

## BAB II

### LANDASAN TEORI

#### 2.1. Tinjauan Pustaka

Rahman (2020) mengimplementasikan algoritma C4.5 pada jurnalnya yang berjudul “Prediksi Kelulusan Mahasiswa Menggunakan Algoritma C4.5 Studi Kasus di Universitas Peradaban. Penelitian ini bertujuan untuk mengetahui tingkat akurasi algoritma C4.5 dalam prediksi kelulusan mahasiswa, menggunakan metode decision tree. Data mahasiswa yang digunakan sebagai atribut adalah IPK, SKS, Umur, Jenis Kelamin. Hasil analisis pada penelitian ini menunjukkan bahwa algoritma C4.5 dapat memprediksi dengan nilai *accuracy* sebesar 88,74%; *precision* sebesar 91,79% dan *recall* sebesar 95,34%.

Tusarwenda (2018) mengimplementasikan algoritma C4.5 pada jurnalnya yang berjudul “Penerapan Data Mining Dengan Algoritma C4.5 dalam Prediksi Penjualan Botol pada CV.Seribukilo”. Penelitian ini dilakukan untuk membantu para penjual untuk memprediksi penjualan barang dagangannya, sehingga mereka dapat menyiapkan atau menyetok barang yang diprediksi akan mengalami kenaikan dalam penjualannya. Penjualan botol dapat diprediksi dengan akurasi yang cukup tinggi dengan menggunakan algoritma C4.5. Hasil pengukuran akurasi data yang diperoleh dari data training dengan *Confusion Matrix* nilainya mencapai 90,59% dan data testing tingkat akurasinya mencapai 88%.

Effendi (2018) mengimplementasikan algoritma C4.5 pada jurnalnya yang berjudul “Prediksi Penjualan Produk Roti Menggunakan Algoritma C4.5 pada PT. Prima Top Boga”. Penelitian dilakukan untuk menjaga kestabilan penjualan pada perusahaan. Data mining digunakan untuk membantu menyelesaikan permasalahan produksi. Algoritma C4.5 adalah algoritma yang mempunyai fungsi untuk memprediksi atau meramalkan sebuah kejadian dimasa mendatang dengan menggunakan data sebelumnya. Prediksi sangat menguntungkan bagi perusahaan agar terjaga kestabilan stok untuk memenuhi produk penjualan yang laku dan kurang laku dipasaran. Algoritma C4.5 mampu

memprediksi hasil penelitian dengan akurasi yang tinggi. Analisa data menggunakan algoritma ini memiliki tingkat akurasi sampai 93%.

Azwanti (2018) mengimplementasikan algoritma C4.5 pada jurnalnya yang berjudul “Analisa Algoritma C4.5 untuk Memprediksi Penjualan Motor pada PT. Capella Dinamik Nusantara Cabang Muka Kuning”. Penelitian ini membahas tentang analisa diperlukan untuk melihat pola dari data penjualan sehingga dapat menghasilkan prediksi penjualan motor yang nantinya akan berguna untuk pendistribusian motor di beberapa wilayah. Dari data konsumen yang begitu banyak, maka dilakukan data mining dengan menggunakan algoritma C4.5. Hasil dari kegiatan mining ini dapat memberikan sebuah pohon keputusan untuk melihat pola prediksi perilaku konsumen membeli motor.

Eska (2016) mengimplementasikan algoritma C4.5 pada jurnalnya yang berjudul “Penerapan Data Mining untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4.5”. Penelitian ini membahas tentang prediksi penjualan wallpaper. Berbagai motif, warna, dan corak wallpaper dapat dipilih. Oleh karena itu penjual harus lebih cermat menyediakan wallpaper mana yang akan banyak peminatnya, sehingga perlu rekomendasi jenis-jenis wallpaper dengan menggunakan metode klasifikasi dilakukan menggunakan data mining algoritma C4.5. data yang diperlukan yaitu data merk wallpaper terbaik, warna, motif, kualitas bahan, ukuran, dan harga. Dari hasil penelitian ini diperoleh kesimpulan yang menjadi faktor tertinggi yang mempengaruhi penjualan adalah faktor jumlah motif wallpaper. Faktor Harga, Ukuran, Kualitas Bahan, dan Warna tidak mempengaruhi pembelian karena wallpaper dengan harga mahal, ukuran yang kecil, kualitas bahan yang baik, dan warna yang sedikit ternyata masih diminati oleh pelanggan.

Mahendra (2018) mengimplementasikan data mining pada jurnalnya yang berjudul “Penerapan Data Mining Pada Peminjaman Buku Di Perpustakaan Universitas Sahid Surakarta Menggunakan Algoritma *FP-Growth*”. Perpustakaan Universitas Sahid Surakarta merupakan perpustakaan perguruan tinggi yang bertugas memberikan pelayanan kepada civitas akademika Universitas Sahid Surakarta, dalam menjalankan tugasnya, perpustakaan Universitas Sahid

Surakarta dibantu dengan aplikasi Usahid Library untuk mencatat proses sirkulasi (peminjaman, pengembalian buku). Selama ini pustakawan Universitas Sahid Surakarta belum melakukan analisa mendalam guna mengetahui karakteristik dari anggota perpustakaan lewat transaksi peminjaman buku yang sudah dilakukan. Dengan mengetahui karakter peminjaman dapat membantu pustakawan dalam menetapkan kebijakan pengembangan perpustakaan antara lain mengetahui pola peminjam. Transaksi tersebut diolah untuk mendapatkan informasi yang bermanfaat dengan melakukan analisa terhadap data peminjaman buku. Proses penelitian dilakukan dengan pendekatan data mining. Teknik data mining yang digunakan adalah *Association Rule*. *Association Rule* adalah hubungan antar item dengan menghitung beberapa kali item tersebut muncul dengan item yang berbeda dalam keseluruhan transaksi. Dan menggunakan metode algoritma *FP-Growth* yang merupakan proses analisa data untuk menemukan suatu pola dari kesimpulan data tersebut. Kesimpulan dari penelitian yang sudah dilakukan untuk periode 2016-2018 cenderung menjadikan kategori buku keperawatan sebagai referensi utama, dan kategori buku yang dipinjam secara bersamaan adalah kategori buku psikologi dan keperawatan karena merupakan top best rule dan kemunculan frekuensinya paling banyak dibanding pola yang lainnya selain itu juga mempunyai nilai *confidence* lebih dari 50%.

## 2.2. Kerangka Pemikiran

Kerangka pemikiran dalam penelitian ini dijelaskan sebagai berikut:

### 1. Latar Belakang Masalah

Belum adanya metode untuk memprediksi pendapatan retribusi Uji KIR pada tahun berikutnya. Selama ini penghitungan target masih dilakukan secara manual, sehingga antara target dan realisasi tidak sesuai.

### 2. Rumusan Masalah

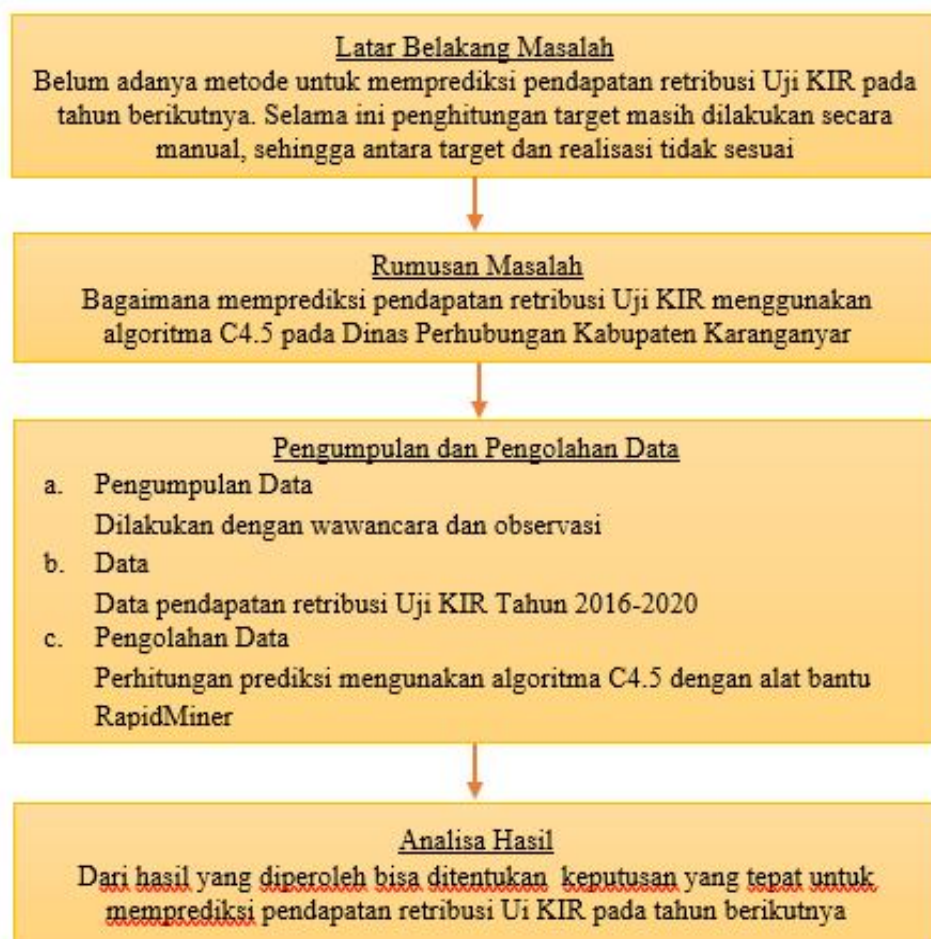
Bagaimana Memprediksi Pendapatan Retribusi Uji KIR Menggunakan Algoritma C4.5 pada Dinas Perhubungan Kabupaten Karanganyar?

### 3. Pengumpulan dan Pengolahan Data

a. Pengumpulan data dilakukan dengan wawancara dan observasi

- b. Data dalam penelitian ini adalah data pendapatan retribusi Uji KIR Tahun 2016-2020, yang diambil langsung pada Dinas Perhubungan Kabupaten Karanganyar melalui bendahara penerimaan.
  - c. Data yang siap diolah akan diproses perhitungan prediksi menggunakan algoritma C4.5 dengan alat bantu Rapidminer.
4. Analisa Hasil

Tahap ini dilakukan analisis hasil pengolahan data berupa prediksi pendapatan retribusi per jenis kendaraan sehingga dari hasil yang diperoleh bisa ditentukan keputusan yang tepat untuk memprediksi pendapatan retribusi Uji KIR pada tahun berikutnya.



Gambar 2.1 Kerangka Pemikiran

## 2.3. Landasan Teori

### 2.3.1. Prediksi

Prediksi atau peramalan pada dasarnya merupakan perkiraan suatu peristiwa di masa mendatang. Menurut Kamus Besar Bahasa Indonesia (KBBI) prediksi adalah hasil dari kegiatan memprediksi atau meramal atau memperkirakan nilai pada masa yang akan datang dengan menggunakan data masa lalu. Prediksi menunjukkan apa yang akan terjadi pada suatu keadaan tertentu dan merupakan *input* bagi proses perencanaan dan pengambilan keputusan.

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan) dapat diperkecil. Prediksi tidak harus memberikan jawaban secara pasti kejadian yang akan terjadi, melainkan berusaha untuk mencari jawaban sedekat mungkin yang akan terjadi (Roza et al., 2020).

Hasil peramalan dikatakan biasa apabila peramalan tersebut terlalu tinggi atau rendah dibandingkan dengan kenyataan yang sebenarnya terjadi. Hasil peramalan dikatakan konsisten apabila besarnya kesalahan dalam peramalan relatif kecil (Roza et al., 2020).

Teknik peramalan dapat dilakukan dengan dua cara, yaitu (Roza et al., 2020):

1. Secara kualitatif adalah cara penaksiran yang menitik beratkan pada pendapat seseorang. Hal ini penting karena hasil peramalan tersebut ditentukan berdasarkan pemikiran yang bersifat intuisi, pendapat dan pengetahuan dari orang yang menyusunnya.
2. Secara kuantitatif adalah cara penaksiran yang menitik beratkan pada perhitungan-perhitungan angka dengan menggunakan berbagai metode statistik. Hasil peramalan yang dibuat sangat bergantung pada metode yang dipergunakan dalam peramalan tersebut.

### **2.3.2. Pendapatan Daerah**

Menurut UU Nomor 33 Tahun 2004 pendapatan daerah yang selanjutnya disebut pendapatan adalah hak Pemerintah Daerah yang diakui sebagai penambah nilai kekayaan bersih dalam periode tahun bersangkutan. Salah satu sumber pendapatan berasal dari Pendapatan Asli Daerah (PAD) yang terdiri dari pajak daerah, retribusi daerah, hasil pengelolaan kekayaan yang dipisahkan dan lain-lain pendapatan asli daerah yang sah (Wulandari & Iryanie, 2018).

### **2.3.3. Retribusi Daerah**

Menurut UU Nomor 28 Tahun (2009) tentang Pajak Daerah dan Retribusi Daerah, pengertian retribusi daerah yang selanjutnya disebut retribusi merupakan pungutan daerah sebagai pembayaran atas jasa atau pemberian izin tertentu yang khusus disediakan dan/atau diberikan oleh pemerintah daerah untuk kepentingan pribadi atau badan.

Retribusi daerah dibagi menjadi 3 jenis yaitu:

#### **1. Retribusi Jasa Umum**

Retribusi Jasa Umum merupakan pungutan atas pelayanan yang disediakan atau diberikan pemerintah daerah untuk tujuan kepentingan dan kemanfaatan umum serta dapat dinikmati oleh orang pribadi atau badan.

Retribusi Jasa Umum dibagi ke dalam 15 bagian, yang meliputi: Retribusi Pelayanan Kesehatan, Retribusi Pelayanan Persampahan/Kebersihan, Retribusi Penggantian Biaya Cetak KTP dan Akta Catatan Sipil, Retribusi Pemakanan dan Pengabuan Mayat, Retribusi Pelayanan Parkir, Retribusi Pelayanan Pasar, Retribusi Pengujian Kendaraan Bermotor/Uji KIR, Retribusi Pemeriksaan Alat Pemadam Kebakaran, Retribusi Penggantian Biaya Cetak Peta, Retribusi Penyediaan dan/atau Penyedotan Kakus, Retribusi Pengolah Limbah Cair, Retribusi Pelayanan Tera/Tera Ulang, Retribusi Pelayanan Pendidikan, Retribusi Pengendalian Menara Telekomunikasi, dan Retribusi Pengendalian Lalu Lintas.

#### **2. Retribusi Jasa Usaha**

Retribusi Jasa Usaha merupakan pungutan atas pelayanan yang disediakan oleh Pemerintah Daerah dengan menganut prinsip komersial, baik itu pelayanan dengan menggunakan/memanfaatkan kekayaan daerah yang belum dimanfaatkan

secara optimal dan/atau pelayanan oleh pemerintah daerah sepanjang belum dapat disediakan secara memadai oleh pihak swasta.

Retribusi Jasa Usaha dibagi ke dalam 11 bagian, yaitu: Retribusi Pemakaian Kekayaan Daerah, Retribusi Pasar Grosir dan/atau Pertokoan, Retribusi Tempat Pelelangan, Retribusi Terminal, Retribusi Tempat Khusus Parkir, Retribusi Tempat Penginapan/Pesanggrahan/Vila, Retribusi Rumah Potong Hewan, Retribusi Pelayanan Kepelabuhan, Retribusi Tempat Rekreasi dan Olahraga, Retribusi Penyeberangan di Air, dan Retribusi Penjualan Produk Usaha Daerah.

### 3. Retribusi Perizinan Tertentu

Retribusi Perizinan Tertentu merupakan pungutan atas pelayanan perizinan tertentu oleh pemerintah daerah kepada pribadi atau badan yang dimaksudkan untuk pengaturan dan pengawasan atas kegiatan pemanfaatan ruang, penggunaan sumber daya alam, barang, sarana, atau fasilitas tertentu guna melindungi kepentingan umum dan menjaga kelestarian lingkungan.

Retribusi Perizinan tertentu dibagi ke dalam 6 jenis, yaitu: Retribusi Izin Mendirikan Bangunan (IMB), Retribusi Izin Tempat Penjualan Minuman Beralkohol, Retribusi Izin Gangguan, Retribusi Izin Trayek, Retribusi Izin Usaha Perikanan dan Retribusi Perpanjangan Izin Memperkerjakan Tenaga Asing (IMTA).

#### **2.3.4. Uji KIR**

Menurut Peraturan Menteri Perhubungan RI Nomor PM 133 Tahun (2015) tentang Pengujian Berkala Kendaraan Bermotor, pengujian kendaraan bermotor disebut juga Uji KIR adalah serangkaian kegiatan menguji dan/atau memeriksa bagian-bagian kendaraan bermotor, angkutan umum, kereta gandengan, kereta tempelan dan kendaraan khusus dalam rangka pemenuhan terhadap persyaratan teknis dan laik jalan.

Uji KIR dilakukan secara berkala setiap enam bulan sekali dengan tujuan untuk:

- a. Memberikan jaminan keselamatan secara teknis terhadap penggunaan kendaraan bermotor, kereta gandengan dan kereta tempelan di jalan.
- b. Mendukung terwujudnya kelestarian lingkungan dari kemungkinan pencemaran yang diakibatkan oleh penggunaan kendaraan bermotor, kereta gandengan dan kereta tempelan di jalan.
- c. Memberikan pelayanan umum kepada masyarakat.

Menurut Peraturan Daerah Kabupaten Karanganyar Nomor 20 Tahun (2019) tentang Perubahan Ketiga Atas Peraturan Daerah Kabupaten Karanganyar Nomor 4 Tahun 2012 Tentang Retribusi Jasa Umum, retribusi Uji KIR adalah pungutan sebagai pembayaran atas pelayanan pengujian kendaraan bermotor sesuai peraturan perundang-undangan yang berlaku. Jenis retribusi Uji KIR meliputi:

- a. Biaya uji berkala kendaraan bermotor.
- b. Biaya penggantian tanda bukti lulus uji.
- c. Biaya penggantian kartu uji dan tanda uji karena rusak/hilang.
- d. Biaya rekomendasi persetujuan numpang uji keluar.
- e. Biaya rekomendasi persetujuan mutasi keluar.

Kendaraan bermotor digolongkan menjadi dua yaitu:

- a. Kendaraan umum

Kendaraan umum adalah setiap kendaraan yang digunakan untuk angkutan barang dan/atau orang dengan dipungut bayaran.

- b. Kendaraan tidak umum

Kendaraan tidak umum adalah setiap kendaraan yang tidak termasuk dalam kendaraan umum (yang tidak dipungut bayaran).



Besarnya tarif retribusi Uji KIR tertera pada Tabel 2.1 sebagai berikut:

Tabel 2.1 Tarif Retribusi Uji KIR

No	Jenis Pelayanan	Tarif Retribusi	
		Umum	Tidak Umum
1	Biaya Uji Berkala Kendaraan Bermotor		
	a. J.B.B $\leq$ 3499 Kg	Rp 40.000	Rp 45.000
	b. J.B.B 3500 Kg s/d 8000 Kg	Rp 50.000	Rp 55.000
	c. J.B.B 8001 Kg s/d 15000 Kg	Rp 60.000	Rp 65.000
	d. J.B.B > 15000 Kg, Kereta Gandengan dan Kereta Tempelan	Rp 70.000	Rp 75.000
2	Biaya Penggantian Tanda Bukti Lulus Uji		
	a. Kartu Uji/Smart Card	Rp 5.000	Rp 5.000
	b. Kertas Berpengaman	Rp 5.000	Rp 5.000
	c. Stiker/Tanda Lulus Uji	Rp 5.000	Rp 5.000
3	Biaya Penggantian Kartu Uji dan Tanda Uji karena		
	a. Rusak	Rp 100.000	Rp 100.000
	b. Hilang	Rp 150.000	Rp 150.000
4	Biaya Rekomendasi Persetujuan Numpang Uji Keluar		
	a. J.B.B $\leq$ 3499 Kg	Rp 35.000	Rp 40.000
	b. J.B.B 3500 Kg s/d 8000 Kg	Rp 45.000	Rp 50.000
	c. J.B.B 8001 Kg s/d 15000 Kg	Rp 55.000	Rp 60.000
	d. J.B.B > 15000 Kg, Kereta Gandengan dan Kereta Tempelan	Rp 65.000	Rp 70.000
5	Biaya Rekomendasi Persetujuan Mutasi Uji Keluar		
	a. J.B.B $\leq$ 3499 Kg	Rp 30.000	Rp 35.000
	b. J.B.B 3500 Kg s/d 8000 Kg	Rp 40.000	Rp 45.000
	c. J.B.B 8001 Kg s/d 15000 Kg	Rp 50.000	Rp 55.000
	d. J.B.B > 15000 Kg, Kereta Gandengan dan Kereta Tempelan	Rp 60.000	Rp 65.000

Jumlah Berat Yang Diperbolehkan yang selanjutnya disebut JBB adalah berat maksimum kendaraan bermotor berikut muatannya yang diperbolehkan menurut rancangannya. JBB dari setiap jenis kendaraan bisa dilihat pada Sertifikat Registrasi Uji Teknis (SRUT).

### 2.3.5. Data Mining

Data mining adalah sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar, yang dapat juga diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Data mining merupakan proses yang menggunakan berbagai teknik dan alat analisis data untuk menemukan hubungan dan pola yang tersembunyi (Wanto et al., 2020).

Ada beberapa tugas yang dapat dilakukan oleh Data Mining dalam proses pemecahan masalah dan pencarian pengetahuan baru di antaranya adalah sebagai berikut (Wanto et al., 2020):

1. Klastering (*Clustering*)

Digunakan untuk mengelompokkan atau mengidentifikasi data yang memiliki karakteristik tertentu. Contoh algoritma: K-Means, K-Medoids.

2. Klasifikasi (*Classification*)

Digunakan untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Contoh algoritma: C4.5, K-Nearest Neighbor.

3. Asosiasi (*Association*)

Digunakan untuk mengatasi masalah bisnis yang khas, yakni dengan menganalisa tabel transaksi penjualan yang mengidentifikasi produk-produk yang sering kali dibeli bersamaan oleh *customer*, misalnya apabila orang membeli sambal, biasanya juga dia membeli kecap. Contoh algoritma: Apriori, Frequent Pattern Growth.

4. Estimasi (*Estimation*)

Digunakan untuk memperkirakan atau menilai suatu hal yang belum pernah ada sebelumnya yang disajikan dalam bentuk hasil kuantitatif. Contoh algoritma: Regresi Linier, Confidence Interval Estimations.

5. Prediksi (*Predictions*)

Digunakan untuk memperkirakan atau meramalkan suatu kejadian yang belum pernah terjadi. Contoh algoritma: Decision Tree, K-Nearest Neighbor.

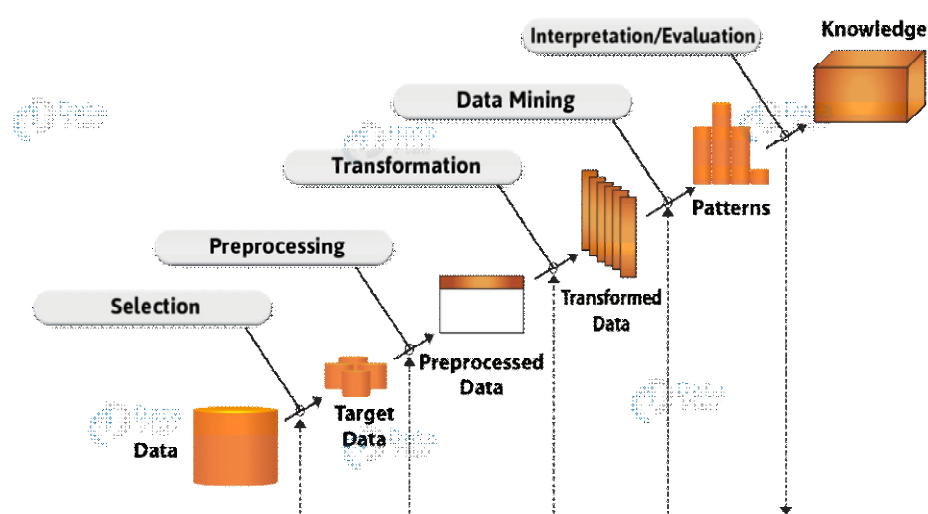
Data mining merupakan suatu proses pencarian pola dari data-data dengan jumlah yang sangat banyak yang tersimpan dalam suatu tempat penyimpanan dengan menggunakan teknologi pengenalan pola, teknik statistik, dan matematika (Saputra & Sibarani, 2020).

Karakteristik data mining adalah sebagai berikut (Saputra & Sibarani, 2020):

1. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
2. Data mining bisa menggunakan data yang sangat besar.
3. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
4. Data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategi.

Data mining sering juga disebut *knowledge discovery in database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining bisa dipakai untuk memperbaiki pengambilan keputusan dimasa depan.

Proses data mining dijelaskan pada Gambar 2.2 sebagai berikut (Saputra & Sibarani, 2020):



Gambar 2.2 Proses Data Mining

Tahapan-tahapan proses data mining, di antaranya:

1. Pembersihan Data (*Data Cleaning*)  
Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan.
2. Integrasi Data (*Data Integrasion*)  
Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru.
3. Seleksi Data (*Data Selection*)  
Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis atau yang cocok dengan data uji yang akan diambil dari *database*.
4. Transformasi Data (*Data Transformation*)  
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining.
5. Proses Mining  
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
6. Evaluasi Pola (*Pattern Evaluation*)  
Untuk mengidentifikasi pola-pola *knowledge based* yang ditemukan. Dalam tahap ini hasil dalam teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai.
7. Presentasi Pengetahuan (*Knowledge Presentation*)  
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

### 2.3.6. Algoritma C4.5

Algoritma C4.5 adalah salah satu metode untuk membuat *decision tree* berdasarkan training data yang telah disediakan. Algoritma C4.5 merupakan pengembangan dari ID3. Beberapa pengembangan yang dilakukan pada C4.5 antara lain bisa mengatasi *missing value*, bisa mengatasi data kontinyu. Pohon keputusan adalah hasil dari proses perhitungan *entropy* dan *information gain*,

setelah perhitungan berulang-ulang sampai semua atribut pohon memiliki kelas dan tidak bisa lagi dilakukan proses perhitungan (Wanto et al., 2020).

Alasan pemilihan algoritma C4.5 adalah, algoritma tersebut mampu menghasilkan sub sistem model base yang dapat digunakan untuk menunjang sistem pendukung keputusan. Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query Language* untuk mencari *record* pada kategori tertentu (Tusarwenda, 2018).

Algoritma C4.5 adalah salah satu metode untuk membuat *decision tree* berdasarkan training data yang telah disediakan. Pohon keputusan banyak sekali perkembangannya tetapi yang sering dipakai adalah ID3 dan algoritma C4.5. Keduanya mempunyai prinsip yang sama dikarenakan algoritma C4.5 merupakan pengembangan dari ID3, tetapi mempunyai perbedaan utama yaitu (Tusarwenda, 2018):

1. Algoritma C4.5 dapat menangani atribut yang kontinyu dan diskrit dan juga menangani data training dengan nilai yang hilang atau data kosong.
2. Hasil yang didapat dari Algoritma C4.5 akan terpangkas setelah dibentuk.
3. Pemilihan atribut yang dilakukan dengan menggunakan *Gain Ratio*.

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 yaitu (Nasrullah, 2018):

1. Menyiapkan data. Data biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas- kelas tertentu.
2. Menghitung nilai *entropy* dan gain.

Untuk menghitung nilai *entropy* digunakan rumus:

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

Keterangan:

S = himpunan kasus

$n$  = jumlah partisi  $S$

$p_i$  = proporsi dari  $S_i$  terhadap  $S$

Untuk menghitung nilai *gain* digunakan rumus:

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

Keterangan:

$S$  = himpunan kasus

$A$  = atribut

$n$  = jumlah partisi atribut  $A$

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

$|S|$  = jumlah kasus dalam  $S$

3. Menentukan akar pohon. Akar diambil dari atribut yang terpilih, dengan nilai *gain* yang paling tinggi.
4. Mengulangi proses di atas sampai semua kasus memiliki kelas yang sama. Proses partisi pohon keputusan akan berhenti saat :
  - a. Semua *record* dalam simpul berada didalam satu kelas yang sama.
  - b. Tidak ada atribut di dalam *record* yang dipartisi lagi.

### 2.3.7. Confusion Matrix

*Confusion matrik* adalah salah satu teknik yang dapat digunakan untuk mengevaluasi *performance* algoritma dari *Machine Learning (ML)*. Pada dasarnya *confusion matrix* mempresentasikan prediksi dan kondisi sebenarnya dari data yang dihasilkan oleh algoritma ML. Berdasarkan *confusion matrik* kita bisa menentukan nilai *Accuracy*, *Precision* dan *Recall* (Nugroho, 2019).

- a. *Accuracy* menggambarkan seberapa akurat model dalam mengklasifikasikan dengan benar.
- b. *Precision* menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model.
- c. *Recall* menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi.

*Confusion matrix* berbentuk tabel yang menggambarkan lebih detail tentang jumlah data yang diklasifikasikan dengan benar maupun salah. Tabel *confusion matrix* dijelaskan pada Gambar 2.3 sebagai berikut:

		Nilai Aktual	
		Positive	Negative
Nilai Prediksi	Positive	TP	FP
	Negative	FN	TN

Gambar 2.3 *Confusion Matrix*

Ada empat nilai yang dihasilkan didalam tabel confusion matrix, diantaranya:

- True Positive (TP)* : Jumlah data yang bernilai Positif dan diprediksi benar sebagai Positif.
- False Positive (FP)* : Jumlah data yang bernilai Negatif tetapi diprediksi sebagai Positif.
- False Negative (FN)* : Jumlah data yang bernilai Positif tetapi diprediksi sebagai Negatif.
- True Negative (TN)* : Jumlah data yang bernilai Negatif dan diprediksi benar sebagai Negatif.

### 2.3.8. RapidMiner

RapidMiner adalah *platform* perangkat lunak ilmu data yang dikembangkan oleh perusahaan ternama sama dengan yang menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran dalam, penambangan teks, dan analisis prediktif. Sejarah RapidMiner, sebelumnya dikenal sebagai YALE (*Yet Another Learning Environment*), mulai dikembangkan pada tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer dari Unit Kecerdasan Buatan Universitas Teknik

Dortmund. Mulai Tahun 2006, perkembangannya didorong oleh Rapid-I, sebuah perusahaan yang didirikan oleh Ingo Mierswa dan Ralf Klinkenberg. Pada tahun 2007, nama perangkat lunak itu berubah dari YALE menjadi RapidMiner.

Pada tahun 2013, perusahaan melakukan rebranding dari Rapid-I menjadi RapidMiner (Ramdhan et al., 2020).

RapidMiner dibangun menggunakan bahasa Java sehingga dapat dijalankan di berbagai sistem operasi seperti: Windows, Linux, Unix serta Mac OS. Sebagai rekomendasi sebaiknya menggunakan sistem 64 bit, dikarenakan jumlah maksimum yang dapat digunakan oleh RapidMiner terbatas pada sistem operasi dengan sistem 32 bit yaitu hanya sebesar 2GB (Ramdhan et al., 2020).

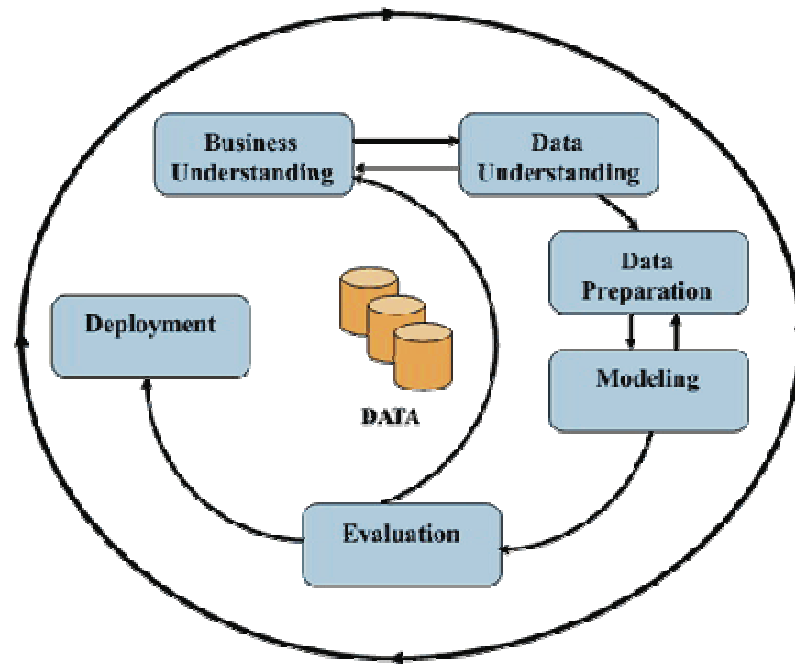
RapidMiner dikembangkan dengan model open core. Terdapat 2 versi edisi dari RapidMiner, yang pertama adalah RapidMiner Basic Edition (gratis), yang dibatasi kemampuannya pada 1 prosesor logis dan maksimal 10,000 baris data, tersedia dengan lisensi AGPL. Versi gratis dapat diunduh melalui tautan <https://rapidminer.com/get-started/>. Kedua adalah RapidMiner versi komersial dengan harga mulai dari \$2.500.

### **2.3.9. Metode CRIPS-DM**

Metode *Cross Industry Standard Process for Data Mining* (CRIPS-DM) merupakan sebuah metodologi yang menerapkan pendekatan terstruktur untuk perencanaan proyek data mining yang sangat ampuh dan sudah teruji dengan baik. Metode ini sangat umum digunakan karena sangat praktis, fleksibel, dan aplikatif untuk memecahkan isu bisnis yang sulit sekalipun. Metode ini merupakan metode andalan yang dapat dijalankan di hampir semua persoalan bisnis data mining (Kuncoro, 2021).

Proses data mining berdasarkan CRIPS-DM terdiri dari enam fase. Fase-fase metode CRIPS-DM dapat dilihat pada Gambar 2.4 sebagai berikut (Kuncoro, 2021):





Gambar 2.4 Proses CRIPS-DM

Proses data mining berdasarkan CRIPS-DM terdiri dari enam fase yaitu:

1. Fase Pemahaman Bisnis (*Business Understanding*)  
 Pada fase ini dibutuhkan pemahaman tentang substansi dari kegiatan data mining yang akan dilakukan. Kejadiannya antara lain; menentukan sasaran atau tujuan bisnis, memahami situasi bisnis, menentukan tujuan data mining dan membuat perencanaan strategi serta jadwal penelitian.
2. Fase Pemahaman Data (*Data Understanding*)  
 Adalah fase mengumpulkan data awal, mempelajari data untuk bisa mengenal data yang akan dipakai. Fase ini mencoba mengidentifikasi masalah yang berkaitan dengan kualitas data.
3. Fase Persiapan Data (*Data Preparation*)  
 Kegiatan yang dilakukan dalam fase ini antara lain: memilih *table* dan *field* yang akan ditransformasikan ke dalam *database* baru untuk bahan data mining.
4. Fase Pemodelan (*Modelling*)  
 Adalah fase menentukan teknik data mining yang digunakan, menentukan *tools* data mining, teknik data mining, algoritma data mining.

5. Fase Evaluasi (*Evaluation*)

Adalah fase interpretasi terhadap hasil data mining yang ditunjukkan dalam proses pemodelan pada fase sebelumnya. Evaluasi dilakukan secara mendalam dengan tujuan menyesuaikan model yang didapat agar sesuai dengan sasaran yang ingin dicapai dalam fase pertama.

6. Fase Penyebaran (*Deployment*)

Adalah fase penyusunan laporan atau presentasi dari pengetahuan yang didapat dari evaluasi pada proses data mining.