

BAB II

TINJAUAN PUSTAKA

2.1. Teori Umum

Bab ini mengenai istilah dan pengertian secara umum yang berhubungan dengan pelaksana penelitian.

2.1.1 *Knowledge Discovery In Database*

KDD adalah cara yang menemukan informasi yg diperoleh dari database. Tabel yang saling berhubungan dalam database. Informasi yang dihasilkan pada proses ini dapat digunakan sebagai basis pengetahuan untuk tujuan pengambilan keputusan. Istilah penemuan pengetahuan (KDD) dan penambangan data dalam basis data sering Dipertukarkan digunakan untuk menggambarkan proses penggalian informasi yang tersembunyi dalam database besar. Padahal, konsep kedua istilah ini berbeda tetapi saling berhubungan. Datamining adalah fase dari keseluruhan proses KDD. (Mardi 2017) Proses KDD secara garis besar dapat dijelaskan sebagai berikut:

1. Penyeleksian

Sebelum fase pembekalan dimulai pada KDD harus melakukan seleksi data dari operasional dataset. Biarkan data spesifik yang digunakan dalam proses penambangan data dalam file yang berbeda dari database operasional.

2. Pembersihan

Sebelum proses penambangan data, perlu untuk membersihkan data yang berfokus pada penemuan pengetahuan (KDD) dalam database. Proses pembersihan meliputi penghapusan data duplikat, pengecekan data yang tidak konsisten, dan koreksi kesalahan pada data, seperti kesalahan pencetakan.

Pengayaan juga dilakukan, yaitu proses “memperkaya” data yang ada dengan data atau informasi lain yang relevan yang diperlukan untuk Knowledge Discovery (KDD) dalam database, seperti data atau informasi eksternal lainnya yang diminta.

3. Perubahan

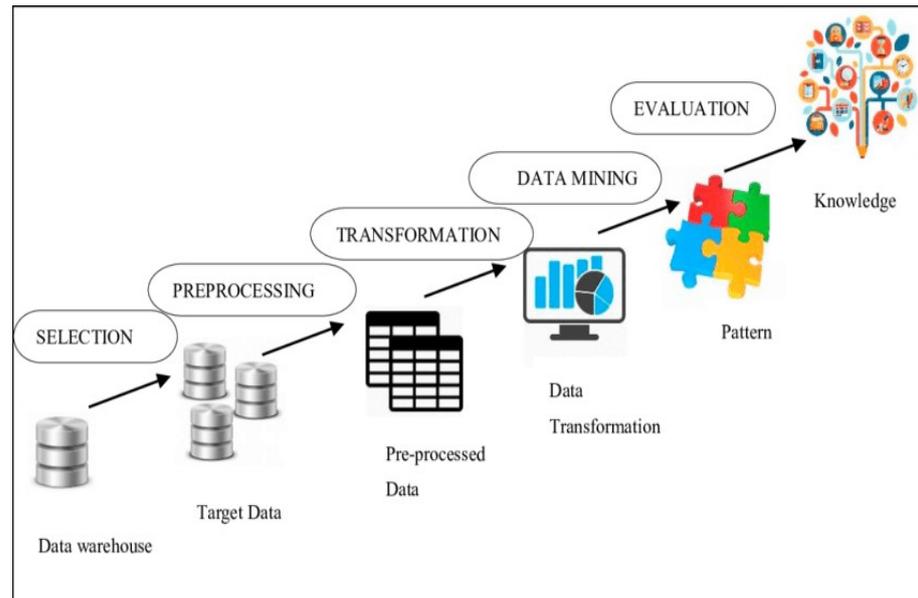
Transformasikan yang dipilih untuk membuat Data yang sesuai untuk operasi penambahan data. Pengkodean dalam Basis Data Penemuan Pengetahuan (KDD) adalah proses kreatif yang sangat bergantung pada jenis atau gaya informasi yang akan dicari dalam basis data.

4. *Data Mining*

Data mining adalah proses menggunakan teknik atau metode tertentu untuk menemukan pola atau informasi menarik dalam data yang dipilih. Teknik, metode atau algoritma dalam data mining sangat beragam. Pilihan metode atau algoritma yang benar sebenarnya tergantung pada tujuan dan proses penemuan pengetahuan yang komprehensif (KDD) dalam database.

5. Evaluasi

Pola data yang di hasilkan dari langkah informasi *data mining* di tampilkan ke wujud yang mudah di pahami pengguna .(Alwafi Ridho Subarkah 2018)



Gambar 2. 1 Langkah langkah KDD

2.1.2 Data Mining

Datamining ialah mempunyai peranan sebagai prediksi, deskripsi, klasifikasi serta asosiasi. Tidak hanya itu peranannya yang merupakan analisis asosiasi klasifikasi serta memprediksi analisis kluster, analisis *outlier*, analisis *trend* serta evolusi. (Novia, Hidayah, and Rozi 2018)

Datamining menganalisis serangkaian tampilan data untuk membuat tautan tak terduga dan merangkum data dengan cara yang berbeda dari sebelumnya. Ini mudah dan nyaman bagi pemilik data. (Utomo and Mesran 2020)

2.1.3 Kelompok *Data Mining*

Berikut ini Berdasarkan tugasnya :

1. Estimasi

Estimasi dan klasifikasi itu sama, kecuali bahwa variabel objektif penilaian merupakan tipe numerik. Model dikonfigurasi dengan harapan nilai yang diberikan sesuai dengan variabel target.

2. Memprediksi

Peramalan, penilaian dan estimasi adalah, kecuali untuk memprediksi hasil nilai di masa depan.

3. PengKlasifikasian

Dalam taksonomi, ada variabel kategoris target. Misalnya, klasifikasi pendapatan yaitu hasil penjualan tinggi, hasil penjualan menengah, dan hasil penjualan rendah.

4. Kluster

Cluster adalah kumpulan *record*, mengamati atau mengamati, dan membuat ruang objek sejenis. Blok adalah rangkaian record yang memiliki kesamaan dan memiliki perbedaan dengan record pada kelompok lainnya. Rakitan tidak sesuai dengan deskripsi, yaitu tidak ada variabel target dalam rakitan.

5. Mengasosiasi

Tugasnya untuk mendapatkan pola yang timbul pada suatu saat. pada global usaha sering dianggap analisis keranjang belanja.(Mardi 2017)



Gambar 2. 2 Pengelompokan Data Mining

2.1.4 Algoritma C4.5

Algoritma C4.5 artinya bentuk algoritma yg dipergunakan sebagai pembagian terstruktur yang serta bersifat prediktif yang pembentukannya adalah pohon keputusan (*Decision Tree*). Berikut langkah – langkahnya :

1. Siapkan data pelatihan yang didapatkan pada historis yang sebelumnya sudah terjadi dan dikompilasi ke dalam bab-bab eksklusif.
2. Tentukan akar pohon yang berasal dari Rute atribut yang didapat dari menghitung nilai gain untuk tiap atribut, dengan nilai gain maksimum menjadi rute pertama. Sebelum menghitung nilai gain dari atribut, terlebih dahulu menghitung nilai entropi. Berikut rumus yang dipakai:
 - a. Rumus menghitung nilai entropy

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

Rumus 2. 1 Rumus Entropy

Keterangan :

S = himpunan

A = kelengkapan yang dipakai

n = jumlah partisi pada bagian S

p_i = proporsi dari S_i ke S

b. Kemudian hitung nilai gain menggunakan persamaan

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

Rumus 2. 2 Perhitungan Gain

Keterangan :

S = himpunan

A = keterangan

n = jumlah partisi di keterangan atau kelengkapan bagian A

$|S_i|$ = jumlah yang ada pada partisi i

$|S|$ = jumlah yang ada pada S

c. Kembali lagi ke cara 2 sampai semua *record* terpartisi.

d. Jika semua record memiliki kelas N yang sama, atribut dalam record tidak dipartisi ulang, dan record di cabang tidak kosong, proses partisi pohon keputusan berakhir..(Mardi 2017)

2.1.5 WEKA (Waikato Environment for Knowledge Analysis)

Menurut pendapat WEKA (Lingkungan Waikato untuk Analisis Pengetahuan) sebagai bagian dari arti pembelajaran mesin yang belum pernah dirilis, WEKA didistribusikan di Universitas Waikato di Selandia Baru yang digunakan untuk pendidikan dan penelitian. WEKA adalah perangkat lunak yang bersifat open source, sehingga siapa pun dapat mengakses perangkat lunak ini.

Pengujian dengan WEKA akan jauh lebih mudah dan lebih mudah untuk membuat pohon keputusan, hampir seperti perhitungan yang dilakukan secara manual. (Elisa & Ervina, 2018) .

2.2. Teori Khusus

Teori ini membahas istilah dan pengertian secara khusus berkaitan dengan pelaksanaan penelitian.

2.2.1 Karyawan

Menurut Ayat 2 Pasal 1 Undang-Undang Nomor 13 Tahun 2003 tentang Ketenagakerjaan, pekerja adalah orang yang dapat menghasilkan barang dan jasa yang dapat memenuhi kebutuhannya dan kebutuhan masyarakat, baik di dalam maupun di luar hubungan kerja. Ini menyatakan Dari pengertian di atas, jelas bahwa penggunaan tenaga kerja adalah tenaga kerja yang mempekerjakan dalam bentuk apapun perusahaan atau individu dengan menggunakan upah. (Novia et al. 2018)

2.2.2 Kinerja Karyawan

Memahami kinerja karyawan Keberhasilan kegiatan organisasi sangat tergantung pada peran karyawan di tempat kerja. Semakin baik Sebaliknya semakin baik kinerja karyawan maka semakin baik pula kinerja perusahaan, tetapi semakin buruk kinerja karyawan maka semakin jauh tujuan yang ditetapkan perusahaan dari tujuan yang dipersyaratkan. Kinerja organisasi berarti bahwa kegagalan merupakan bagian yang tidak terpisahkan dari pelaksanaan tugas organisasi baik di forum pemerintah maupun swasta. Achievements berasal dari bahasa achievement atau pencapaian aktual (actual achievement atau achievement by individu atau organisasi). Kinerja seorang pegawai (job performance) adalah

hasil kerja secara kuantitatif dan kualitatif yang dilakukan oleh pegawai dalam melaksanakan pekerjaan sesuai dengan tanggung jawab yang diberikan kepada pegawai tersebut. (Onainor 2019)

Faktor Kinerja Karyawan adalah kegiatan yang paling umum dalam suatu organisasi, untuk melakukan sesuatu yang berhubungan dengan pekerjaan, jabatan, atau peran dalam organisasi. Faktor kinerja terdiri dari faktor internal dan eksternal dan faktor internal (kecenderungan), yaitu faktor yang berhubungan dengan karakteristik seseorang. Sedangkan faktor eksternal adalah faktor yang mempengaruhi kinerja seseorang dari lingkungan. Sikap, sikap dan perilaku rekan kerja, bawahan atau manajer, fasilitas kerja dan budaya organisasi, dll. Berikut faktor yang mempengaruhi hasil kinerja karyawan, yaitu :

1. Kualitas kerja adalah tingkat baik buruknya pekerjaan yang diterima seorang pegawai yang dapat ditinjau dari segi ketelitian, keanggunan, keterampilan, dan kemampuan.
2. Kerja Tim (Teamwork), Cari tahu bagaimana seorang karyawan bekerja dengan menggunakan orang lain untuk menyelesaikan suatu pekerjaan.
3. Kreativitas adalah kemampuan pegawai untuk menyelesaikan pekerjaannya dengan caranya sendiri atau atas prakarsa sendiri, yang disebut kemampuan efektif, efisien dan mampu menciptakan perubahan baru bagi pemulihan dan kemajuan organisasi.
4. Inovasi adalah tentang mewujudkan sesuatu yang baru atau memperbarui yang sudah ada. Inovasi dapat diartikan sebagai suatu bentuk kreativitas yang tidak hanya dapat menghasilkan ide-ide baru dan bermanfaat, tetapi

juga mengimplementasikan, mengkomersilkan dan memberikan dampak yang besar.

5. Mengambil inisiatif adalah melakukan sesuatu atau melakukan sesuatu tanpa harus memahami apa yang harus dilakukan sebelumnya..(Novia et al. 2018)

2.2.3. *Trainer*

Pelatih Hal ini penting untuk pengembangan profesional yang sukses. Penguasaannya terhadap materi latihan tidak cukup menjadi andalan jika tidak dibarengi dengan penggunaan beberapa keterampilan lainnya. Keterampilan lain ini berkaitan dengan penggunaan pemahaman tentang metode dan materi dalam pelatihan dengan menggunakan kegiatan profesional pelatih. Selain itu, memahami pengembangan karir itu sendiri merupakan kompetensi penting dari seorang pelatih yang sukses. Pembina atau coach berperan penting dalam pengembangan dan perubahan organisasi, pencapaian tujuan, kompetensi, pengetahuan, dan kapabilitas yang dibutuhkan peserta. Seorang pelatih memiliki dampak pada pelatihan peserta. Pelatih adalah seseorang yang mengajarkan keterampilan orang untuk mempersiapkan pekerjaan. Pelatih adalah seorang pelatih atau sering disebut menggunakan pelatih yang menerima dan menguji pengetahuan dan kemudian menghasilkan hasil tes dengan menggunakan beberapa elemen pelengkap (yang dapat dikurangi atau ditambahkan) sehingga merupakan metode pembelajaran yang efektif dan tepat, sehingga peserta pelatihan mampu menyerap pengetahuan ini dengan cara yang dapat diterapkan dan dapat melakukan hal itu dengan menggunakan apa yang disajikan atau diarahkan oleh pelatih. (Heriyanto et al. 2020)

2.3. Penelitian Terdahulu

Peneliti juga mencapai beberapa hasil penelitian sebelumnya terkait analisis menggunakan algoritma C4.5, yang dirangkum sebagai berikut:

1. pengkajian penggunaan Data Mining dengan c4.5 untuk mendapatkan hasil kinerja karyawan terbaik yang dilakukan oleh (Novia et al. 2018) Penelitian dilakukan pembagian terstruktur mengenai atau segmentasi bersifat prediktif yang digunakan untuk pohon keputusan untuk membantu mempermudah pihak pusat penjaminan mutu (PPM) dengan memiliki taraf akurasi yang termasuk penjabarannya akurat.
2. Dari Penelitian menentukan kinerja karyawan dalam berkerja dengan perusahaan memakai algoritma C4.5 (Ferdiansyah and Goeirmento 2020) Survey karyawan serta hasil yang didapatkan akurat menjadi pengambil Sebuah keputusan yang memprediksi loyalitas karyawan kepada perusahaan.
3. Dari Survey Dampak Kompensasi Terhadap Kinerja Karyawan di Kantor Cabang Persada dan PT Arminareka Perdana Surabaya di Indonesia (Wati 2019) Menggunakan hipotesis asosiatif dan komperatif membuktikan nilai yang besar dan signifikan .
4. Dari Penelitian Algoritma Klasifikasi Kandidat C4.5 untuk Kompetisi Cerdas di SMP dengan Aplikasi weka (Ardiansyah 2019) Klasifikasi data kandidat potensial sebagai sarana penerapan algoritma C4.5 terbukti memiliki akurasi yang diperoleh dari evaluasi klasifikasi tersebut.

5. Dari Penelitian Analisa Kinerja Algoritma C4.5 untuk menentukan target keuntungan (2017, Nofitri, and Sembiring 2017) Mendapatkan informasi tambahan bagi pengelola perusahaan di saat menggunakan rule menjadi alat dalam memprediksi pencapaian profit
6. Dari Penelitian Menerapkan penambahan data untuk memprediksi hasil penjualan menggunakan algoritma C4.5 (Eska 2018) Dalam memakai data mining algoritma C4.5 menggunakan cara dagang yang meliputi pemilihan atribut menjadi akar dan cabang-cabang dan setiap cabang akan pada ulangi sampai Beberapa aturan dapat diperoleh dari penyelesaian pohon keputusan dengan kelas yang sama.
7. Sebuah studi dari tentang penggunaan data mining dengan algoritma C4.5 pada klasifikasi fasilitas medis di negara bagian Indonesia. (Algoritma et al. 2020) Dengan analisis data mining, hasilnya menjadi model pohon keputusan, menentukan peringkat Indeks Pembangunan Manusia, dan menggunakan KFold Cross Validation untuk mendapatkan skor akurasi.
8. Dari Penelitian penggunaan Data Mining pada menentukan mobil yang akan dipilih Dengan memakai Algoritma C4.5 (Harahap 2018) Pada hal ini penjabaran pembelian mobil berdasarkan penerapan algoritma C4.5 membantu memberi bahan pertimbangan pembeli buat memilih mobil yang akan di belinya.
9. Dari Penelitian Penerapan algoritma data mining C4.5 untuk mendapatkan catatan kinerja instruktur STT Harapan Medan, Pemilihan catatan kinerja instruktur menurut tiga atribut input (publikasi,

komitmen, pendidikan) dan satu atribut output (nilai kinerja (rendah, cukup, tinggi)).

10. Dari Survey penerapan algoritma data mining C4.5 untuk mengetahui kepuasan pelanggan di Grand Zuri Hotel Dumai (Desyanti 2019) Setelah di lakukan penelitian menggunakan algoritma C4.5 diperoleh yang akan terjadi Jika pelayanannya sangat puas.
11. Dari Penelitian Penerapan Algoritma C4.5 Dalam Memprediksi Ketersediaan Uang Pada Mesin ATM (Syahputra, Hartono, and Rosnelly 2021) Model pohon keputusan yang di hasilkan merupakan menjadikan variabel saldo sebagai akar. Lalu melakukan uji coba dengan metode *K-Fold Cross Validation* hasilnya tingkat akurasi lebih tinggi.
12. Dari Penelitian Menentukan faktor kesesuaian penerimaan karyawan menggunakan algoritma C4.5 (pohon keputusan) pada perusahaan outsourcing staff PT (Sahputra et al. 2020) Dengan memakai metode algoritma C4.lima atribut yang di pakai dalam peneltian bedasarkan Usia, tempat tinggal, jenis kelamin, jabatan, bidang studi, referensi, tingkat pendidikan, dan tes. Saat menggunakan metode solusi, tingkat akurasinya sangat tinggi.
13. Dari penelitian *Predict lung cancer survival based on performance using support vector machine C4.5 algorithm classification techniques for health care analysis* (Pradeep and Naveen 2018) Deretan data penyakit kanker paru-paru beserta dengan data pasien barudi pakai buat mengevaluasi kinerja, perbandingan pada dasarnya akurasi dengan menggunakan algoritma C4.5.

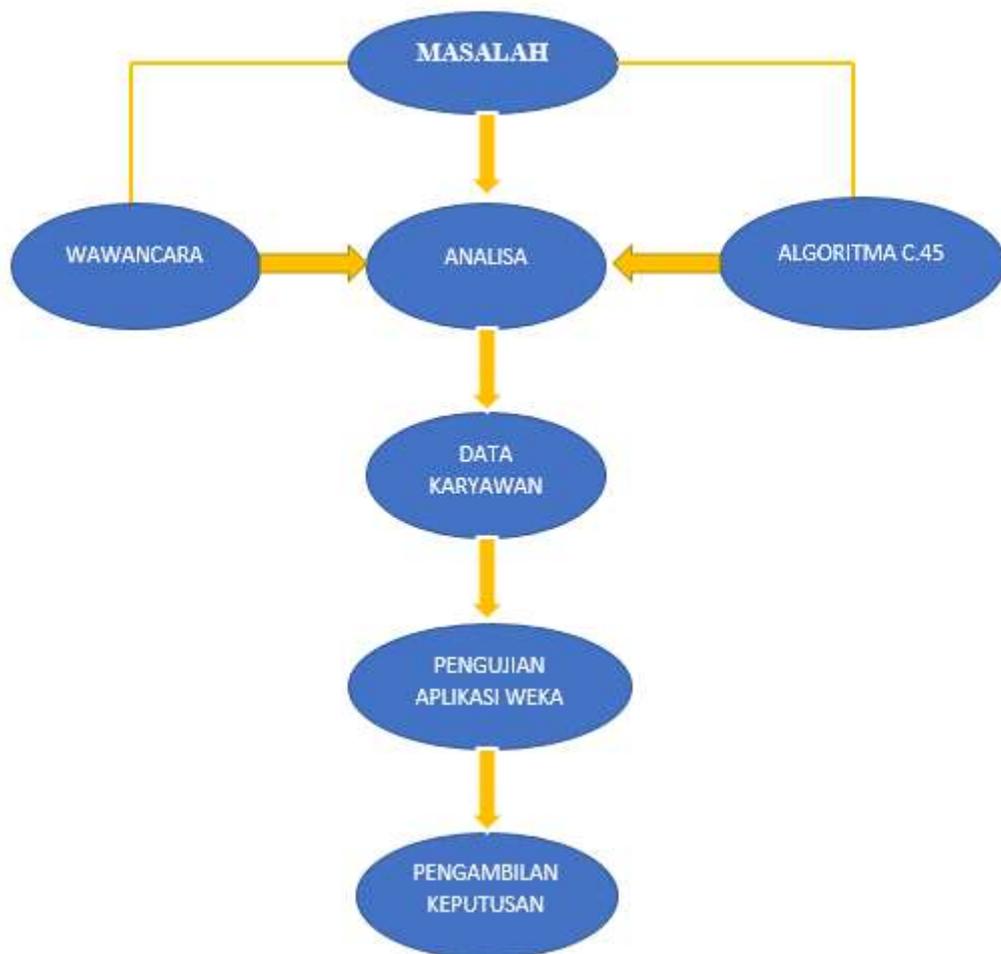
14. Dari Penelitian *Digital transformation in managing agile and employee engagement C4.5 Algorithm* (Elacio, Balazon, and Lacatan 2020)

Penggunaan algoritma C4.5 serta deretan data yang pada kalibrasi membarui pendekatan manual berasal retensi karyawan menjadi pendekatan digital memberikan akurasi yang sangat tinggi.

15. Dari Penelitian *Predicting employee attendance factors using the C4.5 algorithm, random tree, and random forest* (Fahlapi et al. 2020) Dari hasil analisa penulis faktor keterlambatan karyawan menggunakan 3 prosedur pemecahan yaitu bernilai akurasi.

2.4. Kerangka Pemikiran

Dibawah ini adalah kerangka pemikiran dari pengkajian dan juga penjelasannya



Gambar 2. 3 Kerangka Berfikir