam sebuah pola yang terstandarisasi sehingga memudahkan proses selanjutnya yang akan dilakukan oleh peneliti.

2. *Estimation* atau Estimasi

Tahap lanjutan dari deskripsi ialah mengestimasi atau lebih mudah diartikan menjadi proses mengklasifikasi evidensi yang satu dengan evidensi yang lainnya tetapi hal ini tidak sama ketika proses pengestimasian dilakukan pada evidensi bernilai angka. Arketipe harus

terbentuk dari evidensi-evidensi terdeskripsi yang bernilai angka yang juga tidak terlepas dari variabel yang akan diteliti sehingga mampu menghasilkan angka juga yang memudahkan proses memprediksi hasil. Selanjutnya diberikan hasil-hasil numerik sehingga mampu mewakili atau menggambarkan hasil yang didapat dari fenomena yang terjadi yang sedang diteliti atau diamati. Proses ini mampu dianalogikan seperti nilai rapor yang diterima seorang siswa yang mencerminkan hasil pembelajaran selama satu semester. Angka atau nilai yang tertulis pada rapor mengestimasi bagaimana pemahaman siswa tersebut menyerap atau memahami sebuah mata pelajaran kemudian diakumulasi kemudian dibandingkan dengan yang lainnya sehingga didapat peringkat antara satu siswa dengan siswa lainnya hal ini juga mampu dikatakan mempermudah proses membandingkan kemampuan satu siswa dengan siswa lainnya..

3. *Prediction* atau Prediksi

Tahapan selanjutnya yang dilakukan ialah memprediksi arah dari hasil evidensi yang telah di deskripsikan dan diestimasi pada proses sebelumnya tetapi proses ini lebih menitikberatkan pada apa yang akan terjadi di waktu mendatang sehingga menjadi bentuk antisipasi dan menjadi indikator tertentu dalam penentuan keputusan. Apabila secara langsung dihubungkan dengan penelitian yang akan dilakukan oleh peneliti sekarang, kegiatan memprediksi dilakukan dengan keperluan seperti dibawah ini :

- a. Ditemukan peluang-peluang adanya kenaikan atau penurunan pada Indeks Harga Perdagangan Besar bahan bangunan serta yang melengkapi untuk bulan bahkan tahun-tahun mendatang.
- b. Ditemukan peluang-peluang adanya kenaikan atau penurunan pada Indeks Harga Perdagangan Besar bangunan-bangunan jadi untuk bulan bahkan tahun-tahun mendatang.
- c. Ditemukan peluang-peluang adanya kenaikan atau penurunan pada Indeks Harga Perdagangan Besar pekerjaan-pekerjaan umum terutama yang berhubungan dengan listrik dan air bersih atau minum untuk bulan bahkan tahun-tahun mendatang.

4. *Classification* atau Klasifikasi

Pengklasifikasian yang dilakukan atas evidensi-evidensi setelah melalui ketiga proses sebelumnya dengan standar tertentu yang telah ditetapkan sebelumnya sehingga pengklasifikasian terorganisir dengan baik dan memenuhi kriteria yang tepat. Proses ini mampu dianalogikan seperti mengklasifikasikan masyarakat yang layak menerima Bantuan Langsung Tunai (BLT) berdasarkan penghasilan yang diterima masing-masing orang dari pemberi kerja. Masyarakat digolongkan menjadi masyarakat berpenghasilan tinggi, menengah, dan bawah. Apabila secara langsung dihubungkan dengan penelitian yang akan dilakukan oleh peneliti sekarang, kegiatan memprediksi dilakukan dengan keperluan seperti dibawah ini :

- a. Diklasifikasikan Indeks Harga Perdagangan Besar bahan bangunan contohnya baja beton antara toko yang satu dengan toko lainnya.
- b. Diklasifikasikan Indeks Harga Perdagangan Besar bangunan dengan tipe 36 antara satu cluster dengan cluster yang lainnya.
- c. Diklasifikasikan Indeks Harga Perdagangan Besar pasokan listrik dan air antara satu daerah dengan daerah yang lainnya.

5. *Clustering* atau Pengklusteran

Kemudian langkah selanjutnya ialah meng-kluster-kan atau mengelompokkan evidensi-evidensi yang telah diklasifikasikan menjadi satu berdasarkan adanya ciri-ciri yang sama. Proses *clustering* tidak memiliki target tertentu sehingga hanya menitikberatkan pada kesamaan ciri, arkitipe, dan pola dimana apabila semakin mirip antara satu evidens dengan evidens lainnya maka akan menghasilkan angka yang maksimum begitu juga kebalikannya semakin bertolak belakang corak antara satu evidens dengan evidens lainnya akan mengakibatkan adanya nilai yang minimum yang tentunya dipengaruhi juga dengan perhitungan algoritma yang digunakan. Apabila secara langsung dihubungkan dengan penelitian yang akan dilakukan oleh peneliti sekarang, kegiatan meng-kluster-kan dengan keperluan seperti dibawah ini :

- a. Bahan bangunan yang diprediksi akan cenderung naik, menghasilkan bangunan yang cenderung mahal, ditargetkan pada konsumen berkemampuan tinggi sehingga biaya promosi tidak perlu

terlalu besar yang mengakibatkan hasil yang maksimal terutama tingkat keberhasilan menjual rumah.

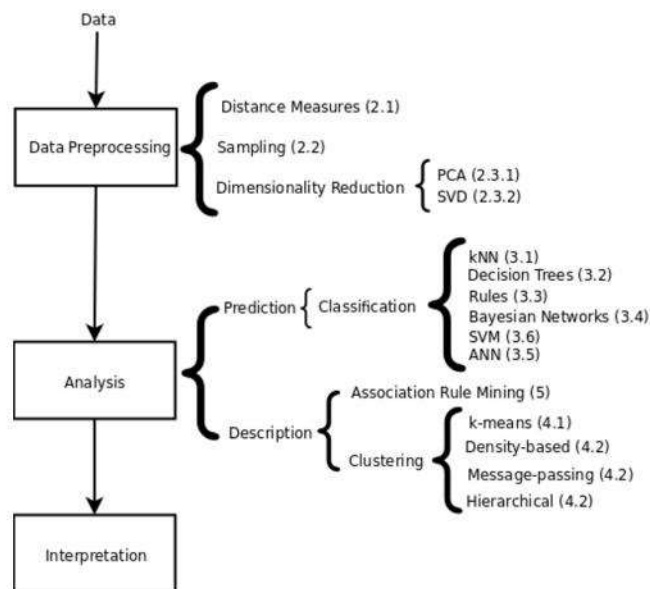
- b. Bahan bangunan yang diprediksi akan cenderung turun, menghasilkan bangunan yang cenderung terjangkau, ditargetkan pada konsumen berkemampuan menengah sampai bawah sehingga biaya promosi tidak perlu terlalu besar yang mengakibatkan hasil yang maksimal terutama tingkat keberhasilan menjual rumah.

6. *Association* atau Asosiasi

Tahapan yang terakhir ialah mengasosiasikan informasi-informasi yang mampu melengkapai atau mendukung sebuah evidensi sehingga menghasilkan putusan yang lebih lengkap. Apabila secara langsung dihubungkan dengan penelitian yang akan dilakukan oleh peneliti sekarang, kegiatan mengasosisasi dengan keperluan seperti dibawah ini:

- a. Bangunan yang berhasil terjual dimintai testimoni mengapa akhirnya konsumen memutuskan untuk membeli bangunan tersebut dengan penilaian dan sudut pandang masing-masing konsumen sehingga mampu dijadikan peningkatan kegiatan penjualan di masa mendatang.
- b. Bangunan yang tidak berhasil terjual dimintai kritik dan saran mengapa akhirnya konsumen memutuskan untuk tidak membeli bangunan tersebut dengan penilaian dan sudut pandang masing-masing konsumen sehingga mampu dijadikan perbaikan kegiatan penjualan di masa mendatang.

Poin demi poin tahapan pada proses pelaksanaan teknik *data mining* yang dijabarkan pada bagaian sebelumnya, mampu diperkuat dengan diagram atau digambarkan secara sederhana seperti di bawah ini.

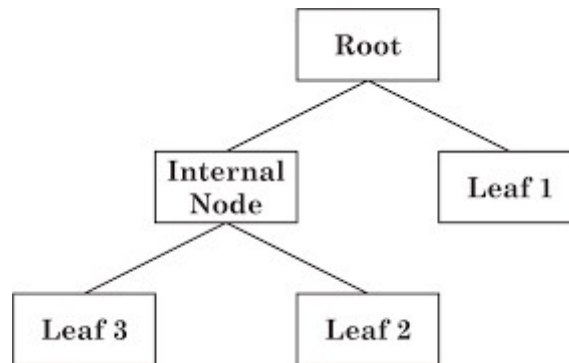


Gambar 2.3 Pengelompokan Teknik *Data Mining*

Sumber: (Halfis, 2019)

2.1.1.4 *Decision Tree*

Dalam standar global keseharian setiap manusia pasti memiliki beragam permasalahan yang sulit untuk dihadapi hingga semakin berkembangnya teknologi manusia mulai mewujudkan suatu permodelan yang terstruktur guna menunjang setiap orang dalam menyelesaikan beragam permasalahan yang ada dimana suatu permodelan yang berhasil ditemukan adalah pohon keputusan yang sering disebut juga dengan decision tree yang digunakan sebagai alat untuk menguraikan permasalahan yang ada dan dapat membantu dalam pengambilan keputusan dari hasil keputusan dari pohon tersebut (Muhammad Rizal, 2019)



Gambar 2.4 Model *Decision Tree*

Sumber: (Takalapeta, 2018)

2.1.1.5 Algoritma C4.5

Penggunaan angka atau kategori dalam melaksanakan pengklasifikasian informasi yang ada sangat efektif dimana pada saat setiap tingkat pemrosesan informasi dilakukan dapat menghasilkan sebuah permodelan yang dapat digunakan untuk mengatur berbagai Langkah penentuan yang berisi informasi yang sudah berlalu hingga informasi yang berlangsung saat ini (Sintia et al., 2019).

Tahapan pendahuluan pada *decision tree* yakni mengklasifikasikan informasi atau embaran atau eksplanasi dari evidensi-evidensi melalui penetapan nilai paling tinggi dari kesemestaaan nilai yang terdapat pada eksplansi evidensi terkait. Kemudian mampu dirumuskan guna mendapatkan nilai gain tertinggi, antara lain seperti dibawah ini:

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

Rumus 2.1 Perhitungan *Gain*

simbol-simbol yang terdapat pada rumus diatas mampu dijelaskan seperti di bawah ini, antara lain:

S mewakili himpunan

A mewakili keterangan atau kelengkapan yang digunakan.

n mewakili jumlah partisi yang terdapat pada keterangan atau kelengkapan bagian A

$|S_i|$ mewakili jumlah dari kasus yang terdapat dalam partisi yang ke- i

$|S|$ mewakili jumlah dari kasus yang terdapat dalam S

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

Rumus 2.2 Perhitungan *Entropy*

simbol-simbol yang terdapat pada rumus diatas mampu dijelaskan seperti di bawah ini, antara lain:

S mewakili himpunan

A mewakili keterangan atau kelengkapan yang digunakan.

n mewakili jumlah partisi yang terdapat pada keterangan atau kelengkapan bagian S p_i adalah proporsi dari S_i kepada S

2.1.2 Teori Khusus

Peneliti juga melakukan pembahasan atas beberapa logika dasar atau argumentasi dasar yang secara khusus bersinggungan langsung dengan variabel-

variabel yang diambil oleh peneliti secara spesifik membahas definisi-definisi dari variabel-variabel tersebut.

2.1.2.1 Penjualan

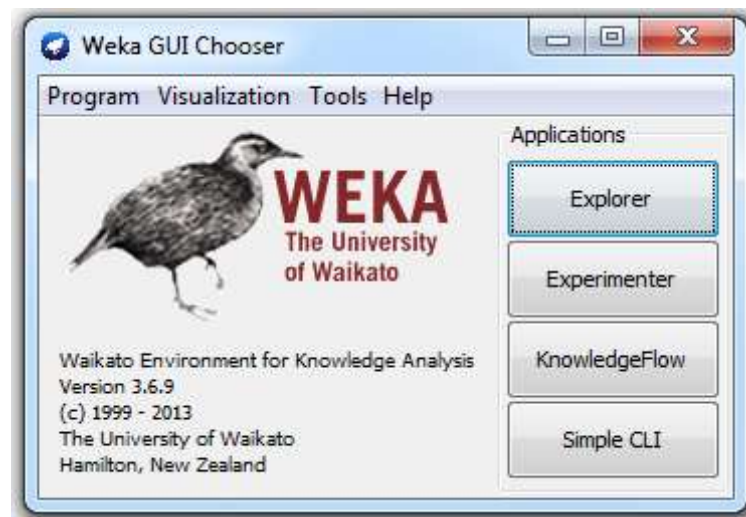
Penjualan adalah pelaksanaan penukaran properti dengan uang dimana dari suatu usaha memberikan barang agar usaha tersebut dapat mendapatkan uang dari orang yang menginginkan property tersebut. Sebuah entitas yang memperdagangkan barang dan jasa dapat dikatakan juga sebagai instansi yang bergerak dibidang bisnis transaksi jual beli yang dapat dipahami juga dengan property yang telah didapatkan dapat diolah yang kemudian dapat disebarakan atau dilakukan pengdistribusian property oleh instansi yang telah dijelaskan sebelumnya (Sandrawira Anggraini, Sarjon Defit, 2018).

2.1.2.2 *Waikato Environment for Knowledge Analysis (WEKA)*

Waikato Environment for Knowledge Analysis atau lebih dikenal oleh masyarakat umum dengan sebutan WEKA, ialah paket tools machine learning praktis. WEKA sendiri pertama kali lahir di Universitas Waikato bertempat di Selandia Baru dan sampai hari masih terus dipugar atau diperbaharui atau di-*update* karena sangat bermanfaat bagi siapapun untuk meneliti. WEKA juga adalah satu dari banyaknya perangkat lunak yang memiliki sifat *open source* atau memiliki kemudahan bagai siapa saja yang ingin memasang atau meng-*install* aplikasi ini di perangkat keras mereka masing-masing seperti komputer, laptop, ataupun tablet. (Azwanti, 2018).

WEKA sendiri juga dikhususkan menjadi *problem solver* bagi siapapun yang mengalami kendala dalam penyelesaian permasalahan di bidang *data mining*

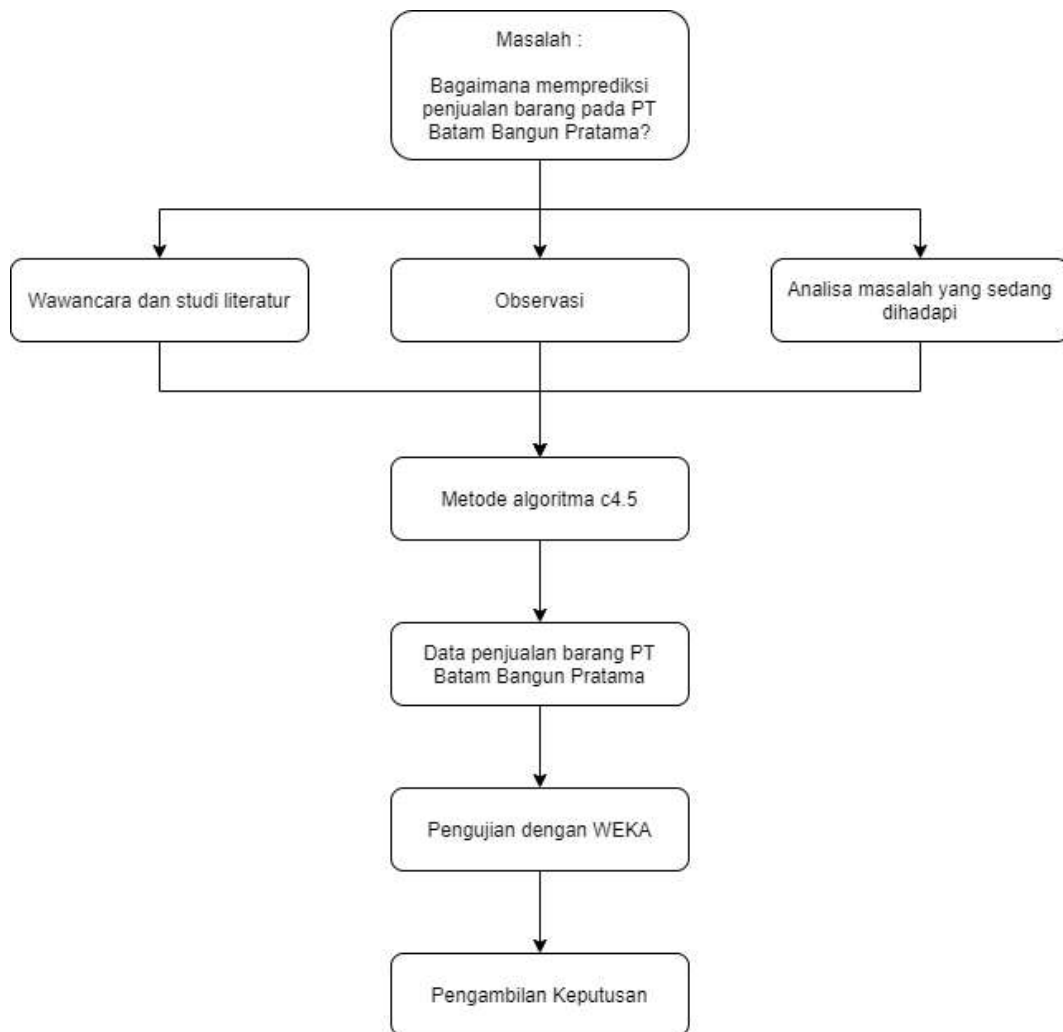
terlebih lagi membantu dalam pengklasifikasian masalah yang memiliki pendekatan *machine learning*. WEKA sudah dilengkapi dengan tahapan-tahapan penganalisisan *data mining* seperti yang dijabarkan pada bagian 2.1.1.3 bahkan WEKA mampu memvisualkan hasilnya. Hasil uji yang dihasilkan oleh WEKA memiliki fleksibilitas juga keefektifitasan yang cukup mempuni dikarenakan *decision tree* yang dihasilkan dari aplikasi ini memiliki tingkat kemiripan hasil yang cukup tinggi ketika seseorang melakukan pengujian dengan cara menghitung manual (Azwanti, 2018).



Gambar 2.2 Tampilan Utama WEKA

2.2. Kerangka Pemikiran

Adapun penelitian ini membentuk kerangka pemikiran sehingga menjadi gambaran plot logika yang dilakukan oleh peneliti guna menerapkan langkah-langkah pembuktian sampai memperoleh hasil yang dijadikan putusan-putusan, antara lain mampu ditelaah dari diagram di bawah ini :



Gambar 2.3 Kerangka Berpikir

Sumber: (Peneliti, 2021)

2.3. Hipotesis Penelitian

Hipotesis mampu didefinisikan sebagai jawaban sementara yang memprediksikan adanya kolerasi dari satu variabel dengan variabel terpilih lainnya melalui satu uji diikuti uji lainnya sehingga didapatkan jawaban yang sesungguhnya atau kebenaran dari permasalahan yang terjadi. Adapun peneliti

telah menetapkan jawaban-jawaban sementara dari penelitian yang akan dijalani, antara lain :

1. Diduga teknik datamining klasifikasi dengan pohon keputusan menggunakan algoritma C4.5 dapat menganalisis penjualan barang konstruksi
2. Diduga hasil analisa penelitian dapat membantu memprediksi penjualan pada PT Batam Bangun Prathama.

2.4. Penelitian Terdahulu

Subbab ini memaparkan emabran mengenai penelitian-penelitian yang telah dilakukan oleh peneliti yang berbeda-beda beserta dengan hasilnya yang bersinggungan atau mencakup hal-hal terkait dengan penganalisan menggunakan algoritma C4.5, antara lain:

1. Sebuah observasi berjudul “Analisa Algoritma C4.5 untuk Prediksi Penjualan Obat Pertanian di Toko Dewi Sri” dari (Kiki Rosita Dewi, 2020) ditemui hasil berikut yakni penerapan jenis algoritma C4.5 telah dapat dibuktikan mampu memprediksi penjualan obat pada Toko Dewi Sri berdasarkan parameter atau patokan tertentu yang telah ditetapkan oleh Toko Dewi Sri.
2. Sebuah observasi berjudul “Komparasi Algoritma C4.5 Dan *Naive Bayes* Yang diKembangkan Menjadi *Web Intellegence* Pada Perhitungan Bonus Tahunan Karyawan di PT.ABC” dari (Ibnu Alfarobi & Taransa Agasya Tutupoly, 2019) ditemui hasil berikut yakni bonus karyawan dari kedua algoritma *Decission Tree C4.5* dan *Naive Bayes*

yang dimana memiliki nilai *Area Under Curve (AUC)* sebesar 0.991 dan *Accuracy* atau *Confusion Matrix* sebesar 99%.

3. Sebuah observasi berjudul “Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Motor Pada PT. Capella Dinamik Nusantara Cabang Muka” dari (Nurul Azwanti, 2018) ditemui hasil berikut yakni penerapan jenis algoritma C4.5 telah dapat dibuktikan mampu mendukung bagian penjualan pada perusahaan dalam menganalisa dan membuat prediksi-prediksi kegiatan penjualan motor PT. Capella Dinamik Nusantara.
4. Sebuah observasi berjudul “Penerapan Algoritma C4.5 Pada Analisis Penentuan Kelayakan Penerima Bonus Tahunan Pegawai Bank Sinarmas Multifinance” dari (Intan Komala Dewi Patwari & Baiq Andriskha Candra Permana, 2019) ditemui hasil berikut yakni penerapan jenis algoritma C4.5 telah dapat dibuktikan mampu mendukung para pihak manajerial atas Bank Sinarmas Multifinance menentukan putusan atas kelayakkan pegawai-pegawainya menerima bonus tahunan.
5. Sebuah observasi berjudul “Prediksi Profit Perusahaan Menggunakan Algoritma C4.5” dari (Elisa, 2018) ditemui hasil berikut yakni penerapan jenis algoritma C4.5 telah dapat dibuktikan mampu merangkai sebuah sistem yang menyokong pengambilan putusan dalam membuat prediksi-prediksi peluang sebuah entitas bisnis berdasarkan kriteria yang telah ditetapkan. Penerapan jenis algoritma C4.5 juga bernilai memiliki efektifitas yang cukup baik.

6. Sebuah observasi berjudul “Prediksi Waktu Kuliah Dengan Algoritma Classifier” dari (Rika Harman & Evan Rosiska, 2020) ditemui hasil berikut yakni penerapan jenis algoritma C4.5 telah dapat dibuktikan mampu memberikan bantuan bagi para mahasiswa mendapat pokok masalah karena adanya keterlambatan penyelesaian masa studi dari yang semestinya.
7. Sebuah observasi berjudul “Penerapan Algoritma C.45 Untuk Kelayakan Penerima Uang Dan Beras” dari (Harman, 2018) ditemui hasil berikut yakni penerapan jenis algoritma C4.5 telah dapat dibuktikan mampu membantu pengambilan proses evaluasi atas kebijakan penerimaan bantuan uang dan beras dari ketidak-adilan dan nepotisme.
8. Sebuah observasi berjudul “*Perfomance evaluation of the GIS-based datamining techniques of best-first decision tree, random forest, and naïve bayes tree for landslide susceptibility modeling*” dari (Chen, Zhang, Li, & Shahabi, 2018) memberikan hasil yakni *random forest model* memiliki keakuratan yang sangat tinggi dalam hal memprediksi sebuah fenomena.
9. Sebuah observasi berjudul “*Reservoir Inflow Forecasting Using ID3 and C4.5 Decision Tree Model*” dari (Charoenporn, 2017) ditemui hasil berikut yakni penerapan jenis algoritma C4.5 dan ID3 telah dapat dibuktikan mampu membantu efisiensi kinerja gudang dan memangkas birokrasi atau alur kerja. Dalam penelitian ini merujuk pada bentuk

perhitungan *decision tree*.