

BAB II

TINJAUAN PUSTAKA

2.1 Teori

Adalah sebuah penjelasan memaparkan teori ataupun pendapat pendapat akan dikajikan peneliti didalam penelitian ini, ada beberapa teori-teori yang akan dijelaskan disini terdiri dari 2 cakupan teori yang besar yaitu teori secara umum (masih secara menyeluruh) dan teori secara khusus (Lebih spesifik), penjelasannya sebagai berikut:

2.1.1 Teori Umum

Ialah penjabaran tentang pendapat umum(sering kita dengar) yang sering digunakan dan yang digunakan untuk pengajuan pada proses penelitian, teori umum disini digunakan untuk mempertegas isi didalam dari teori (pendapat) yang ada yang digunakan peneliti. Ini dilakukan bertujuan agar hasil dari pengkajian yang laksanakan memiliki kualitas lebih tinggi. Beberapa isi yang ada dalam teori umum yang ada antara lain:

2.1.1.1 Karyawan

Yaitu salah satu faktor dalam perusahaan dan sumber daya penting yang dapat dilakukan sebagai alat untuk menggerakkan dan memajukan status perusahaan. Karyawan diperusahaan adalah faktor penting sebagai penggerak utama seluruh rangkaian kegiatan aktivitas yang dilakukan perusahaan untuk sampai pada tujuan yang diinginkan, Untuk hanya sekedar mendapatkan keuntungan ataupun hanya menjaga keberlangsungan siklus perusahaan itu

sendiri. Sukses ataupun tidak perusahaan untuk tetap dalam menjaga keadaan perusahaan dimulai dengan karyawan itu sendiri dalam mempertahankan kelangsungan bisnis yang dijalankan perusahaan dan upaya untuk meningkatkan efektivitas dan efisiensi waktu dan keuangan semaksimal mungkin (Fitria Rachmawati et al., 2018).

2.1.1.2 Sumber Daya Manusia (SDM)

Adalah bagian yang menjadi faktor yang teramat penting untuk menaikkan jumlah produksi dalam kinerja suatu kelompok baik organisasi dan instansi yang akan dijalankan. Dibidang personalia dalam perusahaan terdapat beberapa permasalahan yang bisa ditimbulkan melalui dari hasil seleksi pengangkatan dan penerimaan karyawan yang semula kontrak menjadi karyawan tetap dan kebutuhan akan karyawan baru pada tempat usaha tidak bisa dipastikan jumlah dan kualitasnya, walau dilakukan dan *diplanning* kan dengan sangat baik. Ini dikarenakan berbagai faktor yang muncul melalui faktor internal dan eksternal perusahaan. Maka untuk itu sebelum kita melakukan tahapan penerimaan untuk melakukan ujian karyawan baru pihak HRD harus menetapkan dulu kebutuhan apa saja yang diperlukan baik dari faktor kualitas atau kuantitas calon karyawan (Fitria Rachmawati et al., 2018).

2.1.2 Teori Khusus

ini adalah gagasan yang digunakan dan disampaikan dalam berikut ini :

2.1.2.1 Know Discovery In Database (KDD)

Dibentuk melalui cakupan pengumpulan data, penggunaan kilas balik yang akan menghasilkan aktifitas sesuai kebutuhan, hubungan/pola pada suatu data yang berukuran cukup besar. Metode ini dibuat untuk mendapatkan ilmu yang asalnya dari *database* yang dimiliki. Ada tabel yang bersangkutan dalam *database*. Hasil yang di dapat dari pengetahuan ini yaitu berupa tahapan ini dapat dijadikan untuk dasar ilmu untuk bertujuan dalam pengambilan keputusan yang akan datang(Erlin Elisa & Yunita Tri Utami et al., 2022).

KDD ialah salah satu tahapan yang tidak boleh dianggap enteng pada tahap identifikasi pola data yang mana pola yang mana pola tersebut bersifat sah dan baru juga berguna. KDD adalah proses yang digunakan untuk mencari informasi yang dapat digunakan di *data base*. Seluruh rangkaian proses dalam tahap ini terdiri beberapa langkah, yaitu tentang memahami bidang dalam aplikasi,lalu membuat target kebutuhan data yang dibutuhkan dan telah ditentukan darimana data mentah yang digunakan tersimpan didalam *database*, dilakukan proses *cleaning* data dan pra-pemrosesan data (Fiandra et al., 2017) .

tahapan-tahapan dalam *Knowledge Discovery in Database* (KDD) menurut (Elisa & Ervina et al., 2018) antara lain:

1. *Cleaning Data*

Proses pembersihan (pengurangan) data dari bentuk yang tidak konsisten juga melakukan (*noise*) yang tidak dibutuhkan.

2. *Data Integration*

Tahapan dimana data akan digabungkan melalui banyak jenis asal sumber data yang tidak sama..

3. *Data Selection*

Seluruh data yang dimiliki akan di pilih melalui database menggunakan cara pemilihan dan ataupun memilah jumlah data yang yang dibutuhkan dan disesuaikan dengan tujuan dari analisa.

4. *Data Transformation*

Tahap untuk merubah data dari data mentah jadi data yang sesuai keinginan dan dapat di proses dalam tahapan selanjutnya, yaitu dengan *data mining*

5. *Data mining*

Metode diimplementasikan dan disesuaikan dengan teknik khusus untuk mendapatkan suatu gambaran pola data dari jumlah data yang di olah.

6. *Patten Evaluation*

Merupakan identifikasi pola ataupun bentuk dari sebuah data.

7. *Knowledge Presentation*

Merupakan hasil olah dari *data mining* yang berbentuk dalam informasi yang dihasilkan dan digambarkan lalu bisa sajikan kepada tempat objek penelitian dilakukan.

2.1.2.2 WEKA (Waikato Environment for Knowledge Analysis)

WEKA adalah bagian dari *tool* yang digunakan sebagai pengetahuan unstuck pada mesin, alat ini dijual pada Universitas Waikoto yang berposisi

didaerah bernama Selandia Baru yang digunakan untuk memudahkan pendidikan mesin dan penelitian tentang mesin . *Waikato Environment for Knowledge Analysis* ialah perangkat lunak bersifat bebas akses, maka perangkat aplikasi ini bisa digunakan siapa saja. Pengujian yang dilakukan dengan aplikasi *WEKA* lebih sederhana dan juga lebih mudah pengaplikasiannya untuk mendapatkan bentuk sebuah pohon keputusan, Sedikit mirip sederhananya seperti perhitungan yang dilakukan secara manual dengan *ms.excel* (Erlin Elisa & Yunita Tri Utami et al., 2022).

2.1.2.3 Data Mining

Ilmu ini mulai diperkenal tahun 1990, saat pekerjaan dengan memanfaatkan data yang ada menjadi sesuatu yang berguna dalam banyak bidang keilmuan, dimulai dengan bidang pendidikan, bidang bisnis sampai pengobatan. Adanya ilmu *data mining* diawali dengan jumlah data yang disimpan dalam basis data yang semakin membesar. Sehingga ilmu *Data mining* berperan menjadi prediksi, pendeskripsian, serta pembagian yang memiliki struktur tentang asosiasi. Tidak sampai disitu peranan *data mining* sendiri ialah menganalisis asosiasi dalam pembagian yang terukur mengenai prediksi tentang *cluster* analisis, *outlier* analisis, analisis kejadian terkini (Erlin Elisa & Yunita Tri Utami et al., 2022).

Ilmu ini merupakan beroperasi semi otometric yang mana menggunakannya dengan teknik statistik, matematika, *artificial intelligence*, serta *machine learning* untuk menghasilkan dan pengidentifikasian informasi tentang ilmu potensial yang berguna dan ter *save* didalam *database* yang besar. Pengidentifikasian dalam *data mining* adalah satu dari sekian kesulitan dalam

ilmu *datamining* dan merupakan adalah hal nyata dalam *data mining* membawa berbagai pendukung dan metode dari berbagai bentuk keilmuan yang sudah mapan dan sudah ada lebih dahulu (Sumpena et al., 2018).

Berasal dari beberapa acuan ilmu yang ada, ilmu *data mining* bermaksud membenahi metode tradisional agar bisa menyelesaikan:

1. Data dalam jumlah yang besar
2. Data memiliki berdimensi
3. Kesamaan atau perbedaan sifat antar data yang dimiliki

Karakteristik dalam *Data mining* berikut ini (Elisa & Ervina, 2018) :

1. *Data mining* berkaitan pada suatu temuan yang masih tersembunyi dalam data terdapat pola bentuk data yang tidak kita ketahui sebelum ini.
2. *Data mining* mempunyai jumlah data yang begitu amat besar, dan kebiasaannya data yang begitu amat besar dapat menghasilkan kajian dapat lebih meyakinkan..
3. Bisa digunakan dalam pembuatan sebuah keputusan bersifat kritis, terutama yang berhubungan dengan strategi (rancangan).

2.1.2.4 Pengelompokan Data Mining

Kegunaan *data mining* memiliki tugas yang terbagi jadi enam antara lain (Suyanto, 2017) :

1. ((*Classification*) ialah mengklasifikasikan model/bentuk yang bisa dilihat untuk diterapkan dalam data yang masih baru.
2. (*Clustering*) merupakan pengelompokkan data, yang mana kita belum mengetahui label kelas data, kedalam kelompok data pasti yang sesuai dengan sifat aturan kemiripan data yang dimilikinya.
3. (*Regression*) yaitu yang berfungsi menggambarkan data dengan prediksi kesalahan sekecil mungkin.
4. (*Anomaly Detection*) yaitu mengidentifikasi data yang ada dan tidak lazim ataupun umum, dapat berbentuk *outlier* (pencilan), mengubah data atau deviasi data yang bisa terjadi begitu penting untuk dilakukan dan perlu investigasi atau pengecekan lebih lanjut yang akan dilakukan.
5. (*Association rule learning*) merupakan menggambarkan ketergantungan (*dependency modeling*) untuk mendapatkan hubungan antara variable yang diolah.
6. (*Summarization*) ialah menggambarkan dan menyimpulkan data menjadi yang lebih simpel, meliputi (tampilan data) dan pembuatan hasil laporan dari pengolahan data..

2.1.2.5 Klasifikasi

Merupakan Bagian dari pengolahan datang yang sangat penting di teknik *data mining* yaitu tentang cara mempelajari jumlah data-data yang banyak bisa menghasilkan aturan tentang mengklasifikasikan dan mengenali data baru yang belum disentuh dan belum pernah dipelajari sebelumnya. Teknik ini dijelaskan

sebagai proses yang dilakukan untuk menegaskan satu objek data menjadi bagian salah satu kategori dalam data (*class*) yang terdefinisi (Suyanto, 2017).

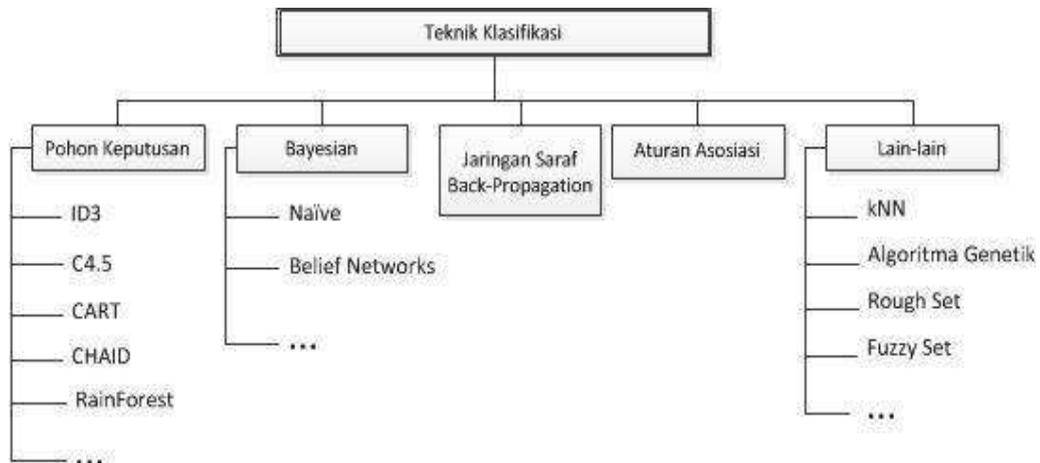
Lalu teknik klasifikasi adalah langkah yang dilakukan agar menggali hasil kekayaan data yang mirip dengan gabungan beberapa objek di *data base* dan dimasukkan pengkelompokkan dalam kelas yang tidak sama. Yang dikelompokkan melalui klasifikasi yang ditetapkan sebelumnya. tahap ini memiliki 2 tahap yaitu yang pertama proses mempelajari data dianalisis dengan metode algoritma klasifikasi sampai bisa menghasilkan bentuk aturan dalam klasifikasi. Lalu tahap selanjutnya adalah tahapan uji coba data untuk menghitung ketepatan data dari aturan klasifikasi (Elisa & Ervina, 2018).

Beberapa bentuk kasifikasi yang lumrah digunakan oleh banyak penelitian (Elisa & Ervina, 2018):

1. Membentuk Deskriptif yang dibutuhkan sebagai bantuan untuk alat menjabarkan dari kelas yang berbeda untuk dapat membedakan suatu objek yang ada.
2. Pembentukan prediksi (Prediktif) biasa dapat digunakan demi memperhitungkan judul untuk kelas agar merecode sesuatu yang belum dan tidak dikenali sebelumnya.

ketika dilakukan proses dalam pengklasifikasian data, ada beberapa metode yang Wajib perlu kita ketahui sebelumnya antara lain adalah teknik Bayesian, teknik jaringan saraf, pola asosiasi, *back propagation*, terakhir *Decision*

tree. Untuk memperjelas adapun teknik klasifikasi sudah digambarkan dan dapat digambarkan berikut:



Gambar 2.1 Pengelompokan Teknik Klasifikasi

Sumber: (Jamhur, 2016)

2.1.2.6 Pohon Keputusan (*Decision Tree*)

Merupakan suatu teknik dalam klasifikasi yang sangat populer dan cara penggunaannya yang sederhana. Teknik ini digunakan untuk mencari bentuk klasifikasi yang kebal terhadap derau. Dikatakan satu teknik *Decision Tree* yang terpopuler saat ini adalah ID3 (*Iterative Dychotomizer Version 3*). Metode ini ialah salah satu bentuk pengaplikasian bentuk dengan pengaplikasian berbagai tahap serta proses yang ada didalam aktivitas pengklasifikasian dan juga aktivitas dalam prediksi(Suyanto, 2017).

Bentuk *Decision Tree* sengaja dibuat menyerupai dengan bentuk yang sama pada sebatang pohon dan ini penjabarannya:

1. Simpul bagian akar

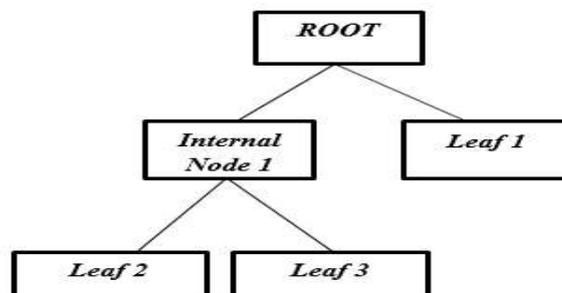
Adalah Bagian yang terletak pada bagian yang paling tinggi dan merupakan struktur awal yang ada dalam pohon keputusan.

2. Simpul internal

Merupakan cabang lanjutan setelah simpul dibagian akar pohon. Tergambar dibawah *Node* bagian akar. Dalam *Node* ini jika memiliki 1 masukan data maka akan menghasilkan *output* data dalam jumlah terbanyak adalah 2 keluaran.

3. Simpul daun

Merupakan *node* yang tempatnya paling diujung dibagian pohon. Pada simpul ini ada 1 masukan tetapi tidak memiliki data *output*.



Gambar 2. 2 Konsep dasar pohon keputusan

Bentuk *Decision Tree* terdiri 3 (tiga) jenis Cabang yang dijelaskan dibawah ini (Elisa & Ervina, 2018) :

1. *Root Node* yaitu simpul faktor utama, simpul ini tidak punya *inputan* tapi biasa juga tidak punya *output* ataupun sebaliknya punya beberapa data *output*
2. *Internal Node* yaitu simpul percabangan, simpul berikut hanya punya 1 masukan dan juga punya beberapa *output*
3. *Terminal Node* atau *Leaf Node* merupakan simpul simpul ujung (akhir), simpul hanya punya *inputan* saja tanpa *output*.

Tugas *Decision Tree* sebagai alat tambahan untuk membantu dan digunakan untuk pengambilan keputusan dalam masa kritis dan yang telah dilakukan pengembangan oleh manusia sejak teori *Decision Tree* makin berkembang. *Decision tree* memiliki kekurangan dan kelebihan yang sudah dijelaskan dalam jurnal (Harryanto & Hansun, 2017) ialah:

1. Kelebihan pohon keputusan
 - a. Tempat pengaplikasian keputusan rumit dapat diganti menjadi lebih mudah..
 - b. Dapat mengurangi jumlah hitungan tidak perlu dikarenakan proses pengujian yang dilakukan hanya beralaskan dari kriteria data yang hanya diperlukan dan tidak menguji data yang tidak memiliki kriteria.
 - c. Proses memilih fitur dari dalam simpul yang tidak sama dan lebih mudah penggunaan serta mudah pengaplikasian. Fitur yang digunakan akan jadi perbedaan pada masing-masing kriteria.

d. Teknik ini bisa mencegah terjadinya masalah baru melalui penggunaan kriteria berjumlah data sedikit pada simpul dalam tanpa harus membatasi hasil dari kualitas keputusan.

2. Kekurangan pohon keputusan

- a. Bisa terjadi *overlap* jika hasil dari keputusan dan penggunaan kriteria memiliki jumlah yang terlalu banyak. Hal ini memicu waktu yang digunakan akan semakin lama serta kalkulasi jumlah memori yang digunakan juga semakin membesar.
- b. Terjadi *error* dalam penghitungan total jumlah dari masing masing tingkat *Decision Tree* besar.
- c. Desain *Decision Tree* yang maksimal akan sulit didapatkan..
- d. Kualitas dari keputusan yang dihasilkan juga terpengaruhi dengan bagaimana kondisi pohon tersebut dibentuk.

2.1.2.7 Algoritma C4.5

Ialah peningkatan bentuk ID3 (*Iterative Dychotomizer Version 3*), peningkatannya dalam ini dapat diatas *missing value*, biasanya yang diatas *continue data* dan *praining data*. Algoritma C4.5 adalah metode pembentukan dalam *Decision Tree* yang didasari oleh kepemilikan data pada waktu sebelum dilakukan penelitian. Metode pokok menjadi dasar algoritma ini ialah pembentukan *Decision Tree* yang didasari oleh pertunjukan atribut yang memiliki total nilai perhitungan *gain* tertinggi berdasarkan dari nilai perhitungan total entropi atribut sebagai tahap awal dari perhitungan algoritma ini ialah

pembentukan *Decision Tree* berawal pemilihan jumlah nilai atribut yang memiliki total nilai *gain* yang paling tinggi berdasarkan nilai perhitungan total nilai entropi atribut tersebut sebagai tahapan atribut klasifikasi. Dalam melakukan pencarian nilai *gain* paling tinggi dapat dilakukan dengan rumus ini (Elisa & Ervina, 2018) .

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

Rumus 2. 1 Perhitungan Gain

Dimana:

S : Himpunan kasus yang ada

A : Atribut

N : Jumlah partisi atribut A

|Si| : Jumlah total kasus pada partisi ke-i

|S| : Jumlah kasus yang ada dalam S

Lalu menghitung jumlah nilai *entropy* bisa dilihat pada rumus.

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

Rumus 2. 2 Perhitungan Entropy

Dimana:

S : Himpunan kasus yang ada

A : Fitur

N : Jumlah partisi S

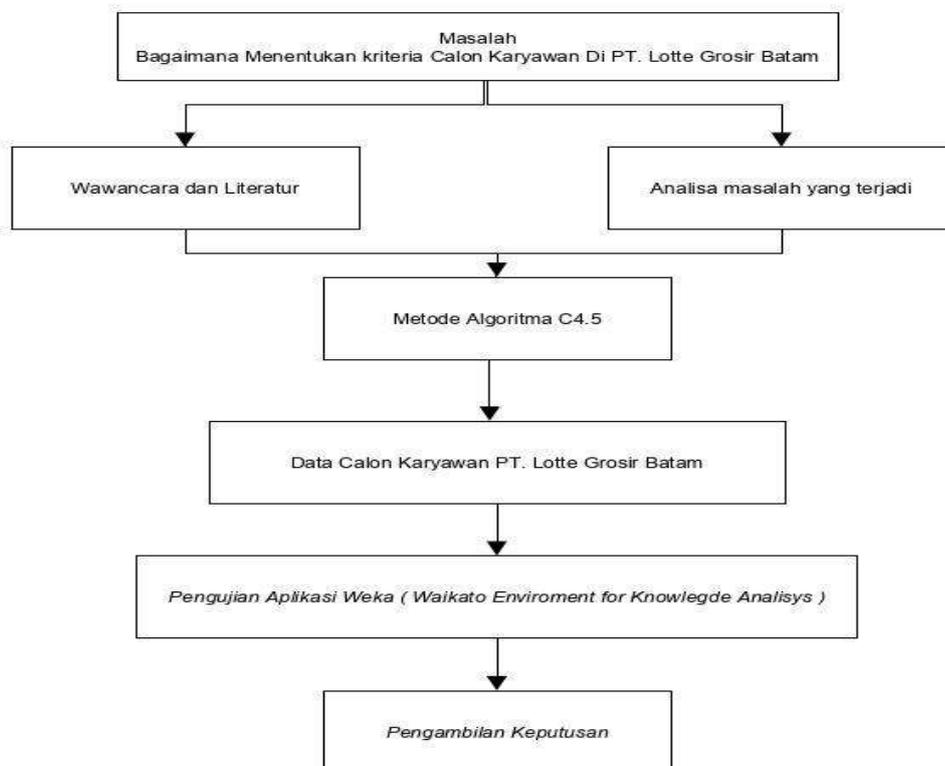
Pi : Proporsi dari Si terhadap S

Dari Hasil jumlah penghitungan nilai *gain* dan nilai entropi ini dapat

diujikan menggunakan *aplikasi WEKA*. Dengan istilah, entropi merupakan pembeda ataupun persamaan. Dalam *data mining*, entropi dijelaskan menjadi sebuah ukuran yang digunakan untuk menghitung kesamaan (*heterogen*) data dalam sebuah kumpulan data data. Semakin mirip jenisnya dalam sebuah kumpulan data maka nilai *entropy* yang hitung semakin besar nilainya (Suyanto, 2017) .

2.2 Kerangka Pemikiran

Merupakan suatu bentuk yang bisa menjelaskan tentang jalan pemikiran yang digunakan pada penelitian tersebut. Berikut adalah gambar Kerangka/pola pikir pada penelitian kali ini :



Gambar 2. 3 Kerangka pemikiran

2.3 Hipotesis Penelitian

Adalah pendapat sementara yang digunakan pada masalah yang dalam proses penyelesaian dan hipotesis ini memiliki sifat praduga/dugaan sementara karena masih harus dibuktikan kebenarannya dengan hasil kesimpulan penelitian ini. Berikut Hipotesis pada penelitian ini:

1. Diduga dengan penggunaan metode *Data Mining* Algoritma C4.5 yang nanti dipakai untuk menganalisis data yang bersumber dari hasil jumlah data calon karyawan.
2. Diduga hasil yang mendasari analisis data penggunaan Algoritma C4.5 diperlukan sebagai acuan pedoman aturan pada menemukan hasil dari kajian masalah ini yaitu dapat ditentukan kriteria karyawan baru yang dibutuhkan perusahaan.

2.4 Penelitian Terdahulu

Pada Penelitian terdahulu yang mana ditemukan beberapa dari hasil penelitian yang bersangkutan dengan menganalisa data dengan penggunaan metode algoritma C4.5 bisa disimpulkan dibawah ini:

1. Pada Penelitian yang berjudul *Prediksi Kinerja Karyawan Berdasarkan Proses Trainer Menggunakan Data Mining* yang dilakukan oleh (Erlin Elisa & Yunita Tri Utami et al., 2022) mengatakan bahwa Teknik *data mining* dengan metode klasifikasi algoritma C4.5 dalam menganalisa untuk menentukan karyawan yang berhak mendapatkan *reward* di PT. Schneider Electric Manufacturing Batam dengan menggunakan penghitungan dengan

algoritma C4.5 menghasilkan jika variabel yang memiliki nilai *gain* paling tinggi adalah faktor paling penting memilih karyawan yang memiliki hak menerima *reward* dari PT. Schneider Electric Manufacturing Batam memaksudkan adalah memiliki Kinerja yang bagus,tingkat disiplin cukup,mencapai target kerja, bisa bekerja sama dan memiliki kemampuan kerja.

2. Penelitian yang berjudul Analisis Penerimaan Karyawan Posisi Field Collecttor Menggunakan Algoritma C4.5 pada PT.Primas Jamintara Jakarta dilakukan oleh (Ayuni Asistyasari & Taufik Baidawi et al., 2017) menyatakan bahwa hasil penghitungan klasifikasi dari data yang di *training* dengan model *Decision Tree* menggunakan algoritma c4.5dan diperbaiki dengan confusion matrix dihasilkan akurasi senilai 71,54%+-9,13%, 80,47% +/-8,16 untuk precision dan recall sebesar 83,78% +/-9,29% serta dengan angka curva ROC 0,721 dan dapat dinyatakan hasil dari penelitian yang dilakukan ini dapat membantu perekomendasiian kepada PT Primas Jamintara Jakarta dalam menentukan kriteria calon pegawai yang sesuai dengan kriteria dan kebutuhan oleh perusahaan.
3. Dalam sebuah penelitian dengan judul Aplikasi Sistem Penilaian Kinerja Guru Dengan Metode Decision Tree Menggunakan Algoritma ID3 (Studi Kasus SLTP Negeri 3 Marga Sakti Bengkulu Utara) yang dilakukan oleh (Rozali Toyib & Surya Ade Saputera et al., 2019) menyatakan bahwa Hasil dari penilai kinerja masing masing Guru ini sangat terpengaruh oleh jumlah penilaian yang dilakukan oleh Kepala sekolah,nilai dari para Siswa,jenjang

Pendidikan serta lama Masa bekerja. Dan untuk mendapatkan jumlah hasil yang layak untuk nilai angket minimal $\geq 2-3$ dan tidak bisa kurang di angka $\geq 0-2$ untuk hasil yang layak bagi semua yang dilakukan dalam penilaian.

4. Penelitian yang berjudul Implementasi *Data Mining* Algoritma C5.0 dalam Memprediksi penerimaan Cleaning Service (CS) Pada PT. ISS Indonesia Medan yang dilakukan oleh (Rikki Piana, 2019) mengatakan bahwa hasil analisa dapat dirancang menggunakan aplikasi c5.0. Lalu dilakukan uji ulang dengan aplikasi rapid *miner*, dengan penggunaan *datamining* dan ditambahkan algoritma C5.0 untuk menganalisa kelayakan hasil prediksi penerimaan *cleaning service* (cs), informasi yang harus dihasilkan mesti bersifat klasifikasi yaitu mengubah data yang sangat besar menjadi *Decision tree* yang merepresentasikan sebuah aturan.
5. Penelitian yang berjudul Prediksi Promosi Jabatan Karyawan Dengan Algoritma C4.5 (Studi Kasus: Apartemen Senayan Jakarta) yang dilakukan oleh (Sunarti, 2019) mengatakan bahwa Teknik algoritma C4.5 yang sudah dilakukan dengan berdasarkan jumlah total data yang ada. dapat ditarik kesimpulan jika algoritma C4.5 punya kemampuan yang bisa digunakan dalam pengambilan sebuah hasil dari keputusan untuk menentukan dalam prediksi pemberian promosi jabatan karyawan. Hasil pendapatan dari penghitungan teknik klasifikasi data menggunakan teknik pohon keputusan.
6. Penelitian yang berjudul Komparasi Analisis penerimaan Karyawan Dengan Metode C4.5, K-NN Dan Naive Bayes Berbasis PSO yang dilakukan oleh (Ayuni Asistiyasari & Yosep Nuryaman, 2019) mengatakan bahwa algoritma

yang paling tepat yang dapat digunakan dalam teknik klasifikasi dalam penerimaan karyawan adalah teknik algoritma Naïve Bayes dioptimalkan dengan PSO yang dievaluasi dengan confusion matrix menyatakan jika tingkat *accuracy* sebesar 91,25% , *precision* 88,50%, dan 94,17% *recall* lalu nilai grafik AUC 0,903.

7. Penelitian yang berjudul Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai yang Layak Mengikuti Assesment Center untuk Clustering Program SDP yang dilakukan oleh (Iin Parlina, Agus Perdana Windarto, Anjar Wanto, & M. Ridwan Lubis et al., 2018) mengatakan bahwa pengkajian yang dilakukan di penelitian ini, literasi *clustering* /pengelompokan pada data daftar Program SDP terjadi dilakukan sebanyak 3 kali literasi. Berdasarkan hasil cluster dengan menerapkan beberapa standar kriteria dari daftar Program SDP menggunakan algoritma K-Means dapat diambil kesimpulan pengelompokan data dengan rata-rata data program SDP yang dapat melakukan Assesment center lanjutan adalah yang lolos dan hasil klasifikasi program SDP yang Hampir lulus diwajibkan membenahi catatan administrasi seperti tentang Kedisiplinan yang dapat dimulai dari bulan juni hingga berakhir pada bulan oktober agar dapat mengikuti Assesment Center selanjutnya, lalu untuk hasil klasifikasi dengan daftar data program SDP yang Tidak lulus harus melakukan perbaikan data dalam hal kedisiplinan selama waktu 1 Tahun untuk mengikuti program SDP Selanjutnya.

8. Penelitian yang berjudul Pemanfaatan Pohon Keputusan dan Graf Dalam Pemetaan Karyawan Berbasis Kompetensi yang dilakukan oleh (Abdul Rahman Kadafi, Eko Setia Budi, & Raden Bagus Dimas Putra et al, 2018) mengatakan bahwa Teori *Decision Tree* dan teori *graf* bisa dipakai didalam pembuatan pola dalam pemetaan karyawan yang berdasarkan kompetensi. Dalam penyusunan pola tata kelola tersebut, setiap tahapan dalam proses menjadi *Node* dan jalur dari satu tahapan ke tahapan selanjutnya ialah pernyataan yang menjawab masalah pada bagian sebelumnya dan dilakukan pada penelitian yang akan datang.
9. Penelitian yang berjudul Pemodelan Klasifikasi dalam Meningkatkan Proses Pemilihan Calon Karyawan Dengan Metode C4.5 Dan Jaringan Syaraf Tiruan yang dilakukan oleh (Henny L & Angga A P et al., 2020) mengatakan bahwa Berdasarkan hasil pengujian, dilanjutkan dengan pengaplikasian metode dengan menggunakan jaringan syaraf tiruan didapatkan pola yang lebih tepat jika di banding dengan algoritma C4.5. Agar pola bisa dipakai secara menyeluruh, spesifiknya pada bidang-bidang dalam kerjaan ini maka baiknya data yang disatukan berasal dari banyak asal perusahaan yang memiliki kemiripan data ataupun yang sama asalnya atau bisa dari perusahaan yang berada di banyak wilayah yang ada..
10. Penelitian yang berjudul Penentuan Penerimaan Pegawai dengan Metode Eksponensial (MPE) yang dilakukan oleh (Tupan T M, Irwansyah & Agus B et al., 2020) mengatakan bahwa teknik dengan Perbandingan Eksponensial (MPE) bisa diterapkan pada penentuan penerimaan karyawan pada perguruan

xyz. Dan bahasa yang bisa dipakai untuk melakukan penghitungan metode yang membandingkan Eksponensial (MPE) adalah bahasa pemrograman dalam bahasa python.

11. Penelitian yang berjudul Penerapan Algoritma C4.5 Dalam Penerimaan Calon Karyawan PT. Telkom Akses Area Lampung Berbasis Web yang dilakukan oleh (Tosy Caesar Kurniawan et al., 2017) mengatakan bahwa Sistem yang dihasilkan saat ini sudah bisa digunakan untuk membantu HR dalam menentukan proses klasifikasi pelamar yang ada sesuai dengan kebutuhan dan sudah ditetapkan. Sistem ini bisa menolong HR dalam hal pengelolaan data pelamar sehingga waktu untuk mengolah data yang perlukan lebih singkat dan efisien.
12. Penelitian yang berjudul Penerapan Algoritma Decision Tree Dalam Penerimaan Guru pada SMK Sijarul Falah Parung yang dilakukan oleh (Siti Nurlela, Lestari Yusuf, Hermanto, Elah Nurlelah, & Mahmud Syarif et al., 2018) mengatakan bahwa Hasil yang didapat dari data yang telah diproses pada rapid miner menunjukkan angka sebesar 66.00% untuk tingkat akurasi guru pada SMK sirajul falah layak menjadi guru.
13. Penelitian yang berjudul Penerapan *Data Mining* Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pelayanan Sistem Informasi di Program Studi Sistem Informasi yang dilakukan oleh (Dimas A R S, M Safii, & Dedi Suhendro et al., 2018) mengatakan bahwa hasil yang diperoleh dengan adanya analisa kepuasan yang dilakukan pada mahasiswa terhadap kualitas pelayanan SIA di STIKOM Tunas Bangsa yang digunakan dengan teknik

Algoritma C4.5, maka didapatkan hasil bahwa tingkat kepuasan mahasiswa STIKOM Tunas Bangsa bisa ukur dengan sangat jelas. Dari beberapa faktor yang dijadikan sbg patokan dalam pengukur untuk penilaian kepuasan mahasiswa maka hal yang dapat diketahui bawah hal kepuasan yang memiliki tingkat kepuasan tertinggi adalah faktor kualitas sistem informasi yang memadai dan disediakan oleh pihak kampus STIKOM Tunas bangsa kepada para mahasiswanya.

14. Penelitian yang berjudul Penerapan *Particle Swarm Optimization* Pada Algoritma C4.5 Seleksi Penerimaan Karyawan yang dilakukan oleh (Agus Wiyatno et al.,2018) mengatakan bahwa hasil percobaan pengimplementasian *particle swarm optimization* pada model algoritma C 4.5 memberikan nilai akurasi yang lebih tinggi yaitu 85,20% jika dibandingkan dengan algoritma C 4.5 80,80 %. dari hasil tersebut terdapat selisih antara 2 model tersebut sebesar 4,4 %. Sementara untuk evaluasi menggunakan ROC Curve untuk kedua model yaitu, untuk algoritma C 4.5 nilai AUCnya adalah 0,878 dengan tingkat perkiraan *Good Classification* dan untuk model algoritma C 4.5 berbasis *particle swarm optimization* nilai AUCnya adalah 0,891 dengan tingkat perkiraan *Good Classification*. Dari hasil perbaikan dengan ROC Curve tersebut terlihat bahwa model algoritma C 4.5 berbasis *particle swarm optimization* memiliki nilai lebih tinggi jika dibandingkan dengan algoritma C 4.5. dari hasil tersebut diperoleh selisih antara kedua model tersebut sebesar 0,13. sehingga dapat dikatakan bahwa

algoritma C 4.5 berbasis *particle swarm optimization* lebih tepat untuk memprediksi tentang penerimaan karyawan dibandingkan Algoritma C4.5.

15. Penelitian yang berjudul Perbandingan Algoritma C4.5 dan *Classification and Regression Tree* (CART) Dalam Menyeleksi Calon Karyawan yang dilakukan oleh(Ng Poi Wong, Florida N.S. Damanik, Christine, Edward Surya Jaya, & Ryan Rajaya et al., 2019) mengatakan bahwa Algoritma dengan CART memiliki tingkat ketepatan serta kesuksesan yang lebih tinggi jika dibandingkan pada algoritma C4.5 dalam menyeleksi calon karyawan baru, sedangkan algoritma C4.5 memiliki tingkat akurasi untuk keputusan dari hasil melakukan seleksi pelamar yang lebih akurat jika disandingkan dengan algoritma CART. Jumlah perbedaan tingkat ketepatan dari algoritma CART dengan algoritma C4.5 tidak terpengaruh oleh banyak data pelamar yang akan diseleksi dan diuji, lalu untuk tingkat perbedaan keakuratan sistem dan tingkat keberhasilan dari keputusan seleksi pelamar kerja baru dengan menggunakan algoritma C4.5 dan algoritma CART akan terpengaruhi oleh banyaknya jumlah data pelamar yang akan uji.
16. Penelitian yang berjudul Penerapan Agoritma C4.5 untuk menentukan kelayakan pelamar bekerja di PT. Indoland Batam yang dilakukan oleh (James tan, Erlin Elisa et al., 2021) mengatakan bahwa hasil dari analisis tentang pencapaian profit perusahaan, dan setelah dibahas, dihitung dan dilakukan pengujian maupun penghitungan secara konvensional dengan mempergunakan teknik *data mining* dan teknik klasifikasi, dibuat pohon keputusan dengan penggunaan algoritma C4.5 dan dilakukan pengujian

menggunakan aplikasi *open source software* WEKA 3.8.4. yang menghasilkan ketepatan melalui prediksi pencapaian keuntungan. Dengan ini dapat dipakaikan dengan menguraikan beberapa pola untuk mengambil keputusan antara lain:

1. Jika calon karyawan tidak memiliki Pendidikan cukup, maka calon karyawan tidak diterima bekerja di PT Indoland.

17. *In a study entitled A Decision Tree Approach for Predicting Student Grade in Research Project Using Weka by (Ertie C. Abana et al., 2019) say that's This study developed a classification model using a decision tree approach in predicted student grades in Research Project. It was limited to the use to only three decision algorithms which include Random Tree, RepTree and J48. The classification rules extracted from the Random Tree decision tree was used to create a software application that can be used by research instructors in identifying students who need academic counseling to improve their performance in research. The resulting accuracy of the classification models after the cross-validation means more samples and more attributes is still need to arrive with a highly accurate prediction.*

For future work, other decision tree algorithms should be use to analyze the data. The software application that can be developed with this kind of study can also be improved by adding a feature like allowing multiple student data to be analyzed at the same time.

18. *In a study entitled Decision Support Model for Employee Recruitment Using Data Mining Classification by (Clarissa Elfira Amos Pah , Ditdit Nugeraha*

Utama et al., 2020) say that's This paper has explain the detail of the data mining process to classify to the prospective employee whether will be hired or not based on experimental or historical data. Researcher found that C.4.5 decision tree algorithm is the most accurates algorithm to be used to produce classification rules. The data mining classification was indeed a very good way to avoid the subjectivity in decision making because it use the historical data, but in this case study, the collected data tends to be small and even though with high accuracy, there are imbalance total of label class in the dataset (total of "Yes" label and "No" label as target). So in the future case study, we can combine with the other method to deal with the imbalance label class. In this study, the scope of the recruitment process is only limited to the Screening stage, so that the data used were limited for employee personal data, education data and working experience data. In the future, the data collection can be expanded to the results of interviews and other test values such as the results of psychological tests, IQ, EQ, Personality test results, English proficiency tests and other related tests to optimize the data mining classification rules.

19. In a study entited *Decision Support System Of Scholarship Grantee Selection Using Data Mining* by (Eka Sugiyarti, Kamarul Azmi Jasmi, Bushrah Basiron, Miftachul Huda, Shankar K., Andino Maselena et al., 2018) say that's The conclusions that can be illustrated from this journal is that a decision support system (DSS) can be generate for the eligibility of scholarship recipients for the students who get a scholarship decision support

system (DSS) with data mining method of C4.5 algorithm can help to better and make better effectiveness than before as well as in earnings reports obtained can be minimized in error when analyze.

20. *In a study entitled A Proposed Decision Tree Classifier for Atherosclerosis Prediction and Classification by (Yousef K Qawqzeh, Mohammad Mahmood Ootom , Fayez Al-Fayez, Ibrahim Almarashdeh, Mutasem Alsmadi, Ghaith Jaradat et al., 2019) say that's This proposed model is developed for atherosclerosis diagnosis as a surrogate measure to CIMT test. The ability of the developed classifier is measured in terms of overall accuracy, sensitivity, specificity, and the AUC measures. The model obtained an overall accuracy of 82.6%, sensitivity of 89.3%, specificity of 69.2%, and AUC of 90.5% respectively. The results were satisfactory enough since the proposed classifier works fine in predicting the atherosclerosis incidents in which it has high accuracy based on decision tree machine learning classification method. Utilizing features extracted from a non-invasive technique, the PPG, to predict atherosclerosis represent a promising technology that has low cost, mobility, smallsize, affordable, and easy setup to be adopted in clinical settings. The proposed decision tree based classifier could assist in atherosclerosis early detection and prevention by predicting its high-risk in an early stage. This without doubt might deliver a fruitful benefit for society and specialists in term of disease prediction and risk preventions.*

21. *In a study entitled Application of Classification Technique of Data Mining for Employee Management System by (Adhana J. Kamatkar, Amarapali Tayade ,*

Amelec Vilorio , and Ana Hernández-Chacín., 2018) say that's Many researches apply data mining technologies on biology, network, and financial data, etc. But we have classified the employment data using Data mining tools. As it is a powerful analytical tool that enables institutions to allocated resources and staff efficiently and proactively manage employees' outcomes. This paper discusses, WEKA data mining toolkit classifier model to predict employee performance based on the employee age, date of joining and number of years of experience. The outcome shows that Experience attribute had the maximum gain ratio, Ages attribute shows a positive effect on performance and employees between the age of 35 years and 60 years have shown better performance. This equips the management to take better decisions regarding the employees who need special attention and to enhance management's