EKSPLORASI DAN KLASIFIKASI ATRISI KARYAWAN MENGGUNAKAN METODE *DECISION TREE* DENGAN PENERAPAN ALGORITMA C4.5

(Skripsi)

Oleh

RISMA NURUL HIDAYATI

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023

ABSTRACT

EXPLORATION AND CLASSIFICATION OF EMPLOYEE ATTRITION USING THE DECISION TREE METHOD WITH THE APPLICATION OF THE C4.5 ALGORITHM

By

Risma Nurul Hidayati

Employee attrition is a significant problem for companies because it can have a negative impact on productivity and company sustainability. To find the information contained in the employee attrition data statistically and visualize one of them using the Exploratory Data Analysis (EDA) method. EDA provides an initial description of the data used, and can assist in the classification stages of data mining. The decision tree method is a data mining technique with the aim of analyzing and modeling the relationship between variables that influence a decision, in this case the problem of employee attrition. One of the algorithms that is often used in the decision tree method is the C4.5 algorithm. One of the advantages of the C4.5 algorithm is that it can handle categorical and numeric data types and uses the gain ratio as the root determination of the model. In this study, the main factor influencing employee attrition decisions is performance ratings. The application of the C4.5 algorithm to the available employee attrition data obtains an accuracy of 88%, a recall of 99% and a precision of 88%, which means that the decision tree method is quite good at classifying data.

Keywords: Employee Attrition, Exploratory Data Analysis, Decision Tree, Algorithm C4.5.

ABSTRAK

EKSPLORASI DAN KLASIFIKASI ATRISI KARYAWAN MENGGUNAKAN METODE *DECISION TREE* DENGAN PENERAPAN ALGORITMA C4.5

Oleh

Risma Nurul Hidayati

Atrisi karyawan menjadi masalah yang signifikan bagi perusahaan karena dapat berdampak negatif pada produktivitas dan keberlanjutan perusahaan. Untuk menemukan informasi yang terkandung pada data atrisi karyawan secara statistik dan visualisasi salah satunya menggunakan metode *Exploratory Data Analysis* (EDA). EDA memberikan gambaran awal tentang data yang digunakan, serta dapat membantu dalam tahapan klasifikasi data mining. Metode *decision tree* merupakan salah satu teknik data mining dengan tujuan menganalisis dan memodelkan hubungan antara variabel yang mempengaruhi sebuah keputusan dalam hal ini masalah atrisi karyawan. Salah satu algoritma yang sering digunakan pada *metode decision tree* adalah algoritma C4.5. Kelebihan algoritma C4.5 salah satunya dapat menangani data bertipe kategorik dan numerik serta menggunakan gain ratio sebagai penentuan akar pada model. Pada penelitian ini, faktor utama yang mempengaruhi keputusan atrisi karyawan adalah peringkat kinerja. Penerapan algoritma C4.5 pada data atrisi karyawan yang tersedia mendapatkan akurasi sebesar 88%, recall sebesar 99% dan precission sebesar 88% yang artinya metode *decision tree* cukup baik dalam mengklasifikasi data.

Kata Kunci: Atrisi Karyawan, *Exploratory Data Analysi*, *Decision Tree*, Algoritma C4.5.

EKSPLORASI DAN KLASIFIKASI ATRISI KARYAWAN MENGGUNAKAN METODE *DECISION TREE* DENGAN PENERAPAN ALGORITMA C4.5

Oleh

RISMA NURUL HIDAYATI

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023

Judul	: **EKSPLORASI DAN KLASIFIKASI ATRISI KARYAWAN MENGGUNAKAN METODE *DECISION TREE* DENGAN PENERAPAN ALGORITMA C4.5**
Nama	: Risma Nurul Hidayati
Nomor Pokok Mahasiswa	: **1917031027**
Jurusan	: **Matematika**
Fakultas	: **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 19740316 200501 1 001

Dr. Ahmad Faisol, S.Si., M.Sc.
NIP. 19800206 200312 1 003

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

   Ketua : **Dr. Aang Nuryaman, S.Si., M.Si.**

   Sekretaris : **Dr. Ahmad Faisol, S.Si., M.Sc.**

   Penguji
   Bukan Pembimbing : **Drs. Nusyirwan, M.Si.**

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Heri Satria, S.Si., M.Si.
NIP. 19711001 200501 1 002

Tanggal Lulus Ujian Skripsi : **11 Juli 2023**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama Mahasiswa	: **Risma Nurul Hidayati**
Nomor Pokok Mahasiswa	: **1917031027**
Jurusan	: **Matematika**
Judul Skripsi	: **EKSPLORASI DAN KLASIFIKASI ATRISI KARYAWAN MENGGUNAKAN METODE *DECISION TREE* DENGAN PENERAPAN ALGORITMA C4.5**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila kemudian hari terbukti bahwa skripsi ini merupakan salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 11 Juli 2023
Penulis,

Risma Nurul Hidayati
NPM. 1917031027

RIWAYAT HIDUP

Penulis bernama lengkap Risma Nurul Hidayati lahir di Klaten pada 29 September 2000. Penulis merupakan anak kedua dari dua bersaudara dari pasangan Bapak Eko Sugiyoto dan Ibu Sri Daryanti.

Penulis mengawali pendidikan di Taman Kanak-kanak Citra Insani Kecamatan Rawajitu Timur pada tahun 2005 sampai dengan 2007, SDS Citra Insani Kecamatan Rawajitu Timur pada tahun 2007 sampai dengan 2013. Kemudian melanjutkan ke Sekolah Menengah Pertama di SMP Negeri 1 Banjar Agung pada tahun 2013 sampai dengan 2016. Selanjutnya ke Sekolah Menengah Atas di SMA Negeri 1 Banjar Agung pada tahun 2016 sampai dengan 2019. Pada tahun 2019 penulis terdaftar sebagai mahasiswa Program Studi S1 Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN).

Pada tahun 2022 penulis melaksanakan Kerja Praktik (KP) di Dinas Komunikasi, Informatika, dan Statistik Provinsi Lampung sebagai aplikasi bidang ilmu di dunia kerja. Pada tahun yang sama, penulis melaksanakan Kuliah Kerja Nyata (KKN) di Sumber Makmur Kecamatan Banjar Margo, Kabupaten Tulang Bawang, sebagai bentuk pengabdian kepada masyarakat. Penulis mengikuti program Merdeka Belajar Kampus Merdeka (MBKM) yaitu Study Independen Bersertifikat di PT. Mitra Talenta Group mengenai *Big Data dan Business Intelligence.*

KATA INSPIRASI

"... Dan Allah beserta orang-orang yang sabar."
(Q.S Al-Baqarah: 249)

"Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya."
(Q.S Al-Baqarah: 286)

"Sesungguhnya bersama kesulitan ada kemudahan."
(Q.S Al-Insyirah: 6)

"Tingkah laku yang baik adalah ketika seseorang tidak membutuhkan imbalan sebagai ganti atas perbuatan baik."
(Abu Bakar Ash-Shiddiq)

"Many of life's failures are people who did not realize how close they were to success when they gave up."
(Thomas Edison)

"Jangan pernah menyerah hanya karena pernah gagal."
(Risma Nurul Hidayati)

PERSEMBAHAN

Alhamdulillahi robbil ‘alamin, dengan mengucap syukur kepada Allah SWT karena atas karunia rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini.

Teriring doa, rasa syukur, dan segala kerendahan hati. Dengan segala cinta dan kasih sayang kupersemhkan karya ini untuk orang-orang yang sangat berharga dalam hidupku:

Ayahku (Eko Sugiyoto) dan Ibuku (Sri Daryanti)
Yang telah mendidik, membesarkanku dan senantiasa mencintaiku dan menyayangiku dengan penuh kasih sayang, terimakasih atas segala usaha, nasihat, dukungan dan selalu mendoakan agar aku menjadi orang yang sukses, mengorbankan segalanya untuk kebahagiaanku dan cita-citaku, kalian merupakan motivasi terbesarku dan aku berjanji akan membahagiakan kalian. Semoga Allah SWT meridhai saya untuk dapat memberikan yang terbaik kepada Ayah, Ibu dan Allah SWT mengganti semuanya dengan Syuga-Nya kelak.
Aamiin Ya Rabbal Alamin.

Kakakku (Husnul Khotimah dan Anas Rosyid Dian)
Untuk kakakku yang selalu mendukung, mendoakan dan memberikan arahan serta semangat dalam menghadapi masalah dan berusaha membahagiakan kedua orang tua. Semoga kita selalu dalam lindungan Allah SWT.
Aamiin Ya Rabbal Alamin

Dosen Pembimbing dan Penguji
Yang telah memberikan ilmu yang bermanfaat, membimbingku tanpa lelah, nasihat-nasihat yang berharga dan kasih sayang yang tulus yang diberikan padaku hingga aku dapat memiliki kesempatan untuk memperoleh ilmu yang sangat berharga selama menempuh pendidkan ini.

Almamater tercinta, Universitas lampung

SANWACANA

Segala puji dan syukur penulis ucapkan kepada Allah SWT atas segala nikmat dan karunia-Nya yang tak terhingga sehingga penulis dapat menyelesaikan skripsi yang berjudul "Eksplorasi dan Klasifikasi Atrisi Karyawan Menggunakan Metode *Decision Tree* dengan Penerapan Algoritma C4.5". Dalam penulisan skripsi ini tidak dapat terselesaikan tanpa adanya bimbingan, bantuan dan dukungan dari berbagai pihak.

Sehingga, dalam kesempatan ini penulis mengucapkan terimakasih kepada:

1. Bapak Dr. Aang Nuryaman, S.Si., M.Si. selaku dosen pembimbing satu serta Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung yang senantiasa membimbing, memberi masukan, saran serta mendukung penulis dalam menyelesaikan skripsi ini.
2. Bapak Dr. Ahmad Faisol, S.Si., M.Sc. selaku dosen pembimbing dua yang memberikan bimbingan, pengarahan, serta saran sehingga penulis dapat mampu menyelesaikan skripsi ini.
3. Bapak Drs. Nusyirwan, M.Si. selaku dosen penguji yang telah memberikan kritik dan saran yang membangun sehingga skripsi ini dapat terselesaikan.
4. Bapak Drs. Tiryono Ruby, M.Sc., Ph.D. selaku pembimbing akademik yang telah memberikan bimbingan dan arahan selama masa perkuliahan.
5. Bapak Dr. Eng. Heri Satria, S.Si., M.Si. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Seluruh dosen, staff, karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Bapak, Ibu serta keluarga tercinta yang selalu mendukung, menemani, memberikan motivasi, mendoakan, serta memberikan semangat sehingga menguatkan penulis dalam menjalani setiap proses meraih gelar sarjana.

8. Terima kasih untuk diri sendiri yang telah berjuang, berproses, dan bertahan hingga dapat menyelesaikan skripsi.
9. Untuk Elsa, Herlina, Novi, Wiranto, Anis, Rizke, Putri, Lina, Nevi, Deni yang selalu mendoakan, memberikan semangat, motivasi, pengertian, pencerahan, serta mendengarkan keluh kesah penulis.
10. Teman-teman Matematika 2019 serta Abang Yunda yang telah membantu serta memberikan semangat kepada penulis yang tidak dapat disebutkan satu persatu.
11. Orang-orang baik yang namanya tidak dapat saya sebutkan satu persatu yang telah menjadi bagian teman terbaik penulis yang selalu memberikan semangat dan menemani penulis dalam keadaan apapun serta telah memberikan pengalaman dan banyak cerita selama masa perkuliahan.

Penulis menyadari bahwa masih banyak kekurangan dalam penulisan skripsi ini. Oleh karena itu, penulis mengharapkan masukan serta saran untuk dijadikan pelajaran kedepannya.

Bandar Lampung, 11 Juli 2023
Penulis,

Risma Nurul Hidayati
NPM. 1917031027

DAFTAR ISI

Halaman

DAFTAR TABEL

Tabel Halaman

DAFTAR GAMBAR

Gambar Halaman

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Setiap perusahaan pasti menginginkan kesuksesan dan keuntungan. Untuk mendapatkan hal tersebut dengan maksimal dibutuhkan sumber tenaga kerja yang memiliki kualitas dan kompetensi yang menjanjikan. Permasalahan pada suatu perusahaan salah satunya tenaga kerja dari segi sumber daya manusia. Agar tujuan perusahaan dapat tercapai, perlu didukung dengan sumber daya manusia yang memenuhi syarat dan kriteria pada perusahaan. Sumber daya manusia merupakan faktor utama dalam perusahaan. Pengurangan karyawan adalah masalah yang dihadapi oleh banyak organisasi, khususnya karyawan yang unggul dan berpengalaman. Bagi karyawan adanya pengurangan karyawan menimbulkan rasa cemas, stress, dan rasa tidak nyaman terhadap lingkungan kerja.

Atrisi karyawan adalah pemberhentian tetap dengan pengurangan yang bertahap pada jumlah staf yang terjadi saat karyawan pensiun, mengundurkan diri atau tidak diganti. Menurut UU No.13 Tahun 2003 tentang Ketenagakerjaan, yang dimaksud dengan pemutusan hubungan kerja (PHK) adalah pengakhiran hubungan kerja karena suatu hal tertentu yang mengakibatkan berakhirnya hak dan kewajiban antara pekerja dan perusahaan (Maringan, 2015).

Exploratory Data Analysis (EDA) adalah salah satu upaya untuk menggali dan mendapatkan informasi lebih dari pengelolaan suatu data. Dengan eksplorasi data akan mempermudah dalam membaca suatu data dan memahami data, sehingga

data yang berukuran besar dapat disajikan dalam sebuah model yang lebih sederhana. *Exploratory Data Analysis* (EDA) dapat membantu perusahaan dalam menentukan faktor-faktor yang meyebabkan atrisi pada suatu perusahaan.

Penggunaan sistem untuk melakukan klasifikasi atrisi karyawan digunakan untuk menangani masalah ketepatan hasil akurasi, dengan menggunakan metode pendekatan yang cocok untuk klasifikasi menggunakan teknik data mining sehingga diharapkan dapat membantu perusahaan memutuskan atrisi. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengidentifikasi informasi yang bermanfaat (Mardi, 2017).

Klasifikasi merupakan salah satu teknik data mining. Klasifikasi bertujuan untuk mengelompokan data menjadi kelas tertentu berdasarkan nilai atribut yang berkaitan dengan objek yang diamati. Pada data mining dikenal beberapa metode untuk proses klasifikasi. Metode tersebut antara lain *Neural Network, Fuzzy, Support Vector Machine,* dan *Decision Tree*. *Decision Tree* merupakan algoritma yang mengambil keputusan berdasarkan aturan yang ditetapkan. Aturan-aturan ini dapat digambarkan sepeti pohon yang daunnya merupakan aturan dan cabangnya merupakan keputusan yang diambil pada algoritma ini. Algoritma C4.5 merupakan struktur pohon keputusan yang sering digunakan karena memiliki kelebihan daripada algoritma yang lain seperti menghasilkan pohon keputusan yang mudah diinterprestasikan serta efisien dalam menangani atribut bertipe kategorik dan numerik.

Banyak penelitian yang menyatakan penggunaan teknik data mining dengan algoritma C4.5 memiliki fungsi untuk memberikan prediksi dengan hasil yang akurat serta mudah dimengerti. Dilihat dari penelitian terdahulu yang dilakukan oleh Anam & Santoso (2018), mengenai klasifikasi penerima beasiswa menggunakan metode Algoritma C4.5 dan *naive bayes*. Hasil penelitian

menunjukkan tingkat akurasi model algoritma C4.5 sebesar 96.40% lebih baik dari tingkat akurasi model algoritma *naive bayes* sebesar 95.11%. Penelitian lain juga dilakukan oleh Wahyuningsih & Utari (2018) tentang perbandingan metode *K-Nearest Neighbor*, *naïve bayes* dan *Decision Tree* untuk prediksi kelayakan Pemberian Kredit. Hasil dari penelitian ini juga algoritma *Decision Tree* masih menunjukan tingkat akurasi tertinggi dilihat dengan hasil akurasi algoritma *Decision Tree* sebesar 92,21%, akurasi algoritma *naïve bayes* sebesar 81,83% dan akurasi algoritma *K-Nearest Neighbor* sebesar 81,82%. Maka dapat disimpulkan bahwa algoritma *Decision Tree* layak untuk digunakan karena memiliki tingkat akurasi tertinggi.

Sebelum diproses menggunakan metode *Decision Tree* dengan algoritma C4.5 terlebih dahulu mengeksplorasi dan menampilkan visualisasi data dengan melihat hasil dalam bentuk tabel maupun grafik yang digunakan dalam memaknai informasi yang terkandung pada data. Seperti yang telah dilakukan oleh Radhi, dkk. (2021), mengenai penjualan barang elektronik menggunakan metode *Exploratory Data Analysis* (EDA) dan metode visualisasi. Hasil penelitian menunjukan barang yang banyak terjual adalah baterai AAA dan barang yang paling sedikit terjual adalah LG Drye (mesin pengering pakaian), sedangkan barang yang paling mahal adalah Macbook Pro Laptop.

Penelitian mengenai atrisi karyawan telah dilakukan oleh beberapa penelitian seperti Setiawan (2020), mengenai atrisi karyawan menggunakan regresi logistik. Dimana pada penelitiannya diperoleh *accuracy* 75% dengan 73% *sensitivity* dan 75% *specificity.*

Berdasarkan hasil penelitian terdahulu, metode *Decision Tree* terbukti efektif untuk mengklasifikasi dengan tingkat akurasi terbaik. Dalam masalah ini penulis tertarik untuk mengeksplorasi dan mengklasifikasi atrisi karyawan menggunakan metode *Decision Tree* dengan penerapan algoritma C4.5 dikarenakan penelitian yang akan dilakukan belum pernah diteli oleh penelitian lain.

1.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengeksplorasi data dalam menganalisis informasi yang terdapat pada kasus atrisi karyawan.
2. Penerapan algoritma C4.5 dengan metode *Decision Tree* pada kasus atrisi karyawan.
3. Mengetahui tingkat akurasi pada kasus atrisi karyawan menggunakan metode *Decision Tree* dengan penerapan algoritma C4.5.

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Mengetahui gambaran umum mengenai atrisi karyawan
2. Menambah ilmu pengetahuan khususnya dalam menerapan metode Algoritma C4.5
3. Dapat dijadikan bahan evalusi bagi perusahaan dalam menetapkan atrisi karyawan

II. TINJAUAN PUSTAKA

2.1 Atrisi Karyawan

Karyawan dapat diartikan individu yang bekerja untuk suatu organisasi, perusahaan, atau institusi dengan tujuan untuk memberikan kontribusi dalam mencapai tujuan perusahaan tersebut. Menurut Subri (2002), karyawan merupakan setiap penduduk yang masuk ke dalam usia kerja (berusia di rentang 15 hingga 64 tahun), atau jumlah total seluruh penduduk yang ada pada sebuah negara yang memproduksi barang dan jasa jika ada permintaan akan tenaga yang mereka produksi, dan jika mereka mau berkecimpung dan berpartisipasi dalam aktivitas itu.

Pemutusan hubungan kerja (PHK) adalah suatu tindakan yang dilakukan oleh pihak pengusaha atau perusahaan untuk mengakhiri hubungan kerja dengan karyawan atau pekerja. Proses PHK harus dilakukan sesuai dengan peraturan perundang-undangan yang berlaku di negara setempat dan sering kali melibatkan pemberian pemberitahuan tertentu kepada karyawan, memberikan hak-hak karyawan, seperti kompensasi atau tunjangan terkait, dan mengikuti prosedur yang ditetapkan untuk memastikan perlakuan yang adil terhadap karyawan yang terkena PHK.

Atrisi merupakan sebuah fenomena yang terjadi dibeberapa perusahaan dalam mengurangi jumlah karyawan yang dimilikinya dalam jangka waktu tertentu, didalam keadaan ini, perusahaan tidak berusaha untuk mencari pengganti dari

posisi yang ditinggalkan. Istilah atrisi sering juga digunakan bersamaan dengan *turnover* (pergantian). Namun perbedaannya pada *turnover* berkaitan dengan pemutusan hubungan kerja atau posisi yang ditinggalkan diisi kembali oleh karyawan baru. Beberapa alasan yang menyebabkan terjadinya atrisi karyawan yaitu pengurangan tenaga kerja, kinerja yang buruk, pelanggaran etika, dll.

2.2 *Exploratory Data Analysis* (EDA)

Menurut Mujilahwati (2021), Visualisasi data yaitu menggambarkan data secara nyata baik dalam bentuk tabel, *bar chart, pie chart, line chart, scatter plot, heatmap* ataupun histogram. Untuk melakukan visualisasi data juga beragam cara, salah satunya memanfaatkan EDA.

Exploratory Data Analysis (EDA) merupakan suatu metode eksplorasi data yang digunakan untuk menjelajahi dan meringkas data pengamatan sebelum melibatkan model statistik formal atau teknik prediksi (Nissa, dkk., 2020). Teknik aritmatika statistik dasar yang sering digunakan dalam EDA yaitu statistik deskriptif. Tujuan EDA yaitu untuk mencari pola data. Hal ini berkaitan dengan konsep data mining yaitu untuk mengekplorasi pola dari suatu data.

Dengan menggunakan teknik aritmatika statistik dasar dan visualisasi data, EDA membantu para analis data untuk mengidentifikasi pola menarik, tren, atau hubungan dalam data yang dapat digunakan untuk proses pengambilan keputusan lebih lanjut. Hal inilah yang digunakan dalam memperkaya analisis data, membantu mengoptimalkan hasil klasifikasi dengan pendekatan data mining (Wahyuni, dkk., 2019).

2.3 *Machine Learning*

Menurut Restoningsih (2020), *Machine Learning* (ML) adalah cabang dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada pengembangan algoritma dan model statistik yang memungkinkan komputer untuk belajar dari data dan melakukan tugas tertentu tanpa harus secara eksplisit diprogram. Dalam konteks AI, *Machine Learning* adalah pendekatan yang digunakan untuk mengajarkan komputer untuk "belajar" dari pengalaman atau data yang diberikan, sehingga mereka dapat melakukan tugas-tugas yang cerdas atau memberikan prediksi berdasarkan pola yang teridentifikasi dalam data tersebut.

2.3.1. Bagian *Machine Learning*

Menurut Mitcel (1997), sistem pembelajaran mesin terdiri dari tiga bagian utama, yaitu:

1. Model. Model *machine learning* adalah representasi matematis dari suatu konsep atau masalah yang ingin dipecahkan. Contoh model populer termasuk *neural networks, decision trees, support vector machines,* dan *regression models.*
2. Algoritma. Algoritma adalah langkah-langkah yang digunakan untuk melatih dan mengoptimalkan model *machine learning*. Algoritma ini mendefinisikan bagaimana model akan belajar dari data, menemukan pola atau hubungan yang ada dalam data, dan membuat prediksi atau keputusan berdasarkan pola-pola tersebut.
3. Training. Training adalah proses di mana model *machine learning* "belajar" dari data yang diberikan. Dalam tahap ini, model melihat contoh-contoh data, menyesuaikan parameter-parameternya, dan mencoba meminimalkan kesalahan atau selisih antara hasil prediksi dan nilai yang sebenarnya.

2.3.2 Jenis *Machine Learning*

Banyak hal yang dipelajari pada *machine learning*, akan tetapi pada dasarnya ada 3 hal pokok yang dipelajari dalam *machine learning* (Mitcel, 1997):

1. *Supervised machine learning*

   Model dilatih menggunakan data yang telah dilabeli dengan target atau keluaran yang diinginkan. Tujuan dari model adalah untuk belajar memetakan masukan (input) ke keluaran yang benar. Contohnya klasifikasi gambar di mana model dilatih menggunakan gambar-gambar yang sudah diberi label dengan kategori yang tepat.
2. *Unsupervised machine learning*

   Model digunakan untuk menemukan pola atau struktur yang tersembunyi dalam data yang tidak memiliki label atau keluaran yang diinginkan. Contohnya klasterisasi data di mana model memisahkan data menjadi kelompok-kelompok berdasarkan kesamaan fitur.
3. *Reinforcement machine learning*

   Model belajar melalui interaksi dengan lingkungannya. Model ini menerima umpan balik dalam bentuk *reward* atau hukuman sebagai respon terhadap tindakan yang diambil. Tujuannya adalah untuk mengoptimalkan keputusan dan tindakan yang menghasilkan *reward* maksimal dalam konteks tertentu. Contohnya permainan komputer di mana model belajar untuk memenangkan permainan melalui pengalaman bermain dan mendapatkan *reward* atau hukuman berdasarkan hasil tindakan.

Machine learning telah banyak digunakan dalam berbagai bidang, seperti pengenalan wajah, pengenalan suara, penerjemahan otomatis, deteksi anomali, prediksi pasar, analisis risiko, dan masih banyak lagi

2.4 Data Mining

Menurut Larose (2005), Data mining adalah salah satu bidang yang berkembang pesat karena besarnya kebutuhan akan nilai tambah dari database dalam skala besar yang makin banyak terakumulasi sejalan dengan pertumbuhan teknologi informasi.

Sedangkan pengertian data mining menurut Bramer(2007) yaitu:

1. Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses merupakan data yang sangat besar.
3. Tujuan data mining adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

2.4.1. Proses KDD

Data mining merupakan salah satu langkah dari serangkaian proses *iterative* KDD (*Knowledge Discovery in Database*) (Kusrini & Emha, 2009). Menurut Fayyad (1996), tahapan-tahapan KDD sebagai berikut:

1. *Data Selection*

   Tahap ini melibatkan pemilihan data yang relevan dari sumber data yang tersedia. Data yang dipilih harus sesuai dengan tujuan analisis yang akan dilakukan. Pemilihan data yang baik akan mempengaruhi kualitas pengetahuan yang ditemukan pada tahap-tahap berikutnya.

2. *Pre-processing/ Cleaning*

   Tahap ini melibatkan pembersihan, integrasi, transformasi, dan reduksi dimensi data mentah yang terpilih. Pembersihan data melibatkan identifikasi dan penanganan nilai yang hilang, duplikat, atau tidak valid. Integrasi data mencakup penggabungan data dari berbagai sumber yang berbeda.

Transformasi data melibatkan konversi data ke format yang sesuai dan normalisasi. Reduksi dimensi data bertujuan mengurangi kompleksitas dengan memilih fitur-fitur yang paling relevan untuk analisis.

3. *Transformation*

   Tahap ini melibatkan transformasi data yang telah diproses menjadi bentuk yang lebih sesuai untuk proses analisis selanjutnya. Transformasi ini dapat meliputi pemilihan atribut, agregasi, pemfilteran, penggabungan, dan penyederhanaan data. Tujuan dari tahap ini adalah untuk mempersiapkan data agar sesuai dengan teknik atau algoritma data mining yang akan digunakan.

4. Data mining

   Tahap ini merupakan inti dari proses KDD. Pada tahap ini, teknik dan algoritma data mining diterapkan pada data yang telah diproses dan ditransformasi untuk mengidentifikasi pola, hubungan dan pengetahuan yang tersembunyi dalam data. Teknik data mining yang umum digunakan antara lain adalah clustering, klasifikasi, regresi, asosiasi, dan analisis urutan.

5. *Interpretation/ Evaluation*

   Tahap ini melibatkan evaluasi hasil dari tahap data mining. Evaluasi dapat dilakukan dengan menggunakan metrik yang relevan tergantung pada tujuan analisis. Evaluasi juga mencakup interpretasi dan validasi hasil untuk memastikan bahwa pengetahuan yang ditemukan berguna, valid, dan dapat dipahami oleh pemangku kepentingan. Jika hasil evaluasi tidak memenuhi harapan, tahap-tahap sebelumnya mungkin perlu diperbaiki atau diulang.

2.4.2. Peran Utama Data mining

Tugas data mining melibatkan berbagai teknik dan metode untuk menganalisis data guna mendapatkan wawasan berharga.

Berikut adalah peran utama dalam data mining:

A. *Description* (Deskripsi)

   Deskripsi dalam data mining berfokus pada pemahaman dan penjelasan tentang karakteristik data yang ada. Tujuannya adalah untuk mengidentifikasi pola, tren, dan statistik yang relevan dari data tersebut. Deskripsi juga dapat melibatkan visualisasi data dalam bentuk grafik atau diagram untuk memberikan pemahaman yang lebih baik tentang distribusi atau hubungan antar variabel.

B. *Estimation* (Estimasi)

   Estimasi dalam data mining berkaitan dengan penggunaan model statistik atau matematis untuk memperkirakan nilai yang tidak diketahui atau tidak terukur. Estimasi membantu dalam mengisi kekosongan data dan memperoleh informasi yang berguna dari data yang tersedia.

C. *Prediction* (Prediksi)

   Prediksi dalam data mining adalah proses memprediksi nilai atau kejadian di masa depan berdasarkan pola dan tren yang terlihat dari data historis. Ini melibatkan penggunaan teknik seperti regresi, analisis deret waktu, atau algoritma pembelajaran mesin untuk membangun model yang dapat memprediksi hasil yang mungkin.

D. *Classification* (Klasifikasi)

   Klasifikasi adalah tugas data mining yang melibatkan pengelompokan objek atau data ke dalam kelas atau kategori yang telah ditentukan sebelumnya. Metode klasifikasi mempelajari pola dari data yang diketahui dan membangun model untuk mengklasifikasikan data baru ke dalam kelas yang tepat.

E. *Clustering* (Pengklusteran)

   Pengklasteran dalam data mining adalah proses mengelompokkan objek atau data yang serupa ke dalam kelompok atau *cluster* yang berbeda. Tujuannya adalah untuk menemukan pola-pola alami dalam data yang tidak diketahui sebelumnya. Metode pengklasteran mencoba untuk meminimalkan perbedaan antara objek dalam *cluster* yang sama dan memaksimalkan perbedaan antara *cluster* yang berbeda.

2.5 Klasifikasi

Menurut Kusrini & Emha (2009), klasifikasi adalah proses pengelompokan benda berdasarkan ciri-ciri persamaan dan perbedaan. Dalam proses klasifikasi, terdapat target variabel kategori misalnya, penggolongan pendapatan dapat dipisahkan dalam tiga kategori yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

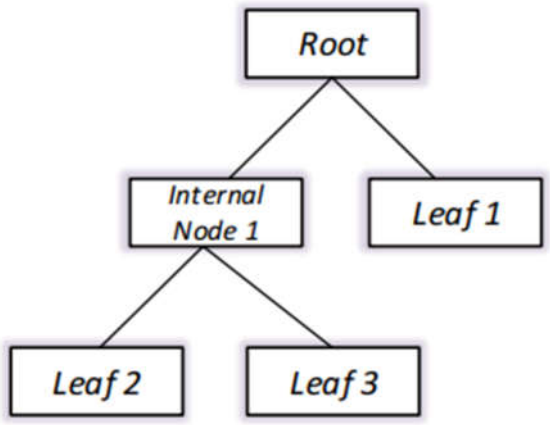
Metode-metode yang telah dikembangkan untuk menyelesaikan kasus klasifikasi yaitu pohon keputusan (*Decision Tree*), *naive bayes*, jaringan saraf tiruan, Analisis statistik, algoritma genetik, *rough sets*, pengklasifikasi *K-Nearest Neighbor*, metode berbasis aturan, *memory based reasoning*, *support vector machine*

2.6 Pohon Keputusan

Decision Tree (pohon keputusan) merupakan salah satu metode klasifikasi yang kuat dan terkenal. Metode pohon keputusan mengubah fakta yang besar menjadi pohon keputusan yang merepresentasikan aturan-aturan agar dapat dengan mudah untuk diinterpretasikan. Kegunaan lain dari pohon keputusan yaitu untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah variabel input dengan sebuah variabel target (Berry & Linoff, 2004). Proses pada *Decision Tree* adalah mengubah bentuk data tabel menjadi sebuah model *tree*. Model *tree* akan menghasilkan *rule* dan disederhanakan (Basuki & Syarif, 2003).

Jadi dapat disimpulkan bahwa, proses pada pohon keputusan yaitu mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule*,

serta menyederhanakan *rule.* Pohon keputusan juga disebut sebagai diagram alir yang berbentuk seperti struktur pohon yang mana setiap internal *node* menyatakan pengujian terhadap suatu atribut, setiap cabang menyatakan output dari pegujian tersebut sedangkan *node* daun (*leaf node*) menyatakan distribusi kelas. *Node* yang paling atas dikatakan sebagai *node* akar (*root node*). Teknik ini terdiri dari kumpulan *decision node*, dan dihubungkan oleh cabang, yang bergerak ke bawah dari *root node* sampai berakhir di *leaf node* (Yusuf, 2007). Algoritma yang dapat digunakan dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C4.5. Konsep dasar pohon keputusan terlihat pada Gambar 1.



Gambar 1. Konsep dasar pohon keputusan

Menurut Ratniasih (2016), metode pohon keputusan memiliki beberapa kelebihan, yaitu:

1. Interpretabilitas. *Decision tree* dapat dengan mudah diinterpretasikan dan dijelaskan kepada orang *non*-teknis. Struktur pohon keputusan membuatnya mudah untuk memahami alur pengambilan keputusan yang dilakukan oleh model.
2. Mengatasi data *non-linear*. *Decision tree* dapat mengatasi data yang tidak memenuhi asumsi linearitas. Model ini mampu menangani hubungan *non-*

linear antara fitur dan target dengan membagi data ke dalam subset berdasarkan ambang batas yang berbeda.

3. Skalabilitas. *Decision tree* efisien secara komputasi, terutama untuk dataset yang besar. Proses pembelajaran dan pengujian pada *decision tree* memerlukan waktu yang relatif singkat dibandingkan dengan beberapa algoritma pembelajaran mesin lainnya.
4. Menangani fitur campuran. *Decision tree* dapat dengan mudah menangani data dengan fitur campuran, baik fitur kategorikal maupun numerik, tanpa perlu banyak *pra*-pemrosesan data.
5. Tidak memerlukan asumsi distribusi. *Decision tree* tidak memerlukan asumsi tentang distribusi data atau variasi kesalahan, sehingga dapat digunakan secara luas dalam berbagai jenis data.

Adapun kekurangan pada metode pohon keputusan (Ratniasih, 2016):

1. *Overfitting.* *Decision tree* cenderung rentan terhadap *overfitting*, terutama jika tidak ada langkah-langkah yang diambil untuk mencegahnya. Pohon yang terlalu kompleks dapat menghafal data pelatihan dan tidak dapat digeneralisasi dengan baik pada data yang belum pernah dilihat sebelumnya.
2. Ketidakstabilan. *Decision tree* sangat sensitif terhadap perubahan kecil dalam data pelatihan. Perubahan kecil pada dataset pelatihan dapat menghasilkan pohon keputusan yang berbeda secara signifikan, yang dapat mengarah pada ketidakstabilan model.
3. Keterbatasan variabel korelasi. *Decision tree* cenderung tidak baik dalam menangkap hubungan kompleks antara fitur yang saling terkait. Jika hubungan antara fitur penting tergantung pada interaksi mereka, *decision tree* mungkin tidak dapat menggambarkannya dengan baik.
4. Tergantung pada pemilihan fitur. Keputusan yang dihasilkan oleh *decision tree* sangat tergantung pada fitur yang digunakan dalam pembuatan pohon. Jika fitur yang relevan tidak termasuk dalam dataset, model tidak dapat mempelajari hubungan tersebut.

5. Kesulitan menangani data numerik. *Decision tree* secara alami lebih baik dalam menangani data dengan fitur kategorikal daripada data numerik. Algoritma ini dapat menghadapi kesulitan dalam memahami pola yang rumit dalam data numerik

2.7 Algoritma C4.5

Algoritma C4.5 adalah kelompok algoritma *Decision Tree*. Algoritma ini memiliki input berupa *training samples* dan *samples*. *Training samples* dapat berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya. Sedangkan *samples* ialah field-field data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data (Sunjana, 2010). Algoritma C4.5 merupakan bagian dari kelompok algoritma pohon keputusan dan termasuk kedalam 10 algoritma yang paling populer.

Menurut Quinlan (1993), Algoritma C4.5 merupakan salah satu metode untuk membuat *Decision Tree* berdasarkan *training* data yang telah disediakan. Algoritma C4.5 dibuat oleh Ross Quinlan dimana algoritma ini merupakan pengembangan dari ID3 yang juga dibuat oleh Quinlan. Beberapa pengembangan yang dilakukan pada algoritma C4.5 antara lain dapat mengatasi *missing value*, dapat mengatasi *continue* data, dan *pruning*. Pohon keputusan telah banyak melakukan perkembangan tetapi yang sering digunakan adalah ID3 dan C4.5. Keduanya mempunyai prinsip yang sama dikarenakan Algoritma C4.5 adalah pengembangan dari algoritma ID3.

Perbedaan utama pada algoritma C4.5 yaitu :

1. C4.5 dapat menangani atribut bertipe kategorik dan numerik dan juga dapat menangani data *training* dengan nilai yang hilang atau data kosong.

2. Hasil yang didapat dari Algoritma C4.5 akan terpangkas setelah dibentuk pohon.
3. Pemilihan atribut pada algoritma C4.5 yang dilakukan dengan menggunakan *gain ratio*.

Ide dasar dari algoritma C4.5 yaitu pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau memiliki nilai *gain ratio* tertinggi berdasarkan nilai *entropy* atribut tersebut sebagai poros atribut klasifikasi. Pohon keputusan adalah struktur data berhierarki yang digunakan untuk menggambarkan keputusan dan konsekuensinya. Setiap simpul dalam pohon mewakili atribut atau fitur, dan setiap cabang mewakili nilai yang mungkin dari atribut tersebut. Daun pohon mewakili kelas atau label keputusan.

Menurut Larose (2005), Ada beberapa tahapan dalam membuat pohon keputusan dalam algoritma C4.5 yaitu:

1. Mempersiapkan data *training*.
   Data training mencakup fitur atau atribut yang akan digunakan dalam proses pengambilan keputusan, serta label atau keputusan yang akan diprediksi oleh pohon keputusan.
2. Menghitung akar dari pohon.
   Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai *gain ratio* dari masing-masing atribut, dimana nilai *gain ratio* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain ratio* dari masing-masing atribut, terlebih dahulu mencari nilai *entropy*. Apabila nilai *entropy* 0 maka akan bisa langsung ditarik kesimpulan. Artinya sudah tidak terdapat cabang pada atribut tersebut.

   Untuk menghitung nilai *entropy* digunakan rumus:

$$Entropy(S) = \sum_{i=1}^{n} -p_i \, log_2 \, (p_i) \qquad (2.1)$$

dimana :

S = himpunan kasus

n = jumlah partisi S

p_i = proporsi S_i terhadap S

3. Menghitung nilai *Gain Ratio*. Sebelum itu, terlebih dahulu mencari nilai *Gain Information* dan *Split Information* menggunakan rumus:

$$Informasi\ Gain(A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} Entropy(S_i) \qquad (2.2)$$

dengan :

S = himpunan kasus

A = fitur

n = jumlah partisi atribut A

$|S_i|$ = jumlah kasus pada partisi ke-i

$|S|$ = jumlah kasus dalam S

$$Split\ Info\ (S, A) = -\sum_{i=1}^{c} \frac{|S_i|}{|S|} log_2 \frac{|S_i|}{|S|} \qquad (2.3)$$

Sehingga didapat nilai *Gain Ratio* menggunakan rumus:

$$Gain\ Ratio = \frac{Informasi\ Gain\ (A)}{Split\ Info\ (S, A)} \qquad (2.4)$$

4. *Gain Ratio* terbesar akan menjadi *root* dan fitur dari *Gain Ratio* yang dipilih akan menjadi cabang.
5. Ulangi langkah ke 2 dan langkah ke 4 sehingga semua atribut terpartisi.
6. Proses partisi pohon keputusan akan berhenti pada saat semua atribut dalam simpul N mendapat kelas yang sama dan tidak ada atribut di dalam atribut yang dipartisi lagi. Artinya tidak ada atribut di dalam cabang yang kosong.

Perbedaan *Gain Ratio* dengan *Gain Information* yaitu:

a. *Gain Ratio* biasa digunakan untuk menentukan atribut yang betipe numerik dan kategorikal.

b. *Gain Ratio* hanya dimiliki oleh algoritma C4.5.
c. *Gain Ratio* bisa memberikan nilai sedikit lebih spesifik daripada *gain information.*

2.8 *Confusion matrix*

Confusion matrix merupakan alat evaluasi yang digunakan dalam *machine learning* dan statistik untuk mengevaluasi kinerja suatu model klasifikasi. Menurut Hidayatullah, dkk. (2019), *confusion matrix* merupakan matrix yang menyimpan informasi untuk mengetahui performa model dan sebagai acuan performa klasifikasi dari algoritma yang digunakan pada tahap evaluasi.

Tabel 1. *Confusion Matrix*

		Nilai Aktual	
		Positif (0)	Negatif (1)
Nilai Prediksi	Positif (0)	TP	FP
	Negatif (1)	FN	TN

Confusion matrix terdiri dari empat istilah utama:

1. *True Positive* (TP): Jumlah data yang diklasifikasikan dengan benar sebagai positif.
2. *True Negative* (TN): Jumlah data yang diklasifikasikan dengan benar sebagai negatif.
3. *False Positive* (FP): Jumlah data yang salah diklasifikasikan sebagai positif (disebut juga sebagai kesalahan tipe I).
4. *False Negative* (FN): Jumlah data yang salah diklasifikasikan sebagai negatif (disebut juga sebagai kesalahan tipe II).

Confusion matrix ini akan melakukan kalkulasi yaitu:

a. *Recall*

Recall (sensitivitas) mengukur kemampuan model untuk mengidentifikasi secara benar semua sampel positif. Ini adalah perbandingan antara jumlah positif yang berhasil diidentifikasi dengan total jumlah positif yang sebenarnya. Recall dinyatakan dalam persentase dan berkisar antara 0 hingga 100%. Recall yang tinggi menunjukkan bahwa model sangat baik digunakan.

$$recall = \frac{TP}{TP + FN} \tag{2.5}$$

b. *Precision*

Precision (presisi) mengukur kemampuan model untuk memberikan hasil positif yang tepat. Ini adalah perbandingan antara jumlah positif yang berhasil diidentifikasi dengan total jumlah hasil positif yang diberikan oleh model. *Precision* juga dinyatakan dalam persentase dan berkisar antara 0 hingga 100%. *Precision* yang tinggi menunjukkan bahwa model cenderung memberikan hasil positif yang benar-benar relevan.

$$precision = \frac{TP}{TP + FP} \tag{2.6}$$

c. *Accuracy*

Akurasi mengukur sejauh mana model klasifikasi mengklasifikasikan dengan benar keseluruhan sampel, baik positif maupun negatif. Ini adalah perbandingan antara jumlah prediksi yang benar dengan total jumlah sampel yang dinilai. Akurasi juga dinyatakan dalam persentase dan berkisar antara 0 hingga 100%. Namun, akurasi mungkin tidak menjadi metrik yang baik jika dataset memiliki ketimpangan kelas (*imbalance*), di mana jumlah sampel positif dan negatif tidak seimbang.

$$accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{2.7}$$

d. *F-Measure*

F-Measure (F1 *Score*) adalah ukuran gabungan dari *recall* dan *precision*, yang memberikan nilai rata-rata harmonik dari kedua metrik tersebut. *F-Measure* menggabungkan kedua metrik ini menjadi satu skor tunggal yang mencerminkan keseimbangan antara *recall* dan *precision*. Skor *F-Measure* berkisar antara 0 hingga 1, dan semakin tinggi skor *F-Measure*, semakin baik performa model.

$$F - Measure = 2 * \frac{precision * recall}{precision + recall} \quad (2.8)$$

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun akademik 2022/2023 bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh dari *website* Kaggle dengan judul "IBM HR *Analytics Employee Attrition* & *Performance*", data tersebut berisikan tentang atrisi karyawan. Dimana memiliki 8 atribut yaitu atrisi, peringkat kinerja, jenis kelamin, usia, status pernikahan, peran pekerjaan, pendidikan dan kehadiran. Serta terdapat 1470 observasi.

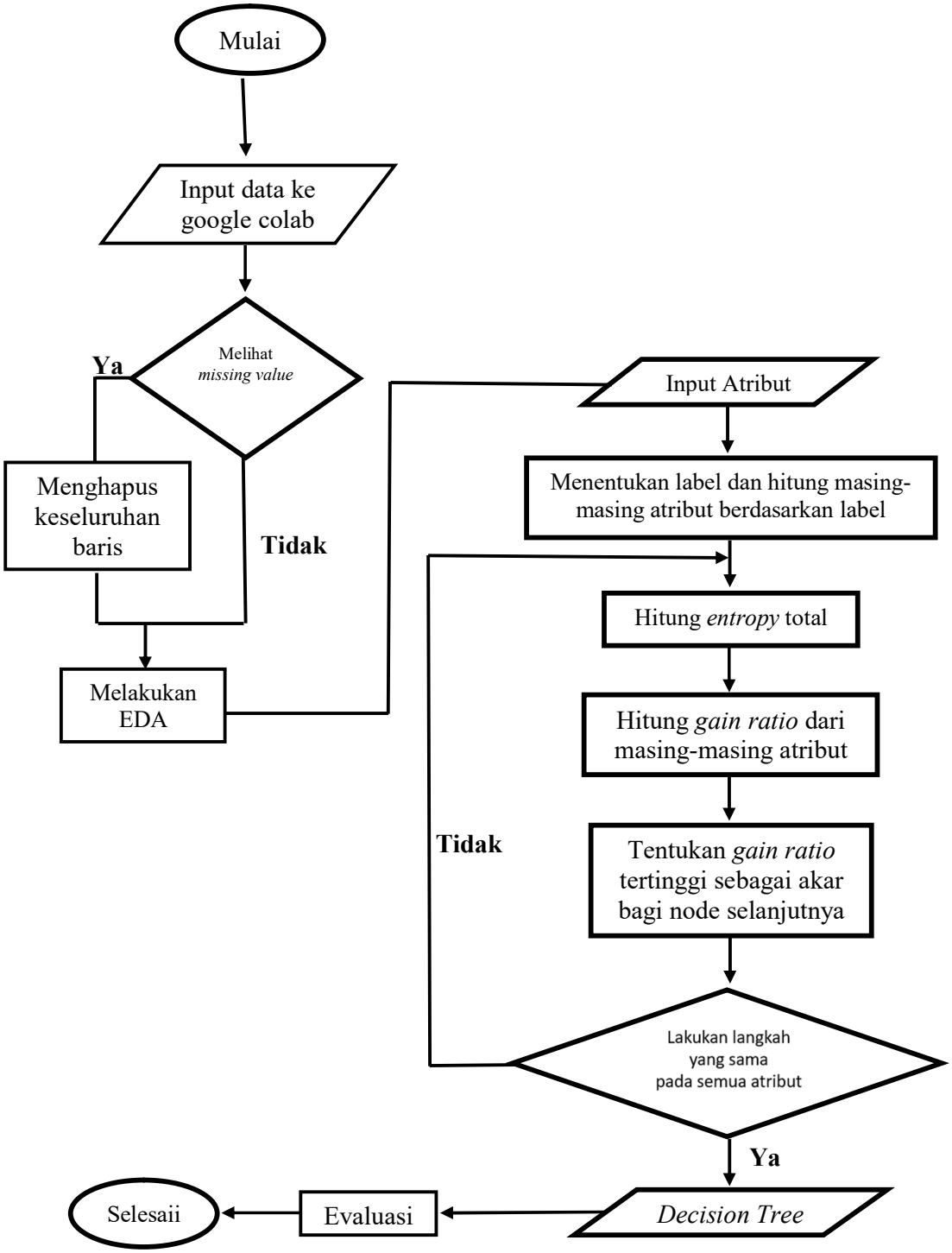
3.3 Metode Penelitian

Tahapan analisis data yang dilakukan dalam penelitian ini sebagai berikut:

1. Melakukan *data visualization*. Pada tahap ini menerapkan berbagai teknik yang disebut sebagai *Exploratory Data Analysis* (EDA) menggunakan *Google Colaboratory* sebagai berikut:
   a. Input data dari *Microsoft Excel* ke *Google Colaboratory.*

b. Melihat *missing value*, jika terdapat *missing value* maka lakukan penghapusan keseluruhan baris, namun jika tidak terdapat *missing value* maka dapat diproses ketahap selanjutnya.
c. Melakukan *Exploratory Data Analysis* (EDA), untuk melihat informasi yang ada pada data atrisi karyawan melalui grafik.

2. Melakukan data mining yang digunakan untuk memilih teknik dan algoritma yang sesuai untuk menemukan pola yang tersembunyi dari data atrisi karyawan dengan menggunakan *Decision Tree* algoritma C4.5 dengan perhitungan secara matematis yang dengan bantuan *Microsoft Excel*. Langkah perhitungan sebagai berikut:
   a. Menentukan label, pada data penelitian ini atribut yang digunakan sebagai label yaitu atrisi.
   b. Menghitung masing-masing atribut berdasarkan label.
   c. Hitung nilai *entropy* total dan *entropy* dari masing-masing atribut berdasarkan rumus 2.1.
   d. Hitung nilai *gain ratio* untuk masing-masing atribut berdasarkan rumus 2.4 Kemudian tentukan nilai *gain ratio* tertinggi.
   e. Atribut dengan nilai *gain ratio* tertinggi maka atribut tersebut dijadikan sebagai akar bagi node selanjutnya.
   f. Ulangi proses c sampai e unruk menentukan *node* selanjutnya hingga semua atribut diproses.
   g. Membuat pohon keputusan dari perhitungan *gain ratio* tertinggi sebagai *node* 1 dan seterusnya pada atribut lainnya.
   h. Evaluasi, pada tahap ini dilakukanya evaluasi dengan menggunakan *confusion matrix.*



Gambar 2. Diagram Alir Metode Penelitian

V. KESIMPULAN

Berdasarkan hasil dan pembahasan dapat disimpulkan bahwa:

1. Perusahaan cenderung mempertahankan karyawan yang sudah ada dan tidak melakukan pengurangan karyawan dalam jumlah yang besar, terbukti pada jumlah karyawan yang diatrisi lebih sedikit daripada karyawan yang tidak diatrisi.
2. Berdasarkan model *Decision Tree* yang terbentuk, didapatkan informasi bahwa faktor utama yang mempengaruhi atrisi karyawan adalah peringkat kinerja dengan cabang terakhir merupakan karyawan yang diatrisi dan tidak diatrisi.
3. Berdasarkan analisa hasil pengujian yang dilakukan menggunakan 75% data *training* dan 25% data *testing* didapatkan tingkat akurasi 88%. Hal ini menunjukan dari total 367 data atrisi karyawan sebagai data *testing* ada sebanyak 323 karyawan yang terklasifikasi secara benar dan 44 karyawan termasuk kedalam klasifikasi yang salah.

DAFTAR PUSTAKA

Anam, C & Santoso, H.B. 2018. Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa. *Energy-Jurnal Ilmiah Ilmu-Ilmu Teknik*. **8**(1): 13-19.

Basuki, A & Syarif, I. 2003. *Decision Tree*. Surabaya: Politeknik Electronika Negeri.

Berry, Michael J.A & Linoff, G.S. 2004. Data mining *Techniques For Marketing, Sales, Customer Relationship Management Second Edition*. United States of America.

Bramer, M. 2007. *Principles of* Data mining. Springer Science, London.

Fayyad, U. M. 1996. *Advances in Knowledge Discovery and* Data mining. Cambridge, Amerika Serikat.

Hidayatullah, A.F., Yusuf, A.A.F., Juwairi, K.P., & Nayoan, R.A.N. 2019. Identifikasi Konten Kasar pada Tweet Bahasa Indonesia. *Journal Linguistik Komputasional*. **2**(1): 1-5.

Kusrini, & Emha T. L. 2009. *Algoritma* Data mining. Yogyakarta.

Larose, D. T. 2005. *Discovering knowledge in data: an introduction to* data mining. New Jersey, Amerika Serikat.

Maringan, N. 2015. Tinjauan Yuridis Pelaksanaan Pemutusan (PHK) Secara Sepihak Oleh Perusahaan Menurut Undang-Undang No. 13 Tahun 2003 Tentang Ketenagakerjaan. *Jurnal Ilmu Hukum Legal Opinion*. **3**(3): 1-10.

Mardi, Y. 2017. Data mining: Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika Penelitian Bidang Komputer Sains dan Pendidikan Informatika*. **2**(2): 213-219.

Mitchell, T. A. 1997. *Machine Learning*. McGraw-Hill.

Mujilahwati, S. 2021. Visualisasi Data Hasil Klasifikasi Naïve Bayes Dengan Matplotlib Pada Python. *Prosiding SNST Fakultas Teknik*. **1**(1): 205-211.

Nisa, N.K., Nugraha, Y., Finola, C.F., Ernesto, A., Kanggrawan, J.I., & Suherman, A.L. 2020. Evaluasi Berbasis Data: Kebijakan Pembatasan Mobilitas Publik dalam Mitigasi Persebaran COVID-19 di Jakarta. *Jurnal Sistem Cerdas*. **3**(2): 84-94.

Radhi, M., Amalia., Sitompul, D.R.H., Sinurat, S.H., & Indra, E. 2021. Analisis Big Data Dengan Metode Exploratory Data Analysis (EDA) dan Metode Visualisasi Menggunakan Jupyter Notebook. *Jurnal Sistem Informasi dan Ilmu Komputer Prima*. **4**(2): 23-27.

Ratniasih, N.L. 2016. Konversi Data Training Tentang Pemilihan Kelas Menjadi Bentuk Pohon Keputusan Dengan Teknik Klasifikasi. *Jurnal Eksplora Informatika*. **4**(2): 145-154.

Setiawan, I., Suprihanto, S., Nugraha, A.C., & Hutahaean, J. 2020. HR Analytics: Employee Attrition Analysis Using Logistic Regression. In *IOP Conference Series: Materials Science and Engineering*. **830**(3): 1-7.

Sunjana, S. 2010. Aplikasi Mining Data Mahasiswa Dengan Metode Klasifikasi *Decision Tree*. *Jurnal Fakultas Hukum*. **2**(1): 24-28.

Quinlan, J.R. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, USA.

Wahyuni, E.D., Arifiyanti, A.A., & Kustyani, M. 2019. Exploratory Data Analysis Dalam Konteks Klasifikasi Data mining. *ReTII.* **1**(1): 263-269.

Wahyuningsih, S., & Utari, D. R. 2018. Perbandingan Metode K-Nearest Neighbor, Naive Bayes dan *Decision Tree* untuk Prediksi Kelayakan Pemberian Kredit. *Konferensi Nasional Sistem Informasi (KNSI).* **8**(1): 619-623.

Yusuf W.Y. 2007. Perbandingan Performasi Algortima *Decision Tree* C5. 0, CART, dan CHAD: Kasus Prediksi Status Resiko Kredit di Bank X. *Jurnal Jurusan Teknik Industri*. **1**(2): 59-62.