



BAB I
PENDAHULUAN

BAB I

PENDAHULUAN

1.1. Latar Belakang

Sistem penilaian kinerja karyawan adalah sebuah sistem yang digunakan untuk menilai kinerja terbaik para karyawannya. Perusahaan melakukan penilaian kinerja terbaik terhadap karyawan untuk mengevaluasi, memotivasi, memverifikasi dan meningkatkan kinerjanya. Hasil kinerja ini menjadi alat untuk membantu pengambilan keputusan seperti promosi, pemberhentian, mutasi, pemberian bonus atau memberikan umpan balik bagi karyawan. (Rully : 2019). Penilaian kinerja ini dilakukan dengan tujuan untuk promosi jabatan atau pengisian jabatan yang kosong. Namun, perusahaan tidak memiliki sistem yang dapat membantu proses penilaian kinerja sehingga mengakibatkan tidak termonitornya kompetensi kinerja dengan baik (Eka, dkk : 2019).

Sistem penilaian kinerja karyawan yang berjalan pada PT. Raharja Sinergi Komunikasi yaitu penilaian kinerja ini dilakukan dengan tujuan untuk promosi jabatan yang kosong. Dengan adanya promosi jabatan karyawan akan bekerja dengan baik dan bertanggung jawab. Permasalahan yang dihadapi dalam penilaian kinerja karyawan adalah perusahaan kesulitan menentukan karyawan yang akan dipromosikan karena belum adanya proses penilaian kinerja yang dilakukan secara pasti dalam hal ini hanya dilakukan secara sepihak. Kualitas kerja karyawan juga seringkali berubah-ubah. Ini sangat berpengaruh pada prospek perusahaan

kedepannya. Kesibukan seorang manager membuat tatap muka dengan karyawan kurang. Tentu saja ini membuat pengawasan pada kinerja karyawan sulit dilakukan.

Penyelesaian masalah penilaian kinerja pada PT. Raharja Sinergi Komunikasi yaitu dengan membuat sistem penilaian kinerja pendekatan data mining. Sistem ini menggunakan algoritma C4.5 yang menggunakan teknik data mining untuk membuat pohon keputusan, algoritma ini dimulai dengan memasukkan data training ke dalam simpul akar pada pohon keputusan. Algoritma C4.5 adalah algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi atau pengelompokan dan bersifat prediktif dan digunakan untuk membentuk pohon keputusan (*Decision Tree*) (Hafizan, dkk : 2020). Pohon keputusan merupakan metode klasifikasi dan prediksi. Pohon keputusan berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target (Hendra, dkk : 2023). Algoritma C4.5 ini menggunakan konsep data mining yang melakukan proses penggalian informasi pada keputusan-keputusan sebelumnya (Tanjung : 2021).

Berdasarkan permasalahan yang telah dijabarkan dan gagasan solusi pemecahan masalah yang ada, maka penulis tertarik mengangkat judul : **“Penerapan Data Mining Dalam Sistem Penilaian Kinerja Karyawan Menggunakan Metode Algoritma C4.5”**. Dengan menerapkannya pendekatan data mining menggunakan metode Algoritma C4.5 penelitian ini diharapkan dapat membantu perusahaan menilai karyawannya secara objektif. Algoritma C4.5 mampu menghasilkan keputusan kompleks menjadi lebih sederhana, sehingga

pengambilan keputusan akan lebih menginterpretasikan solusi dari permasalahan.

1.2. Ruang Lingkup Permasalahan

1.2.1. Identifikasi Masalah

Identifikasi permasalahan dari penelitian ini adalah sebagai berikut :

1. PT. Raharja Sinergi Komunikasi kesulitan menilai karyawan yang akan dipromosikan karena belum adanya proses penilaian kinerja yang dilakukan secara pasti dalam hal ini hanya dilakukan secara sepihak.
2. Kualitas kerja karyawan juga seringkali berubah-ubah, ini sangat berpengaruh pada prospek perusahaan kedepannya.
3. Pengawasan pada penilaian kinerja karyawan sulit dilakukan karena kesibukan seorang manager membuat tatap muka dengan karyawan berkurang.

1.2.2. Rumusan Masalah

Sebagaimana yang telah dikemukakan pada latar belakang masalah di atas, maka rumusan masalahnya adalah sebagai berikut:

1. Bagaimana sistem penilaian kinerja karyawan secara pasti pada PT. Raharja Sinergi Komunikasi?
2. Bagaimana menerapkan metode Algoritma C4.5 dalam penilaian kinerja karyawan pada PT. Raharja Sinergi Komunikasi ?
3. Bagaimana merancang dan membangun sistem penilaian kinerja karyawan pada PT. Raharja Sinergi Komunikasi?

I.2.3. Batasan Masalah

Untuk menghindari meluasnya masalah dari topik pembahasan penelitian ini, maka pembahasan masalah hanya mencakup hal-hal sebagai berikut :

1. Aplikasi hanya untuk sistem penilaian kinerja PT. Raharja Sinergi Komunikasi.
2. Data *input* yang digunakan adalah data pegawai, data jabatan, data perspektif penilaian.
3. Data *output* yang dihasilkan adalah laporan hasil penilaian kinerja.
4. Variabel yang digunakan dalam sistem penilaian kinerja karyawan adalah Masa Kerja, Prestasi, Loyalitas, Kualitas dan Hasil Kerja.
5. Metode yang digunakan adalah metode Algoritma C4.5
6. Bahasa pemrograman yang digunakan yaitu *PHP*.
7. Database yang digunakan yaitu *MySQL*.
8. Perancangan aplikasi pada sistem ini menggunakan pemodelan *Unified Modeling Language* (UML).

1.3. Tujuan dan Manfaat

1.3.1. Tujuan Penelitian

Adapun yang menjadi tujuan dalam penelitian ini adalah :

1. Untuk merancang dan membangun sistem penilaian kinerja karyawan pada PT. Raharja Sinergi Komunikasi.
2. Untuk menerapkan data mining menggunakan metode Algoritma C4.5 dalam penilaian kinerja karyawan pada PT. Raharja Sinergi Komunikasi.

3. Untuk mengetahui kriteria penilaian kinerja karyawan secara objektif pada PT. Raharja Sinergi Komunikasi.

1.3.2. Manfaat Penelitian

Adapun yang menjadi manfaat dalam penelitian ini adalah :

1. Membantu PT. Raharja Sinergi Komunikasi dalam sistem penilaian kinerja karyawan secara pasti yang ditentukan melalui variabel yang sudah ditetapkan.
2. Mengetahui penerapan data mining menggunakan metode Algoritma C4.5 dalam sistem penilaian kinerja karyawan pada PT. Raharja Sinergi Komunikasi.
3. Mempermudah PT. Raharja Sinergi Komunikasi dalam melakukan penilaian kinerja karyawan.

1.4. Metodologi Penelitian

1.4.1. Metode Pengumpulan Data

Pengumpulan data menggunakan beberapa tahapan yang peneliti lakukan untuk mendapatkan data-data yang dibutuhkan yaitu :

1. Studi Lapangan (*Field Research*)

Merupakan suatu metode yang dilakukan langsung ke lapangan untuk mengumpulkan data. Adapun pengumpulan data dilakukan penulis adalah :

- a. Pengamatan (*Observation*)

Yaitu dengan melakukan pengamatan terhadap sistem penilaian kinerja

pada PT. Raharja Sinergi Komunikasi.

b. Wawancara (*Interview*)

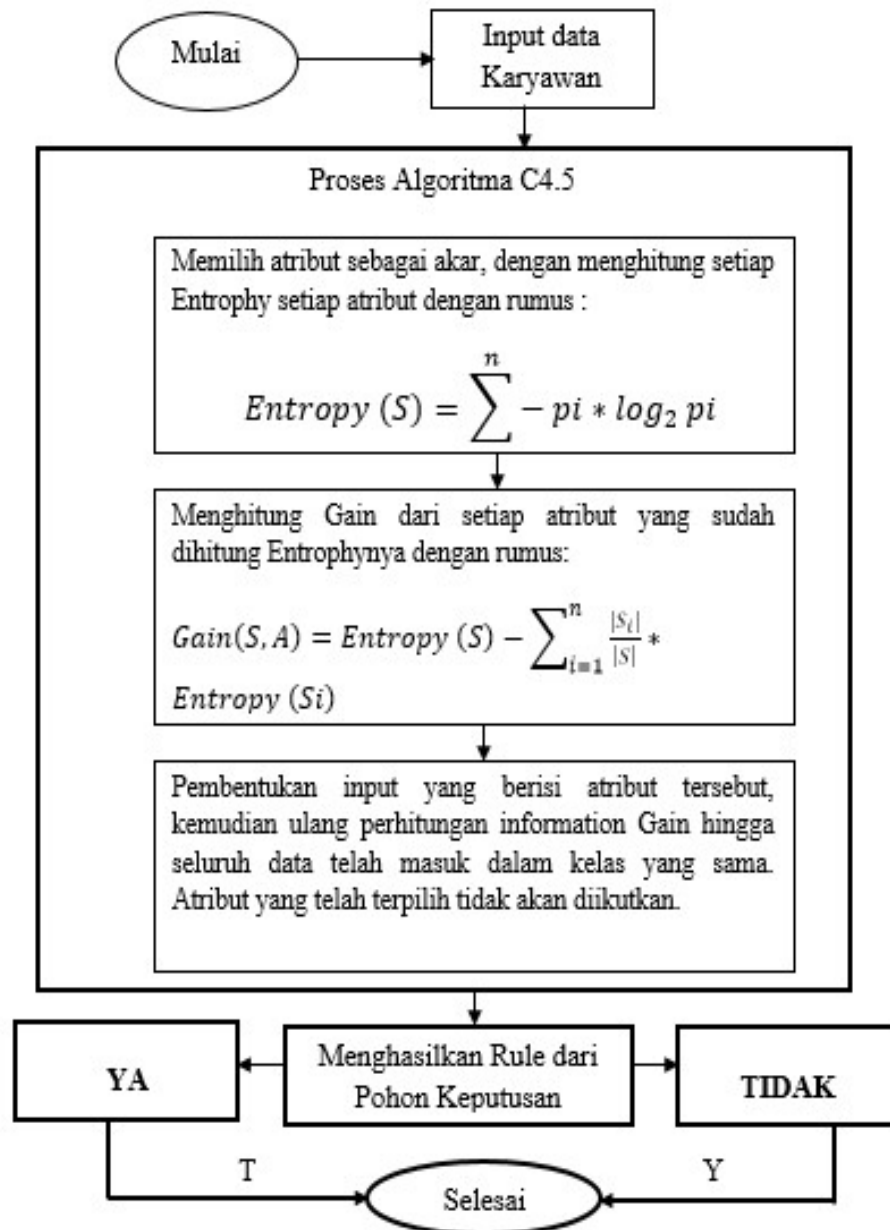
Wawancara merupakan kegiatan untuk mendapatkan sebuah informasi terkait tema yang diangkat dalam sebuah penelitian dengan cara mengajukan pertanyaan kepada perusahaan terkait masalah yang diteliti.

2. Studi Kepustakaan (*Library Research*)

Penulis melakukan studi pustaka untuk mendapatkan referensi yang berhubungan dengan penulisan skripsi dari berbagai sumber bacaan seperti buku panduan dan jurnal terkait yang membahas tentang konsep sistem penilaian kinerja.

1.4.2. Metodologi Penelitian

Metode penelitian yang digunakan pada penelitian ini adalah Algoritma C4.5. Dalam pemodelan Algoritma C4.5 ini dapat dilihat pada Gambar 1.1. berikut:



Gambar 1.1. Pemodelan Algoritma C4.5

Dalam pemodelan algoritma C4.5 dilakukan pada sampel *dataset*, kemudian menghitung *entropy* (S) dari keseluruhan atribut, atribut yang memiliki gain tertinggi sebagai akar/node. Selanjutnya buat cabang untuk

tiap nilai, bagi kasus dalam cabang, ulangi perhitungan *Gain* sampai semua data masuk ke dalam kelas yang sama. Atribut yang telah dipilih tidak lagi dibutuhkan dalam perhitungan, proses pembentukan pohon keputusan terhenti jika sudah tidak ada lagi atribut yang dipartisi dan semua tuple dalam *node* N telah memiliki kelas yang sama.

1.5. Kontribusi Penelitian

Penelitian yang dilakukan oleh Savitri (2021) yang berjudul “Analisa Metode Algoritma C4.5 Untuk Penilaian Kinerja Karyawan PT. PLN Dsitribusi Jakarta dan Tangerang”. Dalam hal ini hasil dari perhitungan algoritma C4.5 memiliki tingkat akurasi sebesar 84,85% dan nilai akurasi prediksi direkomendasi sebesar 84,21%. Dan variabel kuantitas pekerjaan sebagai nilai gain paling tinggi dengan nilai 0,2569, pada variabel komitmen kerja nilai gain paling rendah dengan nilai 0. Metode Algoritma C4.5 hanya membutuhkan waktu 3 atau 4 hari, cara yang dilakukan hanya tinggal memasukan angka-angka yang akan di hitung dengan metode yang sudah ditentukan ke dalam tools.

Penelitian yang dilakukan oleh Nur Fauza Muhidin (2022) yang berjudul “Analisis dan Penerapan Algoritma C4.5 Untuk Memprediksi Kualitas Penelitian dan Publikasi”. Penelitian ini dilakukan membantu program studi untuk menilai kualitas penelitian dan publikasi ilmiah. Dari penelitian ini dapat disimpulkan bahwa menggunakan algoritma C4.5 maka kualitas penelitian dan publikasi ilmiah dapat diprediksi dengan akurasi yang cukup tinggi, hasil pengukuran akurasi data yang diperoleh dari data pengujian pelatihan adalah 81,65%.

Menurut Hijrah, dkk, (2022) pada penelitian yang berjudul “Analisis RapidMiner dan Weka Dalam Memprediksi Kualitas Kinerja Karyawan Menggunakan Metode Algoritma C4.5”. Pengujian data kinerja karyawan pada penelitian ini dilakukan dengan metode C45, dari pengujian menggunakan aplikasi data mining Rapid Miner 7.6.0.0.1 dan Weka 3.8.1. maka diperoleh hasil aplikasi rapidminer lebih baik dari segi Accuracy sebesar 94,12 % dan Recall 95,6 %. Sedangkan aplikasi weka lebih unggul dari segi Precision 98,6 %. Pengujian aplikasi data mining ini dapat menghasilkan nilai yang berbeda jika pengujiannya dilakukan dengan data uji yang berbeda serta algoritma yang digunakan juga berbeda.

Menurut Ikhsan dan Boy (2020) pada penelitian yang berjudul “Evaluasi Penilaian Kinerja Dalam Klasifikasi Data Mining Dengan Metode Naïve Bayes”. Pada penelitian ini hanya memanfaatkan data training guna menghasilkan probabilitas setiap atribut dengan kelas yang berbeda, sehingga perhitungan tersebut menghasilkan nilai yang dapat dioptimalkan untuk menentukan klasifikasi kompeten dan tidak kompeten karyawan dan membantu mengevaluasi penilaian kinerja menjadi lebih akurat.

Menurut Aditia, dkk, (2022) pada penelitian yang berjudul “Analisis Klasterisasi Penilaian Kinerja Pegawai Menggunakan Metode Fuzzy C-Means (Studi Kasus: Pengadilan Tinggi Bandar Lampung). Penetapan nilai yang di peroleh pegawai hanya didasarkan pada pengurutan pusat cluster hasil pengolahan total nilai pegawai menggunakan FCM. Dari hasil analisis pengelompokan FCM dengan 5 cluster dengan 35x iterasi diperoleh fungsi objektif sebesar 111.949781.

Dimana kelompok pertama terdiri dari 940 pegawai, klaster ke dua 692 pegawai, kelompok 23 pegawai, kelompok ke empat terdiri dari 8 pegawai dan kelompok ke lima terdiri dari 17 pegawai. Dari hasil centroid ini telah dianalisa bahwa rata-rata nilai pegawai memiliki nilai sangat baik pada fitur nilai komitmen dan nilai kesidipilinan.

Sedangkan penelitian yang dilakukan oleh penulis dengan judul “Penerapan Data Mining Dalam Sistem Penilaian Kinerja Karyawan Menggunakan Metode Algoritma C4.5 Pada PT.Raharja Sinergi Komunikasi”. Dalam sistem penilaian kinerja ini menggunakan metode tersendiri dalam kategori penilaian dan pemberian nilai yaitu menggunakan metode Algoritma C4.5. Metode Algoritma C4.5 merupakan algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi atau pengelompokan dan bersifat prediktif dan digunakan untuk membentuk pohon keputusan (*Decision Tree*). Dengan menerapkan metode Algoritma C4.5, diharapkan dapat membantu perusahaan menilai kinerjanya secara objektif. Perancangan pada sistem penilaian kinerja ini menggunakan model *Unified Modeling Language*.

1.6. Lokasi Penelitian

Penelitian ini dilakukan pada PT. Raharja Sinergi Komunikasi yang beralamat di Jl.Karya Gg. Kartini No.16, Kota Medan, Sumatera Utara.

1.7. Sistematika Penulisan

Adapun sistematika penulisa dari penelitian ini adalah sebagai berikut :

BAB I : PENDAHULUAN

Pada bab ini berisikan dasar pemikiran, kebutuhan atau alasan yang menjadi ide penulis untuk mengakat judul tersebut menjadi judul skripsi, terdiri dari latar belakang, ruang lingkup masalah, tujuan dan manfaat, metodologi penelitian, keaslian penelitian, lokasi penelitian dan sistematika penulisan.

BAB II : TINJAUAN PUSAKA

Pada bab ini merupakan bab studi literatur yang berisikan tentang teori- teori yang berkaitan langsung dengan permasalahan yang dibahas. Adapun tinjauan pustaka yang diuraikan oleh penulis adalah : metodologi waterfall, UML (*Unified Modeling Language*), pengertian basis data, dan normalisasi.

BAB III : ANALISA DAN DESAIN SISTEM

Pada bab ini berisikan analisa masalah pada sistem yang berjalan, strategi penyelesaian masalah, penerapan metode/algoritma, desain sistem baru, menggunakan *use case diagram*, *class diagram*, *activity diagram* dan *sequence diagram*, *desain database* (normalisasi dan desain tabel) dan desain *user interface*.

BAB IV : HASIL DAN UJI COBA

Pada bab ini berisikan hasil dari sistem penilaian kinerja dan pengujian yang dilakukan pada sistem penilaian kinerja yang sudah

dibangun menggunakan skenario pengujian dan hasil pengujian.

BAB V : KESIMPULAN DAN SARAN

Pada bab ini berisikan tentang kesimpulan dari pemecahan masalah yang telah didefinisikan sebelumnya serta saran berisikan kelemahan sistem yang dibangun dan dianggap penting untuk diperhatikan atau dijalankan pada masa yang akan datang untuk kesempurnaan hasil penentuan.



BAB II

TINJAUAN PUSTAKA

BAB II

TINJAUAN PUSTAKA

II.1. Penelitian Terkait

Adapun penelitian terkait yang akan digunakan sebagai sumber acuan yang relevan dan terkini yaitu:

Menurut (Gustiana, 2021) pada penelitian yang berjudul Penerapan Algoritma C4.5 Dalam Sistem Pendukung Keputusan Evaluasi Kinerja Fasilitator Pamsimas (Studi Kasus Di kabupaten Kampar). Penelitian ini bertujuan membantu dan mempermudah penilai dalam mengevaluasi kinerja fasilitator PAMSIMAS dengan menerapkan Data Mining yaitu Algoritma C4.5 sebagai Sistem Pendukung Keputusan. Metode yang digunakan dalam penelitian ini adalah metode algoritma C4.5 dengan menggunakan atribut hasil evaluasi sebagai atribut label tujuan sebanyak 4 buah sub-atribut, yaitu: Sangat baik, baik, cukup dan buruk. Penelitian ini menghasilkan pohon keputusan dari 20 berupa value atribut tujuan hasil evaluasi, algoritma C4.5 dapat merepresentasikan sebanyak 13 pohon keputusan dari value atribut tujuan hasil evaluasi dan persentasi akurasi dari pengujian adalah sebesar 50%. Kesimpulan dari penelitian ini yaitu algoritma C4.5 membantu mengklasifikasikan hasil evaluasi kinerja berdasarkan atribut waktu, kelengkapan isian data, respon perbaikan dan tanggung jawab dalam tim. Penelitian dapat membentuk pohon keputusan yang dapat memberikan informasi hasil evaluasi kinerja fasilitator PAMSIMAS Kabupaten Kampar.

Menurut (Hidayah & Rozi, 2021) pada penelitian yang berjudul Penerapan Data Mining Dalam Menentukan Kinerja Karyawan Terbaik Dengan Menggunakan Metode Algoritma C4.5 (Studi Kasus: Universitas Mercu Buana Yogyakarta). Pada penelitian ini dilakukan klasifikasi atau segmentasi atau pengelompokan dan bersifat prediktif yang digunakan untuk membentuk pohon keputusan (Decision Tree). Analisa ini akan membantu mempermudah pihak Pusat Penjaminan Mutu (PPM) Universitas Mercu Buana Yogyakarta dalam menentukan dosen terbaik. Penerapan Algoritma C4.5 dalam penilaian dosen terbaik di Universitas Mercu Buana Yogyakarta dalam penelitian ini memiliki tingkat akurasi yang termasuk dalam klasifikasi sangat baik yaitu sebesar 85,52% yang didapat dari hasil uji coba menggunakan tools Rapid Miner dengan 80% data sebagai data training dan 20% data uji.

Menurut (Bachtiar & Mahradianur, 2023) pada penelitian yang berjudul Analisis Data Mining Menggunakan Metode Algoritma C4.5 Menentukan Penerima Bantuan Langsung Tunai. Tujuan peneltian ini untuk memberikan masukan tentang menentukan apakah keluarga tersebut masih berhak menerima atau tidak berhak menerima bantuan BLT, dimana masih ada keluarga yang berhak menerima dan memenuhi syarat dari program bantuan pemerintahan ini. Sumber data yang diperoleh dapat dari Kantor Desa Regei Lestari. Metode yang dipakai pada penelitian merupakan teknik data mining dengan algoritma C4.5 yang diterapkan pada software RapidMiner. Atribut yang dipakai untuk menentukan keluarga yang berhak menerima dan tidak berhak menerima pada program bantuan BLT yaitu Penghasilan, Status,

Pekerjaan, Umur. Hasil dari klasifikasi data menggunakan metode algoritma C4.5 dan pengujian dengan RapidMiner adalah Jika Pekerjaan = buruh tani maka hasilnya adalah menerima dengan nilai 16 item dari menerima dan 0 yang tidak menerima, maka didapatkan faktor yang menentukan yang berhak menerima dan yang tidak berhak menerima bantuan langsung tunai (BLT).

Menurut (Ramadhan et al., 2020) pada penelitian yang berjudul Penerapan Data Mining Menggunakan Algoritma C4.5 dalam Mengukur Tingkat Kepuasan Pasien BPJS. Tujuan penelitian ini adalah untuk mengetahui kualitas pelayanan terhadap kepuasan pasien BPJS yang ditinjau dari aspek Reliability (Kehandalan), Responsiveness (Ketanggapan), Assurance (Asuransi), Emphaty (Empati), Tangibles (Bentuk Nyata) pada pasien BPJS RS. Balimbingan. Pada RS Balimbingan kelima aspek tersebut belum terukur dengan pasti, sehingga pihak RS. Balimbingan merasa kesulitan untuk menentukan aspek mana yang harus ditingkatkan. Dengan menggunakan algoritma C4.5, penulis mencoba mengukur kelima aspek tersebut sehingga terbentuk pohon keputusan (decision tree). Setelah dilakukan perhitungan secara manual, maka dilakukan pembuktian dengan menggunakan software RapidMiner.

Menurut (Meilina, 2022) pada penelitian yang berjudul Penerapan Data Mining Dengan Metode Klasifikasi Menggunakan Algoritma C4.5. Pada penelitian ini penulis menggunakan metode klasifikasi Data Mining yaitu dengan Algoritma C4.5 untuk mengklasifikasikan kegiatan dalam memperoleh kegiatan dengan Skala Sangat Prioritas, Prioritas dan Belum Prioritas. Sampel sumber data diambil dari data investasi kegiatan

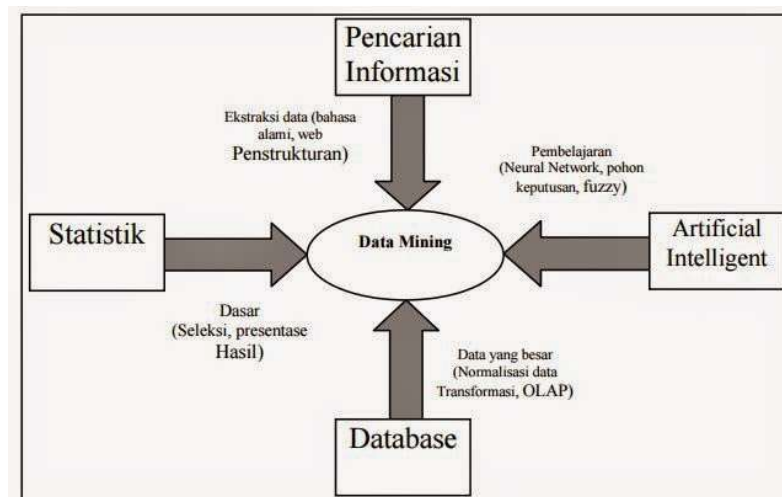
Program KOTAKU di Kabupaten Kuningan. Dari hasil pengujian yang diperoleh untuk nilai Accuracy Algoritma C4.5 adalah sebesar 95,98%. Dengan demikian Algoritma C4.5 merupakan algoritma dan teknik terbaik untuk mengklasifikasikan data investasi kegiatan dalam memperoleh dana bantuan yang bersumber dari pemerintah.

Berdasarkan penelitian diatas ditemukan beberapa perbandingan yang akan diteliti oleh penulis. Pada penelitian ini penulis menggunakan metode klasifikasi Data Mining yaitu dengan Algoritma C4.5 dengan menerapkan sistem berbasis aplikasi yang dapat menentukan hasil pojon keputusan beserta rule dan sistem yang dibangun bersifat online.

II.2. Landasan Teori

II.2.1. Data Mining

Data mining atau *machine learning* bisa dimanfaatkan dan sering digunakan dalam melakukan berbagai bentuk kegiatan analisis untuk menjelaskan ilmu dan pengetahuan secara jelas dan mudah dipahami oleh pikiran yang dimiliki manusia (Hermawan, Sukma, & Halfis, 2019).



Gambar II.1. Bidang Ilmu Data Mining
(Sumber : Hermawan, Sukma, & Halfis, 2019)

Dalam prosesnya, data mining memiliki teknik dalam pengolahan data tersendiri hingga terbentuk sebuah pola, kemudian pola tersebut difungsikan untuk dapat mengenali pola lain yang tidak tersimpan dalam lokasi penyimpanan data dan informasi yang sama. Dalam kegiatan atau proses melakukan prediksi, banyak peneliti melakukan kegiatan tersebut dengan memanfaatkan fungsi dari Teknik data mining. Awal mula istilah data mining yaitu karena data dan informasi yang akan diproses sudah lama dan cara pengolahannya dilakukan seperti kegiatan menambang hingga mendapatkan hasil yang berharga dan bermanfaat (Hermawan et al., 2019).

II.2.2. *Decision Tree*

Akibat dari berbagai jenis permasalahan dan tingkat kesulitan dalam permasalahan yang berbeda dan sering terjadi dalam kehidupan manusia pada akhirnya membuat manusia mulai melakukan berbagai upaya dalam mengembangkan sebuah pola berbentuk sistem agar dapat membantu mereka dalam menyelesaikan berbagai jenis dan bentuk permasalahan yang terjadi dan salah satu hasilnya adalah decision tree atau biasa dikenal dengan istilah pohon keputusan. Decision tree merupakan salah satu bentuk implementasi pola dengan menerapkan berbagai tahapan dan proses yang terdapat dalam kegiatan klasifikasi dan juga kegiatan prediksi. Decision tree menghasilkan alternatif dalam memecahkan masalah dengan hasil yang kuat dan berguna sehingga decision tree terkenal dan banyak digunakan oleh peneliti saat melakukan penelitian dengan teknik data mining.

mengubah data yang ada menjadi sebuah pola berbentuk aturan sehingga dapat dipahami dengan mudah oleh manusia. Dalam tahapannya, decision tree melakukan proses yang dimulai dari keseluruhan bagian-bagian yang terdapat dalam data yang berisi berbagai kondisi dan keterangan yang lengkap dan kemudian diakhiri dengan sebuah keputusan. Arsitektur dari decision tree memang dibuat mirip dengan struktur yang terdapat pada sebuah pohon dan penjelasannya, yaitu:

1. Simpul bagian akar

Simpul dari bagian akar terletak pada bagian paling atas dan pertama dari struktur yang terdapat pada decision tree.

2. Simpul internal

Simpul dari bagian internal ini adalah percabangan lanjut dari simpul bagian akar. Dalam sebuah simpul ini apabila terdapat satu masukan maka dapat mengeluarkan keluaran dengan jumlah maksimal sebanyak dua.

3. Simpul daun

Simpul daun merupakan simpul yang berada pada ujung bagian dari pohon.

Dalam simpul daun terdapat sebuah masukan namun tidak memiliki keluaran

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

1. Kelebihan dari Decision Tree

- a. Dalam penggunaan decision tree apabila terdapat cakupan data yang rumit maka dirubah menjadi lebih sederhana untuk kemudian dijadikan dasar pengambilan keputusan.
- b. Kelebihan dari penggunaan decision tree yaitu dapat menghapus cara

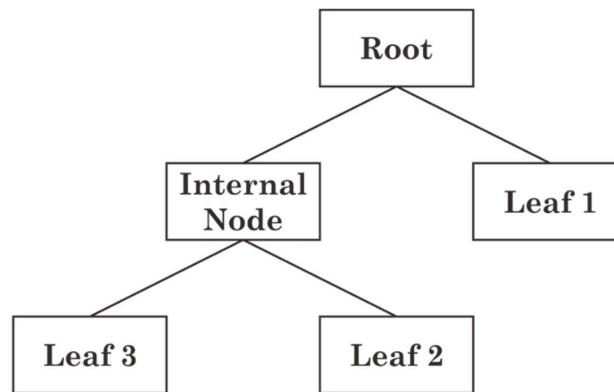
perhitungan yang tidak perlu sehingga dalam melakukan proses dari pengujian hanya menggunakan kriteria sesuai dengan keperluan untuk pengujian saja.

- c. Dalam proses pemilihan akan kelengkapan atau keterangan pada data, bila menggunakan decision tree maka hasil pemilihan terhadap kelengkapan atau keterangan data akan menjadi pembeda dalam mengelompokkan data berdasarkan kriteria lainnya.
- d. Decision tree bertindak untuk menghindari permasalahan lain dengan memanfaatkan penggunaan jenis kelompok atau kategori dengan jumlah sedikit tanpa mengurangi kualitas hasil keputusan akhir.

2. Kekurangan dari Decision Tree

- a. Dalam menggunakan decision tree apabila jumlah dari hasil keputusan yang harus dihasilkan dan kriteria yang digunakan memiliki jumlah yang sangat banyak, maka waktu yang dibutuhkan dalam menghasilkan keputusan akan semakin bertambah serta dibutuhkan memori sistem pendukung yang tinggi.
- b. Dalam decision tree terdapat akumulasi dari jumlah error yang terjadi dalam setiap bagian dari struktur yang terdapat pada decision tree.
- c. Membuat struktur dari decision tree dengan hasil yang maksimal ternyata sulit untuk dilakukan.
- d. Apabila menggunakan decision tree, maka hasil atau kualitas dari keputusan yang dihasilkan sangat ditentukan dari model atau struktur

decision tree yang dibuat.



Gambar II.2. Model Decision Tree
(Sumber : Muhammad Ilham Mubarok ; 2018)

II.2.3. Algoritma C4.5

Algoritma dari C4.5 bermanfaat dalam melakukan proses klasifikasi data yang menggunakan angka atau yang menggunakan kategori. Setelah tahapan dari proses klasifikasi dilakukan maka kemudian hasilnya adalah beberapa aturan dari pola dan dapat dipergunakan untuk melakukan tahapan dari proses prediksi dengan nilai-nilai prediksi yang berasal dari perekaman data yang lama hingga terbaru. Jenis algoritma dari C4.5 berasal dari algoritma ID3, maksudnya adalah algoritma dari C4.5 adalah hasil dari perbaikan lebih lanjut yang dilakukan terhadap algoritma dari ID3 sehingga dalam algoritma C4.5 memiliki kemampuan dalam mengatasi permasalahan saat terjadi ketiadaan data dan mengatasi permasalahan saat terjadi data yang bersifat pecahan.

1. Pada umumnya algoritma dari C4.5 digunakan dalam proses pembangunan struktur dari decision tree, dan demikian penjelasan proses atau tahapannya yaitu: Pertama memilih keterangan data sebagai akar dari decision tree.

2. Kemudian dilanjutkan dengan membuat struktur cabang dari decision tree yang sudah berisi nilai-nilai.
3. Pada tahap ketiga, dilakukan pembagian atau pengelompokkan dari permasalahan yang terdapat pada struktur cabang dari decision tree.
4. Dan tahap akhir adalah melakukan proses pengulangan yang ada pada tahapan pertama hingga ketiga sampai seluruh cabang pada struktur decision tree memiliki kelas dan kategori yang sama.

Dalam melakukan tahap awal dari decision tree yaitu pada saat melakukan pemilihan terhadap kelengkapan atau keterangan data dilakukan dengan cara mencari nilai dari gain dengan hasil yang tertinggi dari keluruhan nilai yang dimiliki oleh kelengkapan atau keterangan yang ada. Demikian rumus untuk melakukan pencarian nilai gain tertinggi, yaitu:

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

Rumus Perhitungan Gain

Keterangan simbolnya, yaitu:

S adalah himpunan

A adalah keterangan atau kelengkapan yang digunakan.

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

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

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

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

Rumus Perhitungan Entropy

Keterangan simbolnya, yaitu:

S adalah himpunan A adalah keterangan atau kelengkapan yang digunakan.

n adalah jumlah partisi yang terdapat pada keterangan atau kelengkapan bagian

$S p_i$ adalah proporsi dari S_i kepada S

II.2.4. Studi Kasus Metode Algoritma C4.5

Penelitian yang dilakukan oleh Alfi Novia Zahrotul Hidayah dan Anief Fauzan Rozi (2021) yang mengangkat judul Penerapan Data Mining Dalam Menentukan Kinerja Karyawan Terbaik Dengan Menggunakan Metode Algoritma C4.5 (Studi Kasus : Universitas Mercu Buana Yogyakarta). Analisa yang dilakukan menggunakan metode Algoritma C4.5 yang menghasilkan output berupa pohon keputusan untuk menentukan penilaian kinerja karyawan (dosen) terbaik. Berikut merupakan tabel rentang penilaian dapat dilihat pada gambar dibawah ini:

Tabel II.1. Rentang Penilaian

Nilai	Keterangan
1	Buruk
2	Kurang baik
3	Cukup baik
4	Baik
5	Sangat Baik

(Sumber : Alfi Novia Zahrotul Hidayah, Anief Fauzan Rozi ; 2021)

Pada proses ini akan dilakukan proses seleksi data yang akan digunakan sebagai atribut penelitian.

A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
3	3	3	3	4	3	4	4	3	3	3	3	4	4	3	3	4	3	3
4	4	5	4	5	5	5	4	4	4	4	4	4	4	4	4	5	4	4
4	4	5	4	5	5	5	4	4	4	4	4	4	4	4	4	5	4	4
4	4	4	4	4	4	4	4	3	4	4	4	3	4	4	4	4	4	4
4	4	5	4	4	4	5	4	4	4	5	4	3	3	4	4	4	3	4
4	4	4	4	5	5	4	4	4	3	3	3	4	4	4	3	4	4	4
4	3	4	4	3	3	4	4	4	4	4	3	3	4	4	3	4	3	4
5	3	4	3	5	5	5	4	3	5	4	4	3	5	4	4	3	4	4
3	4	3	3	5	3	4	4	3	4	3	2	4	3	4	3	1	4	3
3	4	4	4	5	4	3	4	5	5	4	3	4	5	5	5	4	5	4
4	4	3	3	4	4	3	4	4	3	3	3	3	4	4	4	4	4	4
3	3	3	3	4	4	4	3	4	3	3	4	4	3	3	3	4	3	3
3	4	4	3	4	5	5	5	4	4	4	3	3	4	4	4	4	4	4
4	4	4	3	5	5	4	4	3	3	4	3	3	3	3	3	5	3	5
2	3	3	2	4	4	5	5	3	3	3	2	2	2	2	3	3	4	3
3	4	4	4	4	4	4	4	4	4	3	4	4	3	4	4	3	4	4
3	4	4	4	4	4	4	3	4	4	3	3	3	3	3	4	4	4	4
3	4	4	4	5	5	4	4	4	4	3	3	3	3	3	3	4	4	4
1	2	2	2	4	4	4	4	4	4	2	3	3	3	2	2	4	4	4
2	4	3	2	4	4	4	3	4	4	2	3	2	3	2	3	4	3	3
3	3	3	4	4	3	3	3	4	4	4	4	3	4	5	4	5	3	4

Gambar II.3. Data Atribut
(Sumber : Alfi Novia Zahrotul Hidayah, Anief Fauzan Rozi ; 2021)

Data yang diambil pada tahap ini ialah data setiap pertanyaan perkategori lalu perkategori akan dilakukan proses perhitungan untuk mencari rata-rata agar memudahkan proses pemrosesan data pada *Rapid Miner* nantinya seperti pada tabel

Tabel II.2. Rata-rata

A1-A4	B1-B4	C1-C11
3	3,75	3,272727
4,25	4,75	4,090909
4,25	4,75	4,090909
4	4	3,818182
4	4,5	3,636364
3,75	3,5	3,636364
3,75	4,75	3,909091
3,25	4	3,090909

(Sumber : Alfi Novia Zahrotul Hidayah, Anief Fauzan Rozi ; 2021)

Tabel II.3. Perhitungan rata-rata keterangan

Keterangan
3,340909091
4,363636364
4,363636364
3,939393939
4,106060606
4,045454545
3,628787879
4,136363636

(Sumber : Alfi Novia Zahrotul Hidayah, Anief Fauzan Rozi ; 2021)

Data integrasi dilakukan pada atribut-atribut yang mengidentifikasi identitas baik seperti pada tabel berikut :

Tabel II.3. Tabel data Integrasi

Nama	A1-A4	B1-B4	C1-C11	Keterangan
Dosen 1	3	3,75	3,272727	3,340909091
Dosen 2	4,25	4,75	4,090909	4,363636364
Dosen 3	4,25	4,75	4,090909	4,363636364
Dosen 4	4	4	3,818182	3,939393939
Dosen 5	4,25	4,25	3,818182	4,106060606
Dosen 6	4	4,5	3,636364	4,045454545
Dosen 7	3,75	3,5	3,636364	3,628787879
Dosen 8	3,75	4,75	3,909091	4,136363636
Dosen 9	3,25	4	3,090909	3,446969697
Dosen 10	3	3,75	3,272727	3,340909091
Dosen 11	4,25	4,75	4,090909	4,363636364

(Sumber : Alfi Novia Zahrotul Hidayah, Anief Fauzan Rozi ; 2021)

Pada tahap ini akan dilakukan tahap mengolah data menjadi bentuk yang sesuai untuk proses data mining dimana peneliti melakukan perubahan nama atribut sehingga table menjadi seperti gambar dibawah :

Tabel II.4. Hasil transformasi

Nama	Kesiapan Dosen	Kedisiplinan	Penyampaian Soal	Keterangan
Dosen 1	Cukup Baik	Baik	Baik	Baik
Dosen 2	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik
Dosen 3	Sangat Baik	Sangat Baik	Sangat Baik	Sangat Baik
Dosen 4	Baik	Baik	Baik	Baik
Dosen 5	Sangat Baik	Sangat Baik	Baik	Sangat Baik
Dosen 6	Baik	Sangat Baik	Baik	Sangat Baik
Dosen 7	Baik	Cukup Baik	Baik	Baik
Dosen 8	Baik	Sangat Baik	Baik	Sangat Baik

(Sumber : Alfi Novia Zahrotul Hidayah, Anief Fauzan Rozi ; 2021)

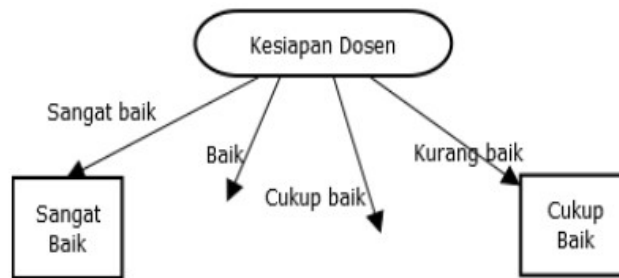
Data yang telah dikumpul, diseleksi dan ditransformasi akan dikelola menggunakan metode decision tree. Atribut data klasifikasi akan dianalisis dengan menggunakan software Rapid Miner. Dataset yang digunakan yaitu data kinerja karyawan (dosen) Universitas Mercu Buana Yogyakarta Periode Gasal 2016. Pembangunan pohon keputusan dilakukan setelah selesai menghitung nilai entropy total, nilai entropy dari tiap-tiap atribut, menghitung nilai information gain, menentukan information gain tertinggi hal ini dapat dilihat dari tabel dibawah ini.

Tabel II.5. Hasil perhitungan *Entrhopy* dan *Gain*

Node	Atribut	Nilai Atribut	Total	Sangat baik	Baik	Cukup baik	Entropy	Gain
TOTAL			1102	528	478	42	1,18876775	
	Kesiapan Dosen							0,508398822
		Sangat Baik	339	326	13	0	0	
		Baik	580	247	329	4	1,03796735	
		Cukup baik	160	9	126	25	0,92340994	
		Kurang baik	21	0	8	13	0	
	Kedisiplinan							0,45619394
		Sangat baik	626	504	120	2	0,73510518	
		Baik	409	78	322	9	0,84870536	
		Kurang baik	7	0	0	7	0	
		Buruk	2	0	1	1	0	
	Penyampaian soal							0,507354987
		Sangat baik	586	499	86	1	0,61942601	
		Baik	469	83	375	11	0,82714973	
		Cukup baik	46	0	17	29	0	
		Kurang baik	1	0	0	1	0	
		Buruk	0	0	0	0	0	

(Sumber : Alfi Novia Zahrotul Hidayah, Anief Fauzan Rozi ; 2021)

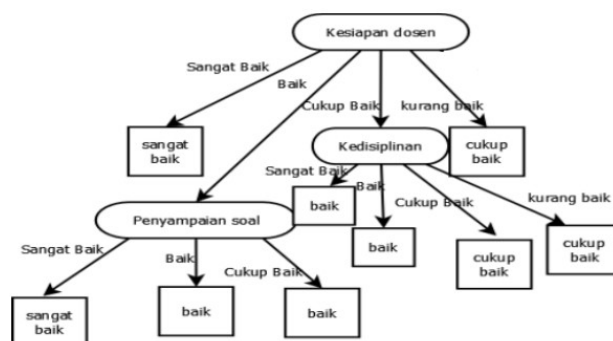
Data kemudian dikelompokkan berdasarkan atribut dan nilai atributnya untuk kemudian dihitung jumlah keseluruhan, kemudian dihitung nilai entropy dan gain untuk masing-masing atribut.



Gambar II.4. Hasil *Decision Tree Root Nude*

(Sumber : Alfi Novia Zahrotul Hidayah, Anief Fauzan Rozi ; 2021)

Dari tree tersebut maka akan dilakukan penghitungan pada kesiapan dosen baik dan cukup baik sebagai tree cabang pertama, Pada perhitungan tahap dua, didapatkan bahawa gain tertinggi terdapat pada penyampaian soal sehingga dilakukan perhitungan pada penyampaian soal. Selanjutnya akan dilakukan perhitungan tree node 1 kesiapan dosen cukup baik. Pada tahap ini akan dilakukan perhitungan tree root pertama kesiapan dosen baik. Dari perhitungan node 1 tersebut maka disimpulkan tidak ada perhitungan node selanjutnya dikarenakan tidak boleh ada atribut perulangan yang akan menghasilkan tree sebagai berikut :



Gambar II.5. Hasil Akhir *decision three*

(Sumber : Alfi Novia Zahrotul Hidayah, Anief Fauzan Rozi ; 2021)

Penelitian yang dilakukan oleh Fitria Rachmawati (2016) yang mengangkat judul Analisis Algoritma C4.5 Untuk Pengangkatan Karyawan Tetap Studi Kasus PT Citra Abadi Sejati Bogor. Analisa yang dilakukan menggunakan metode Algoritma C4.5 yang menghasilkan output berupa pohon keputusan untuk menentukan karyawan Tetap dan Kontrak dengan melihat dari JABATAN, GRADE, MASA KERJA, PENILAIAN KERJA, dan PENDIDIKAN. Contoh dari data karyawan dapat dilihat pada table 1:

Table II.6 : Contoh Data Karyawan

No	Nip	Grade	Nilai	Nama Pegawai	Jabatan	Bagian	Pendidikan	L/P	Tempat Lahir	Tanggal lahir	Tanggal Masuk	Mas Kerja
1	6376	B	RENDAH	Endang Saptaning	Operator	Sample	3	P	Wonogiri	01-Apr-65	17-Oct-05	11 Tahun
2	6304	B	RENDAH	Siti Sopiah	Operator Pattern	CAD Master	3	P	Bogor	12-Jul-85	26-Feb-05	11 Tahun
3	895	B	RENDAH	Sutarmi	Recorder	Sewing Staff B	3	P	Klaten	10-Jul-74	13-Oct-97	13 Tahun
4	2360	B	RENDAH	Nining S	Supervisor	Sewing Staff 2	3	P	Bogor	08-Mar-68	25-Mar-02	14 Tahun
5	1090	B	RENDAH	Evi Pristiwati	Q.C Audit	Quality Control 3	3	P	Surabaya	16-Nov-75	16-Mar-98	18 Tahun
6	538	B	RENDAH	Luki Wahyudin	Asistent Supervisor	Cutting 2	3	L	Bogor	17-Feb-75	25-Sep-95	21 Tahun
7	1743	B	RENDAH	Yanah	Asistent Supervisor	Quality Control 2	3	P	Bogor	25-Jun-68	04-May-00	16 Tahun
8	1863	B	RENDAH	Nursetiasih	Staff	Industrial Eng 2	3	P	Bogor	24-Feb-77	11-Oct-00	16 Tahun
9	1160	B	RENDAH	Zainul Hamidah	Asistent Supervisor	Cutting 2	3	P	Sragen	13-May-76	25-Apr-98	18 Tahun
10	1134	B	RENDAH	Hamsinar	Q.C Audit	Quality Control 3	3	P	B. Lampung	11-Jan-76	20-Apr-98	18 Tahun

(Sumber : Fitri Rachmawati ; 2016)

Dari data karyawan yang telah dikumpulkan, maka tahap berikutnya yaitu adalah dilakukan pengolahan data. Proses pengolahan data mining yang dilakukan

pada penelitian ini terdiri dari beberapa tahap yaitu, tahap data selection dan data transformation.

1. Data selection Data yang terkumpul pada table 1 terdapat beberapa atribut yang tidak terpakai dalam proses pengolahan data. Atribut yang tidak memiliki kontribusi terhadap tujuan penelitian, tidak digunakan dalam penelitian ini. Table 2 merupakan hasil dari proses data selection dapat dilihat pada table 2 :

Table II.7 Data Karyawan Setelah Tahap Data Selection

No	Grade	Nilai	Nama Pegawai	Jabatan	Bagian	Pendidikan	L/P	Mas Kerja
1	B	RENDAH	Endang Saptaning	Operator	Sample	3	P	11 Tahun
2	B	RENDAH	Siti Sopiah	Operator Pattern	CAD Master	3	P	11 Tahun
3	B	RENDAH	Sutarmi	Recorder	Sewing Staff B	3	P	19 Tahun
4	B	RENDAH	Nining S	Supervisor	Sewing Staff 2	1	P	14 Tahun
5	B	RENDAH	Evi Pristiwati	Q.C Audit	Quality Control 3	3	P	18 Tahun
6	B	RENDAH	Luki Wahyudin	Asistent Supervisor	Cutting 2	3	L	21 Tahun
7	B	RENDAH	Yanah	Asistent Supervisor	Quality Control 2	3	P	16 Tahun
8	B	RENDAH	Nursetiasih	Staff	Industrial Eng 2	3	P	16 Tahun
9	B	RENDAH	Zainul Hamidah	Asistent Supervisor	Cutting 2	3	P	18 Tahun
10	B	RENDAH	Hamsinar	Q.C Audit	Quality Control 3	3	P	18 Tahun

(Sumber : Fitria Rachmawati ; 2016)

2. Data Transformation Tahap selanjutnya yaitu melakukan proses transformasi dengan cara mengubah pengetahuan dari data real yang terdapat pada table 2. transformasi data ini dilakukan pada atribut Lama Kerja atau masa kerja dan

3. pada atribut Pendidikan. Mengelompokan lama kerja berdasarkan pada data karyawan seperti terlihat pada table 3 :

Table II.8 Klasifikasi Lama Kerja

Lama Kerja (Tahun)	Kriteria
3 Tahun	Masa Promosi
<3 Tahun	Baru

(Sumber : Fitria Rachmawati ; 2016)

Mengelompokan pendidikan berdasarkan pada data karyawan seperti terlihat pada table 4 :

Table II.9 Klasifikasi Pendidikan

Pendidikan	Kriteria
1	SD
2	SMP
3	SMA

(Sumber : Fitria Rachmawati ; 2016)

Hasil dari proses data transformation dapat dilihat pada table 5 :

Table II.10 Data Karyawan Hasil Transformation

No	Grade	Nilai	Nama Pegawai	Jabatan	Pendidikan	Masa Kerja
1	B	Tinggi	Mulyaman Iskandar	Operator	SMA	Masa Promosi
2	H	Rendah	Prisiyatun	Operator Jahit	SMP	Masa Promosi
3	H	Sedang	Elli Susanti	Operator Jahit	SMP	Masa Promosi
4	H	Tinggi	Erlinda	Operator	SMA	Baru
5	H	Rendah	Suryana	Operator Jahit	SMA	Baru
6	H	Rendah	Widha Hendrawan	Operator Jahit	SMP	Baru
7	H	Rendah	Eka Jayanti	Operator Jahit	SMP	Baru
8	K	Rendah	Sally Siregar	Operator	SMA	Baru
9	K	Rendah	Anih Hanapiah	Helper	SMA	Baru
10	K	Rendah	Reni Suherni	Operator Jahit	SMA	Baru
11	K	Rendah	Enah	Operator Jahit	SMP	Baru
12	K	Rendah	Maryani	Operator Jahit	SD	Baru
13	K	Rendah	Nengsih	Operator Jahit	SD	Baru
14	K	Rendah	Nuraini	Operator Jahit	SMA	Baru
15	K	Rendah	Siti Soleha	Operator Jahit	SMP	Baru
16	K	Rendah	Erni Avrianti	Operator Jahit	SD	Baru
17	K	Rendah	Ria Ratnawati	Operator Jahit	SMA	Baru
18	K	Rendah	Tuti Setiawati	Operator Jahit	SMP	Baru

(Sumber : Fitria Rachmawati ; 2016)

Analisis Pengangkatan Karyawan Tetap Menggunakan Algoritma C4.5

Berikut ini data yang digunakan dalam penelitian dapat dilihat pada table 6 :

Table II.11 Data Karyawan

No	Grade	Nilai	Nama Pegawai	Jabatan	Masa Kerja	Hasil
1	B	Tinggi	Mulyaman Iskandar	Operator	Masa Promosi	Di Terima
2	H	Rendah	Prisiyatun	Operator Jahit	Masa Promosi	Di Tolak
3	H	Sedang	Elli Susanti	Operator Jahit	Masa Promosi	Di Terima
4	H	Tinggi	Erlinda	Operator	Baru	Di Terima
5	H	Rendah	Suryana	Operator Jahit	Baru	Di Tolak
6	H	Rendah	Widha Hendrawan	Operator Jahit	Baru	Di Tolak
7	H	Rendah	Eka Jayanti	Operator Jahit	Baru	Di Tolak
8	K	Rendah	Sally Siregar	Operator	Baru	Di Tolak
9	K	Rendah	Anih Hanapijah	Helper	Baru	Di Tolak
10	K	Rendah	Reni Suherni	Operator Jahit	Baru	Di Tolak
11	K	Rendah	Enah	Operator Jahit	Baru	Di Tolak
12	K	Rendah	Maryani	Operator Jahit	Baru	Di Tolak
13	K	Rendah	Nengsih	Operator Jahit	Baru	Di Tolak
14	K	Rendah	Nuraini	Operator Jahit	Baru	Di Tolak
15	K	Rendah	Siti Soleha	Operator Jahit	Baru	Di Tolak
16	K	Rendah	Erni Avrianti	Operator Jahit	Baru	Di Tolak
17	K	Rendah	Ria Ratnawati	Operator Jahit	Baru	Di Tolak
18	K	Rendah	Tuti Setiawati	Operator Jahit	Baru	Di Tolak

(Sumber : Fitria Rachmawati ; 2016)

Dari data yang tertera pada table 4.12 akan di buat pohon keputusan untuk menentukan karyawan Tetap dan Kontrak dengan melihat dari JABATAN, GRADE, MASA ERJA, PENILAIAN KINERJA, dan PENDIDIKAN.

Secara umum Algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut.

- a. Pilih atribut sebagai akar.
- b. Buat cabang untuk tiap tiap nilai.
- c. Bagi kasus dalam cabang.
- d. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut atribut yang ada. Untuk menghitung gain digunakan rumus :

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

Keterangan :

S : himpunan Kasus

A : fitur

n : jumlah partisi S

Pi : proporsi dari Si terhadap S

Entropy merupakan pengukuran ketidakpastian rata-rata kumpulan data ketika kita tidak tahu hasil dari sumber informasi yang kita tidak punya. Ini juga menunjukkan jumlah rata-rata informasi yang akan diterima dari hasil sumber informasi.

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

Keterangan :

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

| S_i | : jumlah kasus pada partisi ke-i |

S | : jumlah kasus dalam S

Hasil Pengolahan Data Dengan Menggunakan Tools RapidMiner

Pengujian terhadap analisa, sangat penting dilakukan untuk menentukan dan memastikan apakah hasil analisa tersebut telah sesuai dengan keputusan yang diharapkan. Untuk menguji kebenaran dari hasil pengolahan data yang dilakukan secara manual, maka dapat menggunakan salah satu software aplikasi RapidMiner 5.3.0.

RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi.

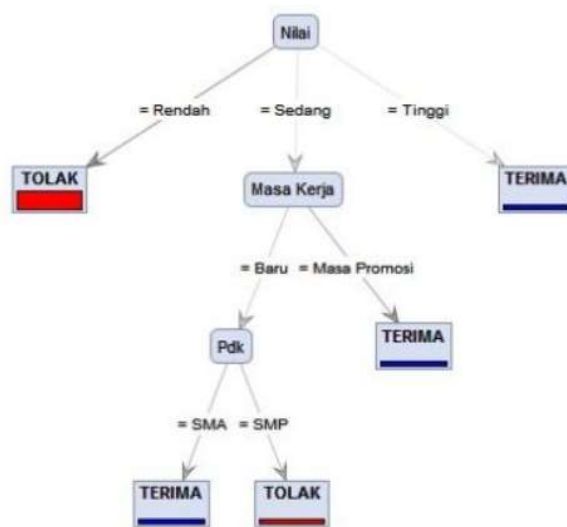
Untuk memudahkan dalam menggunakan RapidMiner untuk membuat decision tree, kita gunakan data sederhana. Pertama-tama data pada table dibuat lagi dalam format excel seperti yang terlihat pada Tabel 7 :

Table II.12 Table Keputusan Dalam Format XLS

No	Grade	Nilai	Nama Pegawai	Jabatan	Pendidikan	Masa Kerja	Hasil
1	B	Tinggi	Mulyaman Iskandar	Operator	SMA	Masa Promosi	Di Terima
2	H	Rendah	Prisiyatun	Operator Jahit	SMP	Masa Promosi	Di Tolak
3	H	Sedang	Elli Susanti	Operator Jahit	SMP	Masa Promosi	Di Terima
4	H	Tinggi	Erlinda	Operator	SMA	Baru	Di Terima
5	H	Rendah	Suryana	Operator Jahit	SMA	Baru	Di Tolak
6	H	Rendah	Widha Hendrawan	Operator Jahit	SMP	Baru	Di Tolak
7	H	Rendah	Eka Jayanti	Operator Jahit	SMP	Baru	Di Tolak
8	K	Rendah	Sally Siregar	Operator	SMA	Baru	Di Tolak
9	K	Rendah	Anih Hanapiah	Helper	SMA	Baru	Di Tolak
10	K	Rendah	Reni Suherni	Operator Jahit	SMA	Baru	Di Tolak
11	K	Rendah	Enah	Operator Jahit	SMP	Baru	Di Tolak
12	K	Rendah	Maryani	Operator Jahit	SD	Baru	Di Tolak
13	K	Rendah	Nengsih	Operator Jahit	SD	Baru	Di Tolak
14	K	Rendah	Nuraini	Operator Jahit	SMA	Baru	Di Tolak
15	K	Rendah	Siti Soleha	Operator Jahit	SMP	Baru	Di Tolak
16	K	Rendah	Erni Avrianti	Operator Jahit	SD	Baru	Di Tolak
17	K	Rendah	Ria Ratnawati	Operator Jahit	SMA	Baru	Di Tolak
18	K	Rendah	Tuti Setiawati	Operator Jahit	No	Baru	Di Tolak

(Sumber : Fitria Rachmawati ; 2016)

Setelah data yang kita punya dibuat dalam bentuk tabel format xls, Selanjutnya lakukan Importing Data kemudian tarik atau drag pada main process kemudian klik Import Configuration Wizard untuk import data.



Gambar II.6 : Hasil Graphic View Decision Tree

(Sumber : Fitria Racahmawati ; 2016)

Selain menampilkan hasil decision tree berupa graph atau tampilan pohon keputusan, RapidMiner juga menyediakan tool untuk menampilkan hasil berupa teks view dengan mengklik button Text View seperti yang tampak pada Gambar 2 :


```

Tree

Nilai = Rendah: TOLAK {TERIMA=0, TOLAK=566}
Nilai = Sedang
| Masa Kerja = Baru
| | Pdk = SMA: TERIMA {TERIMA=2, TOLAK=0}
| | Pdk = SMP: TOLAK {TERIMA=0, TOLAK=2}
| Masa Kerja = Masa Promosi: TERIMA {TERIMA=11, TOLAK=0}
Nilai = Tinggi: TERIMA {TERIMA=12, TOLAK=0}

```

Gambar II.7 : Hasil Text View Decision Tree

Dari hasil penelitian perhitungan Algoritma C4.5 dengan menggunakan Tools RapidMiner terdapat perbedaan sedikit hasil, yaitu:

a. Pada perhitungan Algoritma C4.5 manual atau dengan cara menghitung dengan rumus dapat mengetahui secara detail atribut mana yang di jadikan node awal yang paling penting dalam mengambil keputusan pengangkatan karyawan kontrak menjadi tetap dengan mengetahui gain yang tertinggi, dan dapat mengetahui ada beberapa karyawan yang semula kontrak berdasarkan penilaian menjadi karyawan tetap sesuai kebutuhan yang ada pada perusahaan tersebut.

b. Pada perhitungan menggunakan tools RapidMiner dari hasil penelitian tidak berbeda jauh hanya memunculkan hasil view Graphic dari data keryawan yang dimasukan memakai format xls, juga tidak dapat mengetahui information gain secara detail.

II.2.5. Kinerja Karyawan

Kinerja karyawan adalah kuantitas dan / atau kualitas pekerjaan yang dihasilkan oleh karyawan PT Indoland Batam, seperti standar kerja, dalam kurun waktu tertentu, berdasarkan peraturan yang ditetapkan atau berlaku, prosedur operasi, standar, dan fungsi di dalam perusahaan. Indikator yang digunakan untuk mengukur kinerja karyawan adalah:

1. Jumlah pekerjaan yang di hasilkan
2. Kualitas pekerjaan yang di hasilkan
3. Ketepatan waktu menyelesaikan pekerjaan
4. Kehadiran pada hari dan jam kerja
5. Kemampuan kerja sama

II.2.6. PHP (*Hypertext Preprocessor*)

PHP merupakan script untuk pemrograman *script web server-side*, *script* yang membuat dokumen HTML secara *on the fly*, maksudnya dokumen HTML yang dihasilkan dari suatu aplikasi bukan dokumen HTML yang dibuat dengan menggunakan editor teks atau editor HTML.

PHP adalah secara umum dikenal dengan sebagai bahasa pemrograman *script-script* yang membuat dokumen HTML secara *on the fly* yang dieksekusi di server web, dokumen HTML yang dihasilkan dari suatu aplikasi bukan dokumen HTML yang dibuat dengan menggunakan editor teks atau editor HTML, dikenal juga sebagai bahasa pemrograman *server side*. (Vivi : 2021).

</html>

<head><title>Contoh</title></head>

```

<body>

<?php echo "Tulisan ini dibuat dengan script PHP";?>

</body>

</html>

```

Kode/script PHP diapit dengan menggunakan tag awal dan tag akhir yang khusus (tag awal), yang memungkinkan pemrogram untuk masuk dan keluar dari mode script PHP. Hal yang dapat dikerjakan PHP secara mendasar antara lain mendapatkan data dari form, menghasilkan isi halaman web yang dinamik, dan menerima *cookies*, namun kemampuan (*Feature*) PHP yang paling diandalkan dan signifikan adalah dukungan kepada banyak *database*.

II.2.7. MySQL

Database secara sederhana, dapat kita sebut sebagai gudang data. secara teori, *database* adalah kumpulan data atau informasi yang kompleks, data-data tersebut disusun menjadi beberapa kelompok dengan tipe data yang sejenis disebut *table/entity*), di mana setiap datanya dapat saling berhubungan satu sama lain atau dapat berdiri sendiri, sehingga mudah diakses. MySQL merupakan *database* yang awalnya hanya berjalan pada sistem Unix dan Linux. Seiring berjalannya waktu dan banyaknya peminat yang menggunakan *database* ini, MySQL merilis versi yang dapat diinstal pada hampir semua platform, termasuk Windows. Lisensi dari MySQL adalah *freeware*. SQL merupakan kependekan dari kata "*Structured Query Language*". SQL merupakan suatu bahasa permintaan yang terstruktur yang

melekat pada satu *database* atau SMBD tertentu, sedangkan MySQL merupakan *databasenya*

Fungsi MySQL akan sering digunakan saat membuat website dinamis yang membutuhkan *database* MySQL, Misalnya untuk membuat query, manipulasi table, menghubungkan ke *Server* MySQL, menyimpan data, dan lain-lain. Semua contoh fungsi MySQL di bawah ini bisa bekerja dengan baik, jika sudah terhubung dengan *Server* MySQL. Adapun argumen fungsi `mysql_connect`-nya adalah :

1. Nama *Server* MySQL adalah „localhost“
2. Nama *User* MySQL adalah „mysql_user“
3. Password MySQL adalah „mysql_password“

Seiring berjalannya waktu dan banyaknya peminat yang menggunakan *database* ini, MySQL merilis versi yang dapat diinstal pada hampir semua platform, termasuk Windows. Lisensi dari MySQL adalah *freeware*. Dengan kata lain, MySQL merupakan SMBD-nya dan SQL adalah perintah atau bahasa yang melekat di dalam SMBD tersebut. Struktur dasar dari ekspresi SQL terdiri dari tiga klausa, yaitu : *select*, *from* dan *where*. (Muhammad : 2021).

II.2.8. Basis Data

Basis data adalah kumpulan data yang saling berhubungan secara logis dan didesain untuk mendapatkan data yang dibutuhkan oleh suatu organisasi. Basis Data merupakan data yang terintegrasi, yang diorganisasi untuk memenuhi kebutuhan para pemakai di dalam suatu organisasi. Berdasarkan diatas,

berintegrasi sehingga dapat memenuhi kebutuhan user dalam perusahaan atau organisasi.

Sistem basis data merupakan basis data dengan para pemakai yang menggunakan basis data secara bersama-sama, personil yang merancang dan mengelola basis data, teknik-teknik untuk merancang dan mengelola basis data, serta sistem computer yang mendukungnya. Sistem basis data adalah suatu sistem menyusun dan mengelola *record-record* menggunakan komputer untuk menyimpan atau merekam serta memelihara data operasional lengkap sebuah organisasi/perusahaan sehingga mampu menyediakan informasi yang optimal yang diperlukan pemakai untuk proses mengambil keputusan. Berdasarkan diatas, dapat disimpulkan bahwa sistem basis data merupakan penggunaan basis data yang dimana sistem dapat menyusun dan mengola record secara komputerisasi yang dapat memudahkan sebuah perusahaan/organisasi untuk proses pengambilan keputusan. (Alvin dan Catur : 2020).

II.2.9. Web

Web adalah sebagai kumpulan halaman-halaman yang di gunakan untuk menampilkan informasi teks, gambar diam atau gerak, animasi, suara dan atau gabungan dari semuanya baik yang bersifat statis maupun dinamis yang membentuk satu rangkaian bangunan yang saling terkait dimana masing-masing dihubungkan dengan jaringanjaringan halaman (Hyperlink). Pengertian web yaitu sebuah dokumen yang ditulis dalam hype text markup language (HTML) yang dapat diakses melalui protocol hype text tranfer protocol (HTTP) yang merupakan

protokol untuk menyampaikan informasi dari sebuah pusat situs web untuk ditampilkan dihadapan pengguna program pembaca informasi yang ada pada situs web. (Syahril Hasan & Nurlaila Muhammad : 2020).

II.2.10. Unified Modeling Language (UML)

Unified Modeling Language (UML) adalah bahasa spesifikasi standar yang dipergunakan untuk mendokumentasikan, menspesifikasikan dan membangun perangkat lunak. UML merupakan metodologi dalam mengembangkan sistem berorientasi objek dan juga merupakan alat untuk mendukung pengembangan sistem. UML saat ini sangat banyak dipergunakan dalam dunia industri yang merupakan standar bahasa pemodelan umum dalam industri perangkat lunak.

UML yaitu satu kumpulan konvensi pemodelan yang digunakan untuk menentukan atau menggambarkan sebuah sistem perangkat lunak yang terkait dengan objek. UML merupakan suatu kumpulan teknik terbaik yang telah terbukti sukses dalam memodelkan system yang besar dan kompleks. UML tidak hanya digunakan dalam proses pemodelan perangkat lunak, namun hampir dalam semua bidang yang membutuhkan pemodelan.

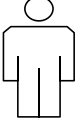
II.2.10.1. Use Case Diagram

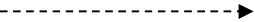
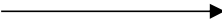
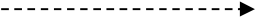
Use Case Diagram merupakan pemodelan untuk sistem informasi yang akan dibangun. *Use case* mendeskripsikan sebuah interaksi antara satu atau lebih aktor dengan sistem informasi yang akan dibangun. *Use case* digunakan untuk mengetahui fungsi apa saja yang ada pada sebuah sistem informasi dan siapa saja

yang berhak menggunakan fungsi-fungsi tersebut.

Adapun simbol *use case diagram* terdapat pada Tabel.II.1.

Tabel II.13. Simbol Use Case Diagram

Simbol	Deskripsi
<i>Use Case</i> 	<i>Use case</i> menggambarkan fungsionalitas yang disediakan sistem sebagai unit-unit yang saling bertukar pesan antar unit atau aktor: biasanya dinyatakan dengan menggunakan kata kerja diawal di awal frase nama use case.
Aktor/actor 	Aktor adalah orang, proses, atau sistem lain yang berinteraksi dengan sistem informasi yang akan dibuat di luar sistem informasi yang akan dibuat itu sendiri.

Asosiasi /association 	Komunikasi antara aktor dan use case yang berpartisipasi pada use case atau use case memiliki interaksi dengan actor.
Ekstensi / extend <<extends> 	Relasi use case tambahan ke sebuah use case dimana use case yang ditambahkan dapat berdiri sendiri walau tanpa use case tambahan itu; mirip dengan prinsip inheritance pada pemrograman berorientasi objek biasanya use case tambahan memiliki nama depan yang sama dengan use case yang ditambahkan misal Arah panah mengarah pada use case yang ditambahkan.
Generalisasi / generalization 	Hubungan generalisasi dan spesialisasi (umum – khusus) antara dua buah use case dimana fungsi yang satu adalah fungsi yang lebih umum dari lainnya.
Menggunakan / include 	Include berarti use case yang tambahan akan selalu melakukan pengecekan apakah use case yang ditambahkan telah dijalankan sebelum use case tambahan dijalankan

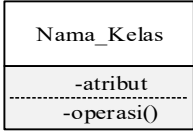
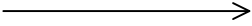
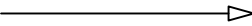
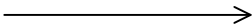
(Sumber : Sinta Maria, 2021)

II.2.10.2. *Class Diagram*

Class Diagram menggambarkan struktur sistem dari segi pendefinisian kelas-kelas yang akan dibuat untuk membangun sistem. Kelas memiliki apa yang

disebut atribut dan metode atau operasi. Adapun simbol *class diagram* terdapat pada Tabel.II.8.

Tabel II.14. Simbol Class Diagram

Simbol	Deskripsi
Kelas 	Kelas pada struktur sistem.
Antar muka/interface 	Sama dengan konsep interface dalam pemrograman berorientasi objek.
Asosiasi / association	Relasi antar kelas dengan makna umum, asosiasi biasanya juga disertai dengan multiplicity.
Asosiasi berarah / directed association 	Relasi antar kelas dengan makna kelas yang satu digunakan oleh kelas yang lain, asosiasi biasanya juga disertai dengan multiplicity.
Generalisasi 	Relasi antar kelas dengan makna generalisasi-generalisasi- spesialisasi (umum khusus).
Kebergantungan / dependency 	Relasi antar kelas dengan makna kebergantungan antar kelas
Agregasi /aggregation 	Relasi antar kelas dengan makna semua-bagian (whole- part).

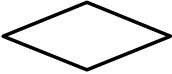
(Sumber : Sinta Maria, 2021)

II.2.10.3. Activity Diagram

Activity Diagram menggambarkan *workflow* (aliran kerja) atau aktivitas dari sebuah sistem atau proses bisnis atau menu yang ada pada perangkat lunak.

Penekanan pada diagram aktivitas adalah menggambarkan aktivitas sistem atau aktivitas yang dapat dilakukan oleh sistem, bukan apa yang dilakukan actor. Adapun simbol *activity diagram* terdapat pada Tabel.II.9.

Tabel II.15. Simbol *Activity Diagram*

Simbol	Diskripsi
Status awal 	Status awal aktivitas sistem, sebuah diagram aktivitas memiliki sebuah status awal.
Aktivitas 	Aktivitas yang dilakukan sistem, aktivitas biasanya diawali dengan kata kerja.
Percabangan / decision 	Asosiasi percabangan dimana jika ada pilihan aktivitas lebih dari Satu
Penggabungan / Join 	Asosiasi penggabungan dimana lebih dari satu aktivitas digabungkan menjadi satu..
Status akhir 	Status akhir yang dilakukan sistem, sebuah diagram aktivitas memiliki sebuah status akhir.

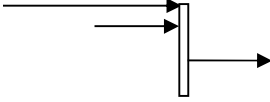
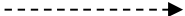
(Sumber : Sinta Maria, 2021)

II.2.10.4. *Sequence Diagram*

Sequence Diagram menggambarkan kelakuan objek pada use case dengan mendeskripsikan waktu hidup objek dan message yang dikirimkan dan diterima

antar objek. Adapun simbol *sequence diagram* terdapat pada Tabel.II.10.

Tabel II.16. Simbol *Sequence Diagram*

Simbol	Deskripsi
Aktor 	Orang, proses, atau sistem lain yang berinteraksi dengan sistem informasi yang akan dibuat di luar sistem informasi yang akan dibuat itu sendiri.
Garis hidup / lifeline 	Menyatakan kehidupan suatu objek.
Waktu aktif 	Menyatakan objek dalam keadaan aktif dan berinteraksi, semua yang terhubung dengan waktu aktif ini adalah sebuah tahapan yang dilakukan di dalamnya
Pesan tipe create << create >> 	Menyatakan suatu objek membuat objek yang lain, arah panah mengarah pada objek yang dibuat. arah panah mengarah pada objek yang memiliki operasi/metode, karena ini memanggil operasi/metode maka operasi/metode yang dipanggil harus ada pada diagram kelas sesuai dengan kelas objek yang berinteraksi.
Pesan tipe send 1 : masukan 	Menyatakan bahwa suatu objek mengirimkan data/masukan/informasi ke objek lainnya, arah panah mengarah pada objek yang dikirim.
Pesan tipe return 1 : keluaran 	Menyatakan bahwa suatu objek yang telah menjalankan suatu operasi atau metode menghasilkan suatu kembalian ke objek tertentu, arah panah mengarah pada objek yang menerima kembalian.

(Sumber : Sinta Maria, 2021)



BAB III

ANALISA DAN DESAIN SISTEM

BAB III

ANALISIS DAN DESAIN SISTEM

III.1 Analisis Masalah

Sistem penilaian kinerja karyawan yang berjalan pada PT. Raharja Sinergi Komunikasi yaitu penilaian kinerja ini dilakukan dengan tujuan untuk promosi jabatan yang kosong. Dengan adanya promosi jabatan karyawan akan bekerja dengan baik dan bertanggung jawab. Permasalahan yang dihadapi dalam penilaian kinerja karyawan adalah perusahaan kesulitan menentukan karyawan yang akan dipromosikan karena belum adanya proses penilaian kinerja yang dilakukan secara pasti dalam hal ini hanya dilakukan secara sepihak. Kualitas kerja karyawan juga seringkali berubah-ubah. Ini sangat berpengaruh pada prospek perusahaan kedepannya. Kesibukan seorang manager membuat tatap muka dengan karyawan kurang. Tentu saja ini membuat pengawasan pada kinerja karyawan sulit dilakukan.

Penyelesaian masalah penilaian kinerja pada PT. Raharja Sinergi Komunikasi yaitu dengan membuat sistem penilaian kinerja pendekatan data mining. Sistem ini menggunakan algoritma C4.5 yang menggunakan teknik data mining untuk membuat pohon keputusan, algoritma ini dimulai dengan memasukkan data training ke dalam simpul akar pada pohon keputusan. Algoritma C4.5 adalah algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi atau pengelompokan dan bersifat prediktif dan digunakan untuk membentuk pohon keputusan (*Decision Tree*).