

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 Penelitian Terkait

##### 2.1.1 Penelitian Sebelumnya

Berikut adalah Tabel yang berisi penelitian terdahulu yang digunakan dalam penelitian ini. Tabel dibawah ini akan menjelaskan isi dari penelitian terdahulu

Tabel 2.1 Penelitian Sebelumnya

NO	Pengarang (Tahun)	Judul	Dataset	Metode	Hasil
1	Nurul Afni,Melan Susanti,Zulfajri (2020)	Prediksi Kepuasan Tenant Pada Gedung Wisma Keiai Menggunakan C4.5	50 tenant yang didapatkan dari Kuisisioner Pada Gedung wisma Keiai	Klasifikasi Menggunakan Algoritma C4.5	Hasil dari Penelitian ini dapat diperoleh kesimpulan yang menyatakan bahwa dengan menggunakan algoritma C4.5 diperoleh akurasi sebesar 100%. Dengan ini algoritma C4.5 bisa dikatakan sinkron digunakan untuk prediksi kepuasan Tenant pada Gedung Wisma Keiai
2	Fanny Adelia,Suhasda,Rafiqa Dewi (2019)	Penentuan Tingkat Kepuasan Pasien Bpjs Terhadap Kualitas Pelayanan di Rumah Sakit dengan Menggunakan Algoritma C4.5	50 Responden Pasien BPJS Rumah Sakit Tk. IV 01.07.01 Pematangsi antar	Klasifikasi Algoritma C4.5	Bedasarkan penelitian ini dapat diperoleh pernyataan bahwa dengan menggunakan algoritma C4.5 memiliki akurasi yaitu 93,33%. Dengan didapatkannya hasil yang sama antara perhitungan yang menggunakan <i>software rapidminer</i> dengan manual maka dapat disimpulkan bahwa penelitian ini berhasil

3	Anief Rufiyanto, Muhammad Rochcham, Abdul	Prediksi Kepuasan Mahasiswa dengan Menggunaka	284 dataset yang didapatkan dari mahasiswa	Algoritma C4.5	Hasil yang didapatkan dari penelitian tentang prediksi kepuasan mahasiswa terhadap pelayanan akademik di
---	-------------------------------------------	-----------------------------------------------	--------------------------------------------	----------------	----------------------------------------------------------------------------------------------------------



NO	Pengarang (Tahun)	Judul	Dataset	Metode	Hasil
	Rohman (2021)	n Algoritma C4.5 terhadap Pelayanan Akademik	Universitas Panandaran		Universitas Panandaran dengan menggunakan metode algoritma C4.5. aplikasi prediksi kepuasan mahasiswa dengan nilai akurasi 87,95% dan nilai AUC 0,995 sehingga termasuk klasifikasi data sangat baik
4	Kiki Rosita Dewi, Kemal Farouq Mauladi, Masruroh (2020)	Analisa Algoritma C4.5 untuk Prediksi Penjualan Obat Pertanian di Toko Dewi Sri	data penjualan obat pertanian pada Toko Dewi Sri diperoleh data	Menggunakan algoritma C4.5	Bedasarkan pengolahan data menggunakan <i>software rapidminer</i> didapatkan hasil dengan akurasi 75%. Dimana tingkat akurasi ini dapat di golongkan cukup baik. Sistem prediksi penjualan menggunakan algoritma C4.5 telah berhasil dibuat dan berfungsi dengan baik dengan fitur menu utama, fitur menu data admin, fitur menu tambah data admin, fitur menu data barang, fitur menu tambah data barang
5	Harry Dhika Tarigan, Fitriana Destiawati, Aswin Fitriansyah (2017)	Implementasi Algoritma C4.5 Terhadap Kepuasan Pelanggan	Objek penelitian diikuti oleh 300 konsumen restoran cepat saji di kota bogor dengan jumlah restoran cepat saji sebanyak 6 restoran	Algoritma C4.5	Dari hasil yang didapatkan dengan menggunakan <i>descission tree</i> dan juga <i>software rapidminer</i> dapat disimpulkan bahwa tingkat akurasi yang dihasilkan adalah sebesar 91%. Tingkat akurasi tersebut dapat digolongkan sebagai klasifikasi yang sangat baik untuk pengolahan data kepuasan pelanggan restoran cepat saji

### 2.1.2 Persamaan dan Perbedaan Penelitian Sebelumnya

Tabel 2.2 ini akan menjelaskan tentang persamaan dan perbedaan dengan penelitian sebelumnya.

Tabel 2.2 persamaan dan perbedaan dengan penelitian sebelumnya

NO	Judul	Pengarang	Persamaan	Perbedaan
1.	Prediksi Kepuasan Tenant Pada Gedung Wisma Keiai Menggunakan C4.5	Nurul Afni, Melan Susanti, Zulfajri (2020)	Menggunakan <i>software rapidminer</i> , dan juga algoritma C4.5 Pengambilan data menggunakan metode kuisioner	Penelitian ini memiliki objek penelitian yaitu Kepuasan Tenant Pada Gedung Wisma Keiai
2	Penentuan Tingkat Kepuasan Pasien Bpjs Terhadap Kualitas Pelayanan di Rumah Sakit dengan Menggunakan Algoritma C4.550	Fanny Adelia, Suhada, Rafiq Dewa (2019)	Pengujian penelitian ini menggunakan klasifikasi Algoritma Decision Tree C4.5 dan juga menggunakan <i>software Aplikasi RapidMiner Studio</i> . Memiliki tema yang sama yaitu tentang Kesehatan	Walaupun memiliki beberapa kesamaan tetapi penelitian ini memiliki objek yaitu Penentuan Tingkat Kepuasan Pasien Bpjs Terhadap Kualitas Pelayanan di Rumah Sakit
3	Prediksi Kepuasan Mahasiswa dengan Menggunakan Algoritma C4.5 terhadap Pelayanan Akademik	Anief Rufiyanto, Muhammad Rochcham, Abdul Rohman (2021)	Menggunakan metode algoritma C4.5 dan juga terdapat pohon keputusan untuk menentukan tingkat akurasi	Perbedaan dapat dilihat dari pengujian dan evaluasi penelitian ini menggunakan Cross Validation dan juga AUC. Objek penelitian ini adalah 284 data set yang di dapatkan dari mahasiswa

NO	Judul	Pengarang	Persamaan	Perbedaan
4.	Analisa Algoritma C4.5 untuk Prediksi Penjualan Obat Pertanian di Toko Dewi Sri	Harry Dhika Tarigan, ,Fitriana Destiawati, ,Aswin Fitriansyah (2020)	Menggunakan <i>software rapidminer</i> , menggunakan metode algoritma C4.5,dan juga terdapat pohon keputusan	Penelitian ini tidak mengambil datanya dengan metodekuisioner tetapi menggunakan data yang sudah terdapat di Obat Pertanian di Toko Dewi Sri
5	Implementasi Algoritma C4.5 Terhadap Kepuasan Pelanggan(2017)	Harry Dhika Tarigan, ,Fitriana Destiawati, ,Aswin Fitriansyah	Pengolahan data dilakukan dengan bantuan perangkat lunak <i>Rapidminer</i> . menggunakan klasifikasi Algoritma Decision Tree C4.5	Objek penelitian diikuti oleh 300 konsumen restoran cepat saji

## 2.2 Landasan Teori

### 2.2.1 Pasca Pandemi Covid-19

Pasca pandemi atau bisa juga disebut endemic adalah kondisi dimana munculnya suatu penyakit yang cenderung konstan atau penyakit tersebut sudah terbiasa dalam suatu area tertentu dan dapat di prediksi (Kemenkes Republik Indonesia,2022). Terjadinya perubahan dari masa pandemi Covid- 19 ke masa pasca pandemi membuat beberapa perubahan peraturan didalam pelayanan kepada pasien. Hal ini juga dapat berpengaruh terhadap kepuasan pasien. Maka dari itu,kesimpulannya adalah Pasca Pandemi Covid-19 adalah suatu kondisi dimana virus Covid-19 mulai mereda penyebarannya dan juga hanya terdapat di beberapa area tertentu di Indonesia.

### 2.2.2 Poli Rawat Jalan

Rawat jalan telah menjadi salah satu bagian terbesar pada pelayanan yang dilakukan di rumah sakit pada waktu yang akan datang. Instansi rawat jalan adalah pelayanan kesehatan pada salah satu unit di rumah sakit yang tidak lebih dari 24 jam dalam memberikan pelayanan pengobatannya, termasuk dalam prosedur terapi dan diagnostik. Instansi rawat jalan adalah sebuah pelayanan terapi & diagnostik dan juga pelayanan pengobatan yang diberikan pada pasien dalam artian pasien tidak membutuhkan perawatan secara intensif atau pengobatan rawat inap dalam mengobati penyakitnya (Anggraeni & Kurniawan, 2019). Proses pelayanan rawat jalan dimulai dari bagian pendaftaran, bagian administrasi atau kasir, bagian rekam medis, bagian keperawatan, bagian penunjang medik, dan bagian farmasi. Oleh karena itu, bila ada satu unit atau bagian yang bekerja tidak maksimal, maka akan sangat mempengaruhi lamanya proses pelayanan rawat jalan secara keseluruhan.

### 2.2.3 Fasilitas Pelayanan Kesehatan

Fasilitas adalah segala sesuatu yang menunjang kenyamanan konsumen dan disiapkan/disediakan oleh penyedia jasa/agen berupa perlengkapan fisik. Dari pernyataan tersebut dapat kita simpulkan bahwa fasilitas pelayanan adalah sarana dan prasarana yang tersedia untuk meningkatkan kepuasan pasien supaya pasien merasa nyaman. Selanjutnya Undang-Undang Nomor 36 Tahun 2009 tentang kesehatan menyatakan bahwa pemerintah daerah dapat menentukan jumlah dan jenis fasilitas pelayanan kesehatan serta pemberian izin beroperasi di daerahnya dengan mempertimbangkan luas wilayah, kebutuhan kesehatan, jumlah dan persebaran penduduk, pola penyakit, pemanfaatannya, fungsi sosial, dan kemampuan dalam memanfaatkan. Untuk mencapai tingkat kesehatan masyarakat yang baik, memerlukan fasilitas pelayanan kesehatan yang dapat menyediakan pelayanan kesehatan yang terjangkau bagi seluruh masyarakat

### 2.2.4 Pengertian Kepuasan Pasien

Kepuasan pasien merupakan salah satu indikator kualitas pelayanan medis. Survei menunjukkan bahwa Sebagian besar masyarakat baik di pedesaan dan

perkotaan kurang puas terhadap pelayanan yang diberikan oleh pihak rumah sakit daerah dan juga puskesmas sehingga banyak masyarakat memilih rumah sakit swasta yang diyakini dapat memberikan pelayanan yang lebih baik (Sharma & Kamra, 2013). Kepuasan pasien merupakan komponen utama atau penting. Jika pasien tidak puas terhadap pelayanan kesehatan yang diberikan, dia tidak akan mencari layanan itu atau menerimanya, walaupun layanan tersebut tersedia, mudah di dapat dan mudah di jangkau (Republik kementerian kesehatan Indonesia, 2016). Berdasarkan uraian pengertian tentang kepuasan pasien yang telah dijelaskan dapat disimpulkan bahwa respon puas atau tidak puasnya seorang pasien terhadap berbagai macam pelayanan yang telah diberikan oleh rumah sakit yang berhubungan dengan keinginan dan kebutuhan pasien

#### 2.2.5 Alat Ukur Kepuasan Pasien

Untuk mengetahui tingkat kepuasan yang dirasakan pelanggan atau penerima pelayanan maka perlu dilakukan pengukuran. Mengukur kepuasan dapat dimulai dengan mengidentifikasi pelanggan, memantau tingkat kualitas yang diinginkan, dan yang terakhir yaitu mengembangkan strategi. Selanjutnya, juga dijelaskan bahwa keinginan pelanggan dapat diperoleh pengalaman sebelumnya, saran dari kerabat, dan informasi dari penyedia jasa. Kepuasan pelanggan dapat dijelaskan melalui sikap pelanggan, tingkat kesukaan dan ketidaksukaan (ketidakpuasan) pelanggan terhadap pelayanan yang pernah didapatkan sebelumnya. Ada beberapa macam metode dalam pengukuran kepuasan pelanggan (Kotler, 2011:231 )

- a. Sistem keluhan dan saran Organisasi yang berorientasi pada pelanggan ( *customer oriented* ) memberikan kesempatan kepada para pelanggannya untuk menyampaikan keluhan dan saran. Misalnya dengan menyediakan kotak saran, kartu komentar, dan hubungan telepon langsung dengan pelanggan.
- b. *Ghost shopping* mempekerjakan beberapa orang untuk berperan atau bersikap sebagai pembeli potensial, kemudian melaporkan semuanya mengenai kekuatan dan kelemahan produk perusahaan dan pesaing berdasarkan pengalaman mereka.
- c. *Lost Customer Analysis* perusahaan seharusnya menghubungi para pelanggan yang telah berhenti membeli agar dapat memahami mengapa hal itu terjadi.

d. Survei kepuasan pelanggan penelitian *survey* dapat melalui pos, telepon dan wawancara langsung. Responden juga dapat diminta untuk mengurutkan berbagai elemen penawaran berdasarkan derajat pentingnya setiap elemen dan seberapa baik perusahaan dalam masing-masing elemen. Melalui survei perusahaan akan memperoleh tanggapan dan umpan balik secara langsung dari pelanggan dan juga memberikan tanda positif bahwa perusahaan menaruh perhatian terhadap para pelanggannya. Kepuasan dapat diukur dengan menggunakan metode yang telah dijelaskan di atas. Data yang diperoleh dari hasil pengukuran tiap - tiap metode mempunyai hasil yang berbeda. Pada penelitian yang menggunakan metode survei kepuasan pelanggan, data / informasi yang diperoleh menggunakan metode ini lebih fokus pada apa yang ingin diteliti sehingga hasilnya pun akan lebih valid

#### 2.2.6 Dimensi Kepuasan Pasien

Meskipun dimensi kepuasan yang dialami masyarakat sangat bervariasi, tetapi dimensi kepuasan yang didefinisikan secara umum mencakup hal-hal berikut (Azwar,1996) :

a. Keterampilan yang berkaitan dengan penerapan standar kode etika profesi. Pelayanan Kesehatan yang diberikan telah sesuai dengan standar dan kode etik profesi yang telah disepakati, atau dengan kata lain pelayanan kesehatan yang diberikan telah relevan dan sesuai dengan standar profesi yang berkompeten dan tidak melanggar dari ketentuan yang telah ada. Ukuran yang digunakan untuk menilai keyakinan masyarakat mengenai tingkat kepuasan yang diterima secara personal (*relationship*), kenyamanan layanan (*amenities*), kepemilikan (*choice*), pengetahuan dan kemampuan teknis (*scientific knowledge and technical skill*), efektifitas layanan (*effectiveness*) dan aktifitas keamanan (*safety*).

b. Kepuasan yang mengacu pada pemenuhan semua persyaratan pelayanan kesehatan. Persyaratan pelayanan medis dinyatakan sebagai pelayanan yang bermutu dan dapat memberikan kepuasan pada penerima jasa apabila pelaksanaan pelayanan yang diajukan atau ditetapkan, yang didalamnya mencakup penilaian terhadap Kepuasan pasien rawat inap mengenai ketersediaan pelayanan kesehatan (*available*), kewajaran.

Pelayanan kesehatan (*appropriate*), kesinambungan pelayanan kesehatan (*continue*), penerimaan pelayanan kesehatan (*acceptabel*), capaian pelayanan



kesehatan (*accessible*), keterjangkauan pelayanan kesehatan (*affordable*), efisiensi pelayanan kesehatan (*efficient*) dan mutu pelayanan kesehatan (*quality*). Memberikan layanan kesehatan yang memenuhi persyaratan kinerja bukanlah hal yang mudah. Untuk mengatasi hal tersebut diterapkan prinsip kepuasan yang terkombinasi secara selektif dan efektif, dalam artian menerapkan dimensi kepuasan kelompok pertama dilakukan secara optimal, sedangkan beberapa dimensi kelompok kedua dilakukan secara selektif yaitu yang sesuai dengan kebutuhan serta kemampuan.

Pelayanan kesehatan diakui apabila memenuhi kebutuhan kepuasan pasien apabila pelayanan yang diberikan mengikuti standar serta kode etik yang disepakati dalam suatu profesi atau dengan kata lain yaitu apabila pelayanan kesehatan yang diberikan telah sesuai dengan standar yang telah ditetapkan oleh profesi yang tidak menyimpang dari kode etik yang berlaku bagi profesi tersebut. Mengemukakan bahwa terdapat 5 dimensi *service quality* untuk menentukan kepuasan pasien yaitu sebagai berikut (A.Pasuraman,2008):

1. Tangible (Kualitas Pelayanan)
2. Assurance (Kepastian)
3. Empati
4. Reliability (Keandalan)
5. Responsiveness (Daya Tanggap)

#### 2.2.7 Data Mining

Menurut penelitian terdahulu data mining merupakan penemuan informasi terbaru dengan cara mencari aturan tertentu dari beberapa data yang besar (Davies, 2004). Istilah data mining terbagi menjadi 2, yaitu *knowledge discovery* dan *pattern recognition*. Istilah tersebut memiliki nilai benarnya tersendiri (Patty, 2010). Disimpulkan bahwa data mining adalah proses menemukan pola dalam sejumlah besar data yang disimpan dalam *database*, data warehouse, dan penyimpanan informasi lainnya. Data mining didefinisikan sebuah proses menemukan pola dalam sebuah data. Proses ini otomatis atau sering kali semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar. Kata *mining* berarti mendapatkan sedikit data

berharga dari jumlah material dasar. Oleh karena itu data mining berasal dari bidang ilmu seperti machine learning, statistik, dan kecerdasan buatan.

### 2.2.8 Tujuan Data Mining

Data mining memiliki tujuan yang dapat dikelompokkan menjadi 2 bagian yaitu:

#### a. Prediksi

Prediksi adalah data yang dikategorikan berdasarkan perilaku atau nilai yang diharapkan. Prediksi juga memiliki beberapa kesamaan dengan klasifikasi. Salah satu contoh tugas prediksi adalah memprediksi jumlah kenaikan angka kasus Covid-19 dalam jangka 1 tahun di provinsi Jawa Timur

#### b. Deskripsi

Deskripsi ini mempunyai tujuan untuk dapat mengidentifikasi sebuah pola secara berulang dalam sebuah data. Deskripsi juga dapat menerjemahkan sebuah pola ke dalam aturan dan kategori yang dapat mudah dipahami untuk dapat meningkatkan pengetahuan sistem secara efektif. Deskriptif juga memiliki tugas yaitu tugas data mining yang sering dibutuhkan dalam Teknik post-processing untuk dapat menjelaskan dan memvalidasi dari proses data mining. Pasca pemrosesan adalah proses yang cukup membuat hasil yang valid dan tersedia oleh pihak yang membutuhkan

### 2.2.9 Metode Data Mining

Data mining terbagi menjadi beberapa kelompok yang berdasarkan pada aktivitas yang telah dilakukan yaitu klasifikasi, asosiasi, *clustering*, *regression*, deskripsi, dan estimasi. Berikut adalah penjelasan beberapa kelompok tersebut.

- **Asosiasi**

Asosiasi adalah metode berbasis aturan yang digunakan untuk menemukan asosiasi dan hubungan variabel dalam satu set data. Biasanya analisis ini terdiri dari pernyataan “if atau then” sederhana. Association banyak digunakan dalam mengidentifikasi korelasi produk dalam keranjang belanja untuk memahami kebiasaan konsumsi pelanggan. Sehingga, perusahaan dapat mengembangkan strategi penjualan dan membuat sistem rekomendasi yang lebih baik.

- **Klasifikasi**

Selanjutnya klasifikasi, merupakan metode yang paling umum digunakan dalam data mining. Klasifikasi adalah kegiatan untuk memprediksi kelas suatu objek.

- **Regression**

Regression merupakan teknik yang menjelaskan variabel melalui proses analisis variabel independen. Sebagai contoh, prediksi penjualan suatu produk berdasarkan korelasi antara harga produk dengan tingkat pendapatan rata-rata pelanggan.

- **Clustering**

Metode clustering. Clustering digunakan dalam membagi kumpulan data menjadi beberapa kelompok berdasarkan kemiripan atribut yang dimiliki. Contoh kasusnya adalah *Customer Segmentation* membagi pelanggan ke dalam beberapa grup berdasarkan tingkat kemiripannya.

#### 2.2.10 Tahapan Data Mining

Karena data mining adalah sebuah untai proses, maka pecah menjadi beberapa tahap. Tahapan tersebut akan bersifat interaktif, pengguna akan terlibat langsung (Eska, 2016). Diawali dari seleksi data dari data sumber ke data target, tahap pre-processing untuk memperbaiki kualitas data, transformasi, data mining serta tahap interpretasi dan evaluasi yang menghasilkan output berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik. Tahapan data mining dibagi menjadi enam bagian yaitu :

1. Pembersihan data (*data cleaning*) sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang 4 duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.
2. Integrasi data (*data integration*) Integrasi data melibatkan penggabungan data dari berbagai *database* yang digabungkan ke dalam *database* baru. Tidak jarang data yang diperlukan untuk data mining berasal dari berbagai macam *database* tetapi juga dari beberapa *database* atau file teks. Integrasi data dilakukan

menggunakan atribut mengidentifikasi entitas unik, contoh atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data harus dilakukan secara detail, karena kesalahan dalam integrasi data bisa memberikan hasil yang tidak tepat dan menyesatkan pengambilan aksi nantinya. Misalnya, saat integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori berbeda, hal ini dapat mengakibatkan korelasi antar produk yang sebenarnya tidak ada.

2. Seleksi Data (*Data Selection*) Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analisis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

3. Transformasi data (*Data Transformation*) data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

4. Proses mining adalah sebuah proses yang paling utama pada saat metode diterapkan untuk mencari pengetahuan tersembunyi dan berharga dari data.

5. Evaluasi pola (*pattern evaluation*), untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai.

6. Presentasi pengetahuan (*knowledge presentation*), merupakan penyajian dan visualisasi pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam

proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining.

#### 2.2.11 Rapid Miner

RapidMiner merupakan perangkat lunak untuk pengolahan data. Rapid Miner adalah perangkat lunak data ilmu pengetahuan yang dikembangkan oleh perusahaan dengan nama yang sama, dan menyediakan lingkungan terintegrasi untuk pembelajaran mesin (*machine learning*), pembelajaran mendalam (*deep learning*), penambangan teks (*text mining*), dan analisis prediktif (*predictive analytics*). Rapidminer adalah salah satu software untuk pengolahan data mining. Kinerja rapidminer text mining adalah berkisar dengan analisis teks, mengekstrak pola-pola dari dataset yang besar dan mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan *database* (Faid et al., 2019). *RapidMiner* mempermudah pengguna dalam menghitung data dalam jumlah yang besar dengan mudah menggunakan operator. Operator ini dapat digunakan untuk memodifikasi data. Data dihubungkan dengan node pada operator kemudian disambungkan ke node hasil untuk melihat hasilnya. Hasil oleh *RapidMiner* pun dapat ditampilkan secara visual dengan grafik (Rahmat et al., 2017).

#### 2.2.12 Algoritma C4.5

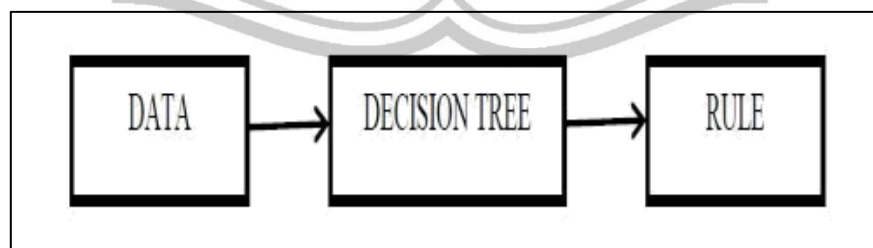
Algoritma C4.5 adalah algoritma yang mempunyai klasifikasi data bertipe pohon keputusan (Ramadhan et al., 2020). Pohon keputusan menyerupai struktur sebuah pohon yang terdapat node internal (bukan daun) yang menjelaskan tentang atribut atribut, setiap cabang mengdeskripsikan hasil dari atribut yang diuji (Adelia et al., 2019). Dan di setiap daun menggambarkan kelas. Teknik pohon keputusan ini juga disebut *decision tree*. Terdapat beberapa aturan dengan tujuan meningkatkan tingkat ketelitian dari sebuah prediksi yang sedang dilaksanakan (Dewi et al., 2020). Algoritma ini mempunyai inputan berupa *trainingsamples*. *Training samples* adalah contoh data yang digunakan sebagai membangun sesuatu *tree* (pohon) yang sudah diuji kebenarannya (Athiyah et al., 2021). Juga terdapat *samples*, *samples* merupakan *field-field* data yang

selanjutnya akan digunakan untuk parameter dalam melaksanakan klasifikasi data.

Untuk membangun sebuah pohon keputusan perlu adanya sebuah persyaratan:

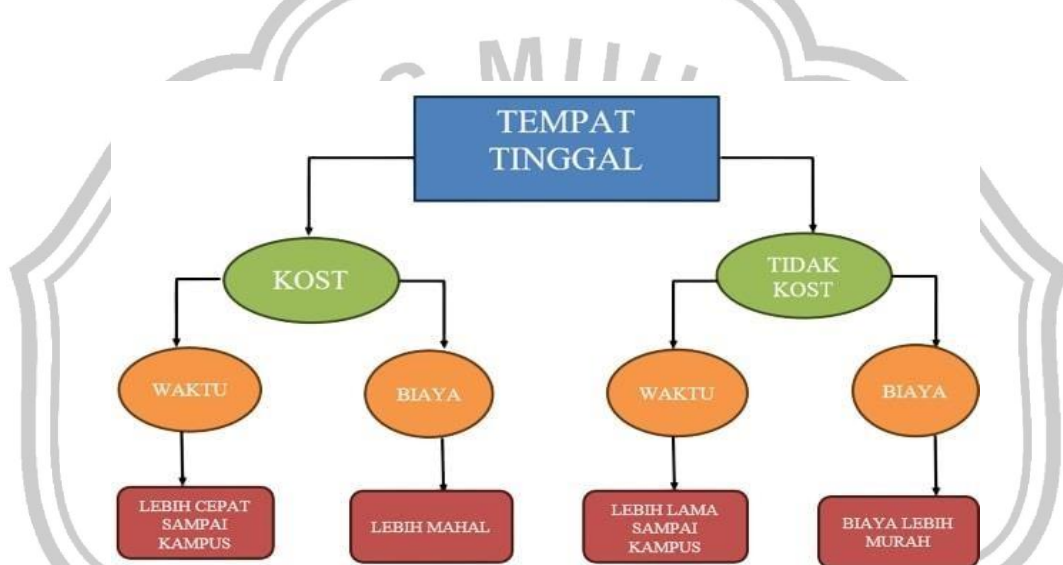
1. Pilih atribut sebagai akar
2. Buat cabang untuk tiap nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk setiap cabang sampai semua memiliki kelas yang sama (Rani, 2016)

Pohon keputusan atau decision tree merupakan teknik data mining yang digunakan untuk mengeksplorasi data dengan membagi kumpulan data yang besar menjadi himpunan record yang lebih kecil dan memperhatikan variabel tujuannya. Teknik ini dapat diterapkan dalam menentukan tingkat kualitas mahasiswa yang melibatkan banyak data (Zega, 2014). Pohon keputusan juga memiliki definisi yaitu salah satu mode klasifikasi yang paling terkenal dan juga paling kuat. Metode pohon keputusan ini mengubah fakta menjadi sebuah pohon keputusan yang memaparkan aturan supaya mudah dipahami oleh manusia. *Decision Tree* juga digunakan untuk mengeksplorasi data dan menemukan hubungan tersembunyi antara sekumpulan variabel input dan variabel tujuan (Undavia et al., 2017). Model pohon keputusan terdiri dari serangkaian keputusan yang membagi populasi besar menjadi aturan-aturan yang lebih kecil dengan mempertimbangkan tujuan dalam bentuk objek. Secara umum, tujuan dari klasifikasi model keputusan berfokus pada perhitungan yang memungkinkan setiap catatan diinputkan ke dalam kategori yang berbeda atau setiap catatan dikelompokkan kedalam kelas. Konsep dari decision tree atau pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan keputusan.



Gambar 2.1 Konsep *Descision Tree*

Klasifikasi pohon keputusan merupakan teknik klasifikasi yang sederhana yang banyak digunakan. Bagian ini menjelaskan cara kerja pohon keputusan dan bagaimana pohon keputusan dibangun. Seringkali untuk mengklasifikasikan objek, ajukan urutan pertanyaan sebelum memutuskan pengelompokannya. *Decision Tree* berkaitan erat dengan algoritma C4.5, sebab dasar dari algoritma C4.5 adalah pohon keputusan. Algoritma C4.5 merupakan salah satu yang digunakan untuk melakukan prediksi klasifikasi atau pengelompokan. Cabang pohon keputusan adalah pertanyaan klasifikasi dan daunnya adalah segmen atau kelas



Gambar 2.2 Contoh Sederhana Pohon Keputusan

Algoritma C4.5 digunakan untuk model klasifikasi, maka dari itu hasil dari pengolahan catatan pengujian berupa pengelompokan data ke dalam kelas kelas. Berikut adalah langkah - langkah algoritma C4.5 yang digunakan sebagai bentuk pohon keputusan sebagai berikut:

- Pilih atribut sebagai root
- Buat cabang untuk setiap nilai
- Bagi tiap cabang kedalam kelas
- Ulangi terus proses untuk semua cabang sampai semua kasus pada tiap cabang memiliki kelas yang sama
- Dalam memilih atribut sebagai *root*, pemilihan didasarkan pada nilai gain tertinggi dari atribut yang telah ada. Selain itu, untuk mendapatkan nilai gain

Ulangi terus proses untuk semua cabang sampai semua kasus pada tiap cabang memiliki kelas yang sama

Dalam memilih atribut sebagai *root*, pemilihan didasarkan pada nilai gain tertinggi dari atribut yang telah ada. Selain itu, untuk mendapatkan nilai gain tertinggi, diperlukannya menghitung nilai entropy pada semua nilai didalam atribut. Entropy berfungsi sebagai parameter guna mengukur varian dari data sampel. Selanjutnya, setelah mengetahui nilai entropy dari data sampel, maka atribut yang memiliki pengaruh terbesar adalah ukuran klasifikasi data. Ukuran tersebut juga disebut information gain. Berikut adalah rumus menghitung entropy pada algoritma C4.5

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

Dengan :

$S$  = jumlah sampel data (*sampling*) / himpunan data set

$n$  = jumlah partisi  $S$

$p_i$  = probabilitas yang berasal dari sum(ya) atau sum(tidak dibagi total kasus/ proporsi dari  $S_i$  terhadap  $S$

Setelah nilai entropy telah ditemukan dari semua kasus maka dilakukanlah analisis pada setiap atribut dan nilai nilainya, selanjutnya hitung entropinya. Langkah berikutnya yaitu menghitung gain. *Gain* ( $S, A$ ) merupakan selisih nilai total entropi dikurangi nilai entropi masing-masing nilai setiap atribut kriteria dikalikan nilai proporsi nilai atribut dibagi jumlah sampel data. Fungsi dari nilai *gain* yang diperoleh adalah untuk mengukur keefektifan setiap atribut *kriteria* dalam mengklasifikasikan data. Untuk rumus menghitung nilai gain yaitu sebagai berikut:

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

Keterangan:

$S$  = jumlah sampel data (*sampling*) / himpunan data set

$A$  = *attribute*

$n$  = Jumlah partisi  $S$

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

$|S|$  = Jumlah kasus dalam  $S$



### 2.2.13 Akurasi, Presisi, dan Recall

Recall adalah suatu tingkat keberhasilan suatu sistem dalam mengambil informasi, sedangkan presisi adalah suatu tingkat ketepatan antara informasi yang diinginkan oleh pengguna dengan respon yang diberikan oleh sistem. Akurasi adalah ukuran tingkat interaksi kedekatan antara nilai yang sebenarnya dan nilai yang diharapkan. Untuk menghitung recall, presisi, dan akurasi membutuhkan salah satu sistem kerja yang dipakai pada umumnya yaitu *confusion matrix*. *Confusion matrix* merupakan suatu metode untuk memberikan informasi tentang hasil dari klasifikasi yang dilakukan oleh sistem yang berguna untuk menganalisis seberapa baik *classifier* mengenali tuple dari kelas yang berbeda. *Confusion matrix* juga dapat diartikan sebagai tabel yang digunakan untuk mencatat hasil kinerja klasifikasi. Berikut adalah tabel *confusion matrix* untuk melakukan klasifikasi dua kelas.

Confusion Matrix		Prediksi	
		False	True
Aktual	False	TN	FP
	True	FN	TP

Tabel 2.3  
Confusion  
Matrix

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Keterangan :

1. TP (True Positive) = banyak data yang benar sebagai kelas positif dan yang teridentifikasi positif
2. FP (False Positive) = jumlah data positif yang teridentifikasi sebagai kelas negatif
3. FN (False Negative) = jumlah data negatif yang teridentifikasi sebagai kelas positif
4. TN (True Negative) = jumlah data yang teridentifikasi dengan benar sebagai kelas negatif

Apabila nilai TP, TN, FN, dan FP telah didapatkan maka nilai tingkat akurasi, presisi, dan

recall dari setiap model dapat dihiung dengan persamaan

