

## BAB II TINJAUAN PUSTAKA

### 2.1 Penelitian Terkait

Penelitian sebelumnya yang menjadi latar belakang penelitian ini dijabarkan pada tabel dibawah ini :

Tabel 2.1. Penelitian Terkait

NO	JUDUL, PENULIS, TAHUN	DATASET	METODE	HASIL	KEKURANGAN	KELEBIHAN
1	PENERAPAN ALGORITMA C4.5 UNTUK PREDIKSI PENGGUNAAN JENIS KONTRASEPSI BERBASIS WEB  RUSDA WAJHILLAH  2017	data para akseptor KB di Bidan Praktek Mandiri dan Posyandu Desa Nyangkowek, Cicurug.	ALGORITMA C4.5	akurasi sebesar 85,38%	Terdapat 130 jumlah kasus yang terdiri dari 104 kasus akseptor pengguna kontrasepsi hormonal dan 26 kasus pengguna kontrasepsi non hormonal	Sistem pakar dibuat agar membantu para pengguna khususnya para akseptor KB dalam mendapatkan informasi mengenai penggunaan jenis kontrasepsi yang tepat, tanpa harus berkonsultasi langsung dengan para tenaga medis
2	PERANCANGAN SISTEM PAKAR UNTUK MENENTUKAN ALAT KONTRASEP	Data pasien diperoleh dari Puskesmas Cijayanti	ALGORITMA C4.5	akurasi 92%	Data sampling sangat sedikit hanya 150 data	Menggunakan pembuatan aplikasi menentukan keputusan.

	SI MENGGUNA KAN ALGORITMA C4.5  NI WAYAN PARWATI SEPTIANI  2014					
3.	SISTEM PENDUKUN G KEPUTUSAN PEMILIHAN METODE KONTRASEP SI DI PUSKESMAS II COLOMADU DENGAN ALGORITMA NAÏVE BAYES	Dataset Kecamatan Colomadu pada tahun 2013  45 sampel	Naïve Bayes	akurasi 82,2 %.	Variabel pada pemilihan metode kontrasepsi dapat ditambah lagi pilihannya	Pemilihan metode kontrasepsi dapat dilakukan dengan menggunakan bantuan sistem pendukung keputusan
4.	Pengaruh Media Terhadap Pengambilan Keputusan Dalam Menjalankan Program Keluarga Berencana	kaggle.com	Algoritma Decision Tree	91,58%	-	Penelitian pada dataset program Keluarga Berencana yang diperoleh dari hasil survei prevelensi Kontrasepsi

	<p>Dengan Algoritma Decision Tree</p> <p>ALI MUSTOPA, SITI KHOTIMATU L WILDAH, GANDA WIJAYA, WINDA GATA, SARIFAH AGUSTIANI</p> <p>2020</p>					<p>Nasional Indonesia kepada wanita yang sudah menikah yang didasarkan pada pengambilan keputusan program keluarga berencana.</p>
5	<p>Metode Clustering Pada Model Algoritma K-Means Untuk Pemilihan Alat Kontrasepsi</p> <p>Diah Ayu Maulida Wati 1</p> <p>2019</p>	<p>data primer, data dan informasi penelitian ini diperoleh dari instansi terkait yaitu data pengguna alat kontrasepsi pada tahun 2017 sebanyak 100</p>	K-Means	<p>48% untuk cluster 1 yang terdiri dari , usia 36 tahun, jumlah anak 2, pendidikan SMA, status kerja tidak kerja. 22%</p>	<p>Hanya menggunakan 100 data yang berada di Rumah Sakit Annisa.</p>	<p>Melakukan pengelompokan berdasarkan pada ciri atau atribut pada variabel penentuan penggunaan alat kontrasepsi.</p>

				<p>untuk cluster 2 yang terdiri dari , usia 46 tahun, jumlah anak 3, pendidikan SMA, status kerja tidak kerja. 30% untuk cluster 3 yang terdiri dari , usia 27 tahun, jumlah anak 2, pendidikan SMA, status kerja tidak kerja.</p>		
--	--	--	--	--	--	--

Hasil performa dari jurnal yang pertama menggunakan satu algoritma yaitu Algoritma C4.5 menghasilkan akurasi sebesar 85,38%. Jurnal kedua menggunakan satu algoritma yaitu Algoritma C4.5 menghasilkan akurasi sebesar 92%. Jurnal ketiga

menggunakan satu algoritma yaitu Naive Bayes menghasilkan akurasi sebesar 82,2%. Jurnal keempat menggunakan satu algoritma yaitu algoritma Decision Tree menghasilkan akurasi sebesar 91,58%. Jurnal kelima menggunakan algoritma K-Means menghasilkan akurasi yang cukup rendah sebesar 48%. Untuk menentukan alat kontrasepsi pada jurnal ini tidak ada algoritma pembandingan dalam penentuan alat kontrasepsi yang berpotensi untuk mengetahui adanya algoritma yang lebih baik hasil tingkat akurasinya.

## 2.2 Landasan Teori

### 2.2.1 Kontrasepsi

Kontrasepsi dapat diartikan sebagai usaha untuk mencegah terjadinya kehamilan, usaha tersebut dapat bersifat sementara ataupun permanen. Berdasarkan maksud dan tujuan kontrasepsi, maka yang membutuhkan kontrasepsi adalah pasangan yang aktif melakukan hubungan seks dan kedua-duanya memiliki kesuburan normal namun tidak menghendaki kehamilan. Keluarga berencana (KB) merupakan tindakan membantu individu atau pasangan suami istri untuk menghindari kelahiran yang tidak diinginkan, mendapatkan kelahiran yang memang diinginkan, mengatur interval di antara kelahiran.

KB merupakan proses yang disadari oleh pasangan untuk menentukan jumlah dan jarak anak serta waktu kelahiran. Tujuan dari KB yaitu meningkatkan kesejahteraan ibu dan anak serta mewujudkan keluarga kecil yang bahagia dan sejahtera melalui pengendalian kelahiran dan pengendalian pertumbuhan penduduk

Indonesia. KB diharapkan dapat menghasilkan produk yang berkualitas, sumber daya manusia yang bermutu dan meningkatkan kesejahteraan keluarga.

## Manfaat Program KB

Menurut WHO (2018) manfaat KB adalah sebagai berikut:

1. Mencegah kesehatan terkait kehamilan

Kemampuan wanita untuk memilih untuk hamil dan kapan ingin hamil memiliki dampak langsung pada kesehatan dan kesejahteraannya. KB memungkinkan jarak kehamilan dan penundaan kehamilan pada wanita muda yang memiliki risiko masalah kesehatan dan kematian akibat melahirkan anak usia dini. KB mencegah kehamilan yang tidak diinginkan, termasuk wanita 16 tahun dan wanita yang lebih tua dalam menghadapi peningkatan risiko terkait kehamilan. KB memungkinkan wanita yang ingin membatasi jumlah keluarga yang ingin dimiliki. Penelitian banyak menunjukkan bahwa wanita yang memiliki lebih dari 4 anak berisiko mengalami kematian ibu. Dengan mengurangi tingkat kehamilan yang tidak diinginkan, KB juga mengurangi kebutuhan akan aborsi yang tidak aman.

2. Mengurangi AKB (angka kematian bayi)

KB dapat mencegah kehamilan dan kelahiran yang berjarak terlalu dekat dan tidak tepat waktu. Hal ini berkontribusi pada tingginya angka kematian bayi tertinggi di dunia.

3. Membantu mencegah penularan *Human Immunodeficiency Virus (HIV)* / *Acquired Immunodeficiency Syndrome (AIDS)*

KB mengurangi risiko kehamilan yang tidak diinginkan diantara wanita yang hidup dengan HIV, mengakibatkan lebih sedikit bayi yang terinfeksi dan anak yatim. Selain itu, kondom pria dan wanita memberikan perlindungan ganda terhadap kehamilan yang tidak diinginkan dan terhadap IMS ( Infeksi menular seksual) termasuk HIV.

4. Memberdayakan masyarakat dan meningkatkan pendidikan

KB memungkinkan masyarakat untuk membuat pilihan berdasarkan informasi tentang kesehatan seksual dan reproduksi. KB memberikan peluang bagi perempuan untuk mengejar pendidikan hingga usia diatas 17 tahun sehingga dapat berpartisipasi dalam kehidupan publik, termasuk mendapatkan pekerjaan yang layak. Selain itu, memiliki keluarga yang lebih kecil memungkinkan orang tua untuk berinvestasi lebih banyak pada setiap anak. Karena faktanya anak-anak dengan lebih sedikit saudara kandung cenderung tetap bersekolah lebih lama daripada mereka yang memiliki banyak saudara kandung.

5. Mengurangi kehamilan remaja

Remaja hamil lebih berisiko diantaranya adalah lebih cenderung memiliki bayi prematur atau bayi berat lahir rendah (BBLR). Bayi yang dilahirkan oleh remaja memiliki angka kematian neonatal yang lebih tinggi. Banyak gadis

remaja yang hamil harus meninggalkan sekolah. Hal ini memiliki dampak jangka panjang bagi mereka sebagai individu, keluarga dan komunitas.

#### 6. Perlambatan pertumbuhan penduduk

KB adalah kunci untuk memperlambat lajunya pertumbuhan penduduk yang tidak berkelanjutan dengan dampak negatif yang dihasilkan pada ekonomi, lingkungan dan upaya pembangunan nasional dan regional (WHO, 2018).

### **Syarat-syarat Kontrasepsi**

Adapun syarat-syarat kontrasepsi yaitu:

1. Aman pemakaiannya dan dapat dipercaya
2. Efek samping yang merugikan tidak ada
3. Kerjanya dapat diatur menurut keinginan
4. Tidak mengganggu hubungan persetubuhan
5. Tidak memerlukan bantuan medis atau kontrol selama pemakaian
6. Cara penggunaannya sederhana
7. Harganya murah agar dapat dijangkau oleh masyarakat luas
8. Dapat diterima oleh pasangan suami istri

### **Pelayanan Metode Kontrasepsi**

- Implant

Implant adalah metode kontrasepsi hormonal yang efektif, tidak permanen dan dapat mencegah terjadinya kehamilan antara tiga sampai lima tahun. Kontrasepsi implant ini memiliki cara kerja menghambat terjadinya ovulasi, menyebabkan selaput lendir endometrium tidak siap untuk

menerima pembuahan (nidasi), mengentalkan lendir dan menipiskan lapisan endometrium dengan efektivitas keberhasilan kontrasepsi implant sebesar 97-99% (BKKBN, 2013).

- Alat Kontrasepsi Dalam Rahim (AKDR) / Intra Uterine Devices (IUD)

a. Pengertian

IUD (Intra Uterin Device) atau nama lain adalah AKDR (Alat Kontrasepsi Dalam Rahim) adalah suatu benda kecil yang terbuat dari plastic yang lentur, mempunyai lilitan tembaga atau juga mengandung hormon dan dimasukkan ke dalam rahim melalui vagina dan mempunyai benang (BKKBN, 2014).

Sangat efektif yaitu 0,5-1 kehamilan per 100 perempuan selama satu tahun pertama penggunaan (Arum dan Sujiyati, 2011).

b. Jenis

Saat ini IUD (Intra Uterin Device) yang umum beredar dan digunakan adalah :

1) IUD (Intra Uterin Device) terbentuk dari rangka plastik yang lentur dan pada lengan dan batang IUD (Intra Uterin Device) terdapat tembaga.

2) IUD (Intra Uterin Device) Nova T, terbentuk dari rangka plastik dan tembaga.

Pada ujung lengan bentuk agak melengkung tanpa ada tembaga, tembaga hanya ada di batangnya.

3) IUD (Intra Uterin Device) Mirena, terbentuk dari rangka plastic yang dikelilingi oleh silinderpelepas hormone progesteron yang bisa dipakai oleh ibu menyusui karena tidak menghambat ASI (Mulyani dan Rinawati, 2013).

- Metode Operasi Wanita (MOW)

Kontrasepsi metode operasi wanita (MOW) atau tubektomi atau juga dapat disebut sterilisasi adalah tindakan penutupan terhadap kedua saluran telur sehingga sel telur tidak dapat melewati saluran telur sehingga sel telur tidak bertemu dengan sperma laki-laki sehingga tidak terjadi kehamilan. Sangat efektif (0,5 kehamilan per 100 prempuan selama tahun pertama penggunaan) dan efektif 6-10 minggu setelah operasi (Triyanto dan Indriani, 2019).

- Metode Operasi Pria (MOP)

Pengertian

Metode operasi pria yang dikenal dengan nama vasektomi merupakan operasi kecil yang lebih ringan dari pada sunat/khitanan pada pria. Bekas operasi hanya berupa satu luka di tengah atau luka kecil di kanan kiri kantong zakar (kantung buah pelir) atau scrotum. Vasektomi berguna untuk menghalangi transport spermatozoa (sel mani) di pipa-pipa sel mani pria (saluran mani pria) (Mega dan Wijayanegara, 2017).

- Metode Kalender

Pengertian

Metode kalender atau pantang berkala adalah cara atau metode kontrasepsi sederhana yang dilakukan oleh pasangan suami istri dengan tidak melakukan senggama atau hubungan seksual pada masa subur atau ovulasi. Metode kalender ini merupakan metode keluarga berencana alamiah (KBA) yang paling tua. Pencetus KBA sistem kalender adalah dr. Knaus (ahli kebidanan dari Vienna) dan dr. Ogino (ahli ginekologi dari Jepang). Metode kalender ini berdasarkan pada siklus haid/ menstruasi wanita. Knaus berpendapat bahwa ovulasi terjadi tepat 14 hari sebelum menstruasi, tetapi 12 atau 16 hari sebelum menstruasi berikutnya. Penelitian kedua ahli ini menjadi dasar dari KBA sistem kalender. Metode ini efektif bila dilakukan secara baik dan benar. Dengan penggunaan sistem kalender setiap pasangan dimungkinkan dapat merencanakan setiap kehamilannya. Sebelum menggunakan metode ini, pasangan suami istri harus mengetahui masa subur. Siklus masa subur pada setiap wanita tidaklah sama. Untuk itu perlu pengamatan minimal 6 kali siklus menstruasi.

- Pil

#### Pengertian

Pil adalah salah satu jenis kontrasepsi oral hormonal yang diminum secara rutin setiap hari untuk mencegah kehamilan. Hormon yang terkandung di dalam pil kb, yaitu hormon estrogen dan progesteron, adalah hormon yang sama yang diproduksi oleh tubuh wanita. Meminum pil kb secara teratur akan membantu menstabilkan level kedua hormon di dalam tubuh, dan hal

ini yang akan membantu dalam pencegahan kehamilan. Pil kb berisi zat yang berguna untuk mencegah lepasnya sel telur dari indung telur wanita. Ada 2 macam kemasan pil, yaitu: kemasan berisi 21 pil dan kemasan berisi 28 pil. Sebelum meminum pil kb, kesehatan ibu perlu diperiksa terlebih dahulu. Jika menurut hasil pemeriksaan ibu bias memakai pil kb barulah ibu dapat meminum pil kb.

Jenis pil kb

- 1) Pil kb kombinasi adalah jenis pil kb yang umum ditemui di pasaran. Pil kb jenis ini mengandung dua jenis hormon, yaitu estrogen dan progesteron.
- 2) Pil kb laktasi adalah jenis yang cocok untuk wanita menyusui atau wanita yang mempunyai alergi terhadap hormon estrogen. Pil kb laktasi hanya mengandung satu jenis hormon, yaitu hormon progesterin.

- Suntik KB

Pengertian

Suntik kb merupakan salah satu metode pencegahan kehamilan yang paling banyak digunakan di Indonesia. Secara umum, suntikan kb bekerja untuk mengentalkan lender rahim sehingga sulit untuk ditembus oleh sperma. Selain itu, suntikan kb juga membantu mencegah sel telur menempel di dinding rahim sehingga kehamilan dapat dihindari.

Jenis suntik kb

1. Suntik kb 3 bulan

Suntikan kb ini mengandung hormon Depo *Medroxyprogesterone Acetate* (hormon progestin) 150 mg sesuai dengan namanya, suntikan pertama ini diberikan setiap 3 bulan (12minggu). Suntikan pertama biasanya diberikan 7 hari pertama periode menstruasi, atau 6 minggu setelah melahirkan. Suntikan kb 3 bulanan ada yang dikemas dalam cairan 3 ml atau 1 ml.

## 2. Suntik kb 1 bulan

Suntikan kb ini mengandung kombinasi hormone *Medroxyprogesterone Acetate* (hormon progestin) dan *Estradiol Cypionate* (hormon estrogen). Komposisi hormon dan cara kerja suntikan kb 1 bulan mirip dengan pil kb kombinasi. Suntikan pertama diberikan 7 hari pertama periode menstruasi atau 6 minggu setelah melahirkan bila tidak menyusui.

- Kondom

Pengertian

Kondom adalah alat kontrasepsi keluarga berencana yang terbuat dari karet dan pemakaiannya dilakukan dengan cara disarungkan pada kelamin laki-laki ketika akan bersenggama.

- Metode Amenorea Laktasi (MAL)

Pengertian

Kontrasepsi yang menggunakan ASI eksklusif sebagai alat untuk mencegah kehamilan, yang berarti bayi hanya diberikan ASI tanpa adanya makanan atau minuman tambahan.

Metode ini merupakan alat kontrasepsi alamiah dengan ketentuan tidak digunakan bersama dengan alat kontrasepsi yang lain (Marmi, 2016).

Metode kontrasepsi MAL menggunakan ASI eksklusif untuk menekan proses terjadinya pelepasan sel telur yang matang (matahari, utami and Sugiiharti, 2018).

- Senggama terputus

Pengertian

Metode kb tradisional, dimana laki-laki mengeluarkan alat kelamin (penis) nya dari vagina sebelum mencapai ejakulasi. Disebut juga sebagai koitus interruptus.

### 2.2.2 Data Mining

Data mining merupakan salah satu teknik dalam pengolahan data yang menemukan hubungan dari data yang tidak diketahui oleh pengguna serta menyajikannya kedalam bentuk yang mudah dipahami sehingga dari hubungan data tersebut dapat dijadikan sebagai dasar dalam pengambilan keputusan (Ridwan et al., 2013). Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan yaitu : Deskripsi, Estimasi, Prediksi, Klasifikasi, Clustering, dan Asosiasi. (Muslehatin et al., 2017).

Definisi umum dari data mining itu sendiri adalah proses pencarian pola-pola yang tersembunyi (*hidden patern*) berupa pengetahuan (*knowledge*) yang tidak diketahui sebelumnya dari suatu sekumpulan data yang mana data tersebut dapat berada di dalam database, data warehouse, atau media penyimpanan informasi yang lain.

Data mining dilakukan dengan tool khusus, yang mengeksekusi operasi data mining yang telah didefinisikan berdasarkan model analisis. Data mining merupakan proses analisis terhadap data dengan penekanan menemukan informasi yang tersembunyi pada sejumlah data besar yang disimpan ketika menjalankan bisnis perusahaan. Kemajuan luar biasa yang terus berlanjut dalam bidang data mining didorong oleh beberapa faktor antara lain:

- 1) Pertumbuhan yang cepat dalam kumpulan data.
- 2) Penyimpanan data dalam data warehouse, sehingga seluruh perusahaan memiliki akses ke dalam database yang andal.
- 3) Adanya peningkatan akses data melalui navigasi web dan internet.
- 4) Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi.
- 5) Perkembangan teknologi perangkat lunak untuk data mining (ketersediaan teknologi).
- 6) Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan (Rahmawati & Merlina, 2018)

Secara umum, metode data mining dapat dibagi menjadi dua :

Deskriptif dan prediktif. Deskriptif berarti data mining digunakan untuk mencari pola-pola yang dapat dipahami manusia yang menjelaskan karakteristik data. Sedangkan prediktif berarti data mining digunakan untuk membentuk sebuah model pengetahuan yang akan digunakan untuk melakukan prediksi (Suyanto, 2017)

Metode yang ada dalam data mining adalah sebagai berikut :

#### 1. *Classification*

Klasifikasi merupakan proses untuk menemukan sekumpulan model yang dijelaskan kelas-kelas data, sehingga model tersebut dapat digunakan untuk memprediksi nilai suatu kelas yang belum diketahui pada sebuah objek. Untuk mendapatkan model, kita harus melakukan analisis terhadap data latih. Sedangkan data uji digunakan untuk mengetahui tingkat akurasi dan model yang telah dihasilkan. Klasifikasi dapat digunakan untuk memprediksi nama atau nilai dari suatu objek data.

#### 2. *Clustering*

Pengelompokan data yang tidak diketahui label kelasnya kedalam sejumlah kelompok tertentu sesuai dengan ukuran kemiripannya. Metode inilah yang digunakan dalam tugas akhir ini.

#### 3. *Association*

Tujuan dari metode ini yaitu untuk menghasilkan sejumlah rule yang menjelaskan sejumlah data yang terhubung kuat dengan yang lainnya.

#### 4. *Regression*

Regression mirip dengan klasifikasi. Perbedaan utamanya adalah terletak pada atribut yang diproduksi nilai yang kontinyu.

#### 5. *Forecasting*

Prediksi (forecasting) berfungsi untuk melakukan prediksi kejadian yang akan diproses berdasarkan data sejarah yang ada.

#### 6. *Sequence Analysis*

Tujuan dari metode ini adalah untuk mengenali pola dari data diskrit sebagai contoh adalah menemukan kelompok gen dengan tingkat ekspresi yang mirip.

#### 7. *Deviation Analysis*

Tujuan dari metode ini adalah untuk menemukan penyebab perbedaan antara data yang satu dengan data yang lain dan biasa disebut sebagai outlier detection. Sebagai contoh adalah apakah sudah terjadi penipuan terhadap pengguna kartu kredit dengan melihat catatan transaksi yang tersimpan dalam basis data perusahaan tersebut.

### 2.2.3 Klasifikasi

Klasifikasi data merupakan suatu proses yang menemukan properti-properti yang sama pada sebuah himpunan obyek di dalam sebuah basis data dan mengklasifikasikannya ke dalam kelas-kelas yang berbeda menurut model klasifikasi yang ditetapkan. Tujuan dari klasifikasi adalah untuk menemukan model dari data latih yang akan membedakan atribut ke dalam kategori atau kelas yang sesuai model.(Ente et al., 2020).

Untuk menggunakan metode klasifikasi tentunya harus menerapkan Algoritma dalam Implementasinya. Algoritma yang akan digunakan adalah Decision Tree. Algoritma C4.5 adalah ekstensi Quinlan untuk algoritma ID3 untuk menghasilkan pohon keputusan (Decision Tree), algoritma C4.5 rekursif mengunjungi setiap node keputusan, memilih split optimal sampai tidak ada perpecahan lanjut yang memungkinkan (Larose, 2005 dalam Novandya, 2017).

Klasifikasi adalah salah satu prediksi teknik data mining yang membuat prediksi tentang data nilai menggunakan hasil yang diketahui yang ditemukan dari kumpulan data yang berbeda. Masalah akurasi dari banyak algoritma klasifikasi adalah diketahui mengalami penurunan informasi saat dihadapi dengan data yang tidak seimbang, misalnya ketika distribusi sampel lintas kelas sangat miring (Misdrum, 2021). Dalam klasifikasi, ada variabel kategoris target, seperti braket pendapatan, yang, misalnya, dapat dipartisi menjadi tiga kelas atau kategori: berpenghasilan tinggi, menengah pendapatan, dan pendapatan rendah. Model data mining memeriksa satu set besar catatan, masing-masing catatan yang berisi informasi tentang variabel target serta satu set input atau prediktor variable. Contoh tugas klasifikasi dalam bisnis dan penelitian meliputi: (Larose & Larose, 2014).

- a. Menentukan apakah transaksi kartu kredit tertentu adalah penipuan
- b. Menempatkan siswa baru pada jalur tertentu yang berkaitan dengan kebutuhan khusus
- c. Menilai apakah aplikasi hipotek adalah risiko kredit yang baik atau buruk
- d. Mendiagnosis apakah ada penyakit tertentu

- e. Menentukan apakah surat wasiat ditulis oleh almarhum yang sebenarnya, atau curang oleh orang lain
- f. Mengidentifikasi apakah perilaku keuangan atau pribadi tertentu menunjukkan kemungkinan ancaman teroris

Klasifikasi yang dilakukan secara manual adalah klasifikasi yang dilakukan oleh manusia tanpa adanya bantuan dari algoritma cerdas komputer. Sedangkan klasifikasi yang dilakukan dengan bantuan teknologi, memiliki beberapa algoritma, diantaranya Naïve Bayes, Support Vector Machine, Decision Tree, Fuzzy dan Jaringan Saraf Tiruan (Wibawa, 2018).

#### 2.2.4 Algoritma C4.5

Algoritma C4.5 adalah salah satu algoritma induksi pohon keputusan yaitu *ID3* (*Iterative Dichotomiser 3*). ID3 ini dikembangkan oleh J. Ross Quinlan. Dalam prosedur algoritma ID3, input merupakan sampel training, table training dan atribut. Algoritma C4.5 merupakan pengembangan dari ID3 (Rajesh, 2011).

Algoritma C4.5 merupakan teknik pohon keputusan yang terkenal dan disukai karena memiliki kelebihan-kelebihan, misalnya:

- a. Dapat mengolah data *numeric* (kontinyu) dan diskret
- b. Dapat menangani nilai atribut yang hilang
- c. Bisa menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat di antara algoritma-algoritma yang menggunakan memori utama di komputer.

Algoritma C4.5 mengkonstruksi pohon keputusan dari data pelatihan, yang berupa kasus-kasus atau record-record (tupel) dalam basisdata. Setiap kasus berisikan

nilai dari atribut-atribut untuk sebuah kelas. Setiap atribut dapat berisi data diskret (*numeric*). C4.5 juga menangani kasus yang tidak memiliki nilai sebuah/ lebih atribut. Akan tetapi, atribut hanya bertipe diskret dan tidak boleh kosong.

Tiga prinsip kerja algoritma C4.5 pada tahap belajar dari data yang bisa dikemukakan, yaitu:

1. Pembuatan pohon keputusan

Obyektif dari algoritma pohon keputusan yaitu mengkonstruksi struktur data pohon (dinamakan pohon keputusan) yang dapat digunakan untuk memprediksi kelas dari sebuah kasus atau record baru yang belum memiliki kelas. C4.5 mengkonstruksi pohon keputusan dengan strategi divide dan conquer. Pada awalnya, hanya dibuat node akar dengan menerapkan algoritma divide dan conquer. Algoritma ini memilih pemecahan kasus-kasus yang terbaik dengan menghitung dan membandingkan gain ratio, kemudian pada node-node yang terbentuk di level berikutnya, algoritma divide dan *conquer* akan diterapkan lagi. Demikian pohon keputusan, dengan simbol kotak menyatakan simpul dan elips menyatakan daun.

2. Pemangkasan pohon keputusan dan evaluasi (opsional)

Karena pohon yang dikonstruksi dapat berukuran besar dan tidak mudah “dibaca”, C4.5 dapat menyederhanakan pohon dengan melakukan pemangkasan berdasarkan nilai tingkat kepercayaan (*confidence level*). Selain untuk pengurangan ukuran pohon, pemangkasan juga bertujuan untuk mengurangi tingkat kesalahan prediksi pada kasus (*record*) baru.

### 3. Pembuatan aturan-aturan dari pohon keputusan (opsional)

Aturan dalam bentuk if-then diturunkan dari pohon keputusan dengan melakukan penelusuran dari akar sampai ke daun. Setiap node dan syarat pencabangnya akan diberikan di if, sedangkan nilai pada daun akan menjadi ditulis di then. Setelah semua aturan dibuat, maka aturan akan disederhanakan (digabungkan atau diperumum).

Ada beberapa langkah-langkah untuk membangun pohon keputusan (Kusrini, 2007).

1. Pilih atribut sebagai akar
2. Buat cabang untuk tiap-tiap nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Perhitungan nilai Gain digunakan rumus Persamaan 1.

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

Dan untuk mendapatkan nilai Gain, harus ditentukan terlebih dahulu nilai Entropy:

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

n: Jumlah partisi S

pi: Proporsi dari terhadap S

### 2.2.5 Naïve Bayes

Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema Bayes. Teorema tersebut dikombinasikan dengan “naive” dimana diasumsikan kondisi antar atribut saling bebas. Pada sebuah dataset, setiap baris/dokumen diasumsikan sebagai vector dari nilai-nilai atribut dimana tiap nilai-nilai menjadi peninjauan atribut (Naafian et al., 2016). Pendekatan dari Teorema Naïve Bayes adalah sebagai berikut :

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

Keterangan :  $P(C_i|X)$  : peluang dokumen X pada kategori  $C_i$ .

-  $P(X|C_i)$  : peluang pada kategori  $C_i$ , dimana kata pada dokumen X muncul pada kategori tersebut.

-  $P(C_i)$  : peluang dari kategori yang diberikan, dibandingkan dengan kategori- kategori lainnya yang dianalisa.

-  $P(X)$  : peluang dari dokumen tersebut secara spesifik. Pada pengembangannya,  $P(X)$  dapat dihilangkan karena nilainya tetap, sehingga saat dibandingkan dengan tiap kategori, nilai ini dapat dihapus.

### 2.2.6 Precision, Recall dan Akurasi

Dalam “dunia” pengenalan pola (pattern recognition) dan temu kembali informasi (information retrieval), precision dan recall adalah dua perhitungan yang banyak digunakan untuk mengukur kinerja dari system/ metode yang digunakan. *Precision*

adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem.

*Recall* adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

*Accuracy* didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual. Tahap berikutnya melakukan pengujian terdapat tiga bagian.

1. Akurasi

Didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual.

2. Presisi

Tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem.

3. Recall

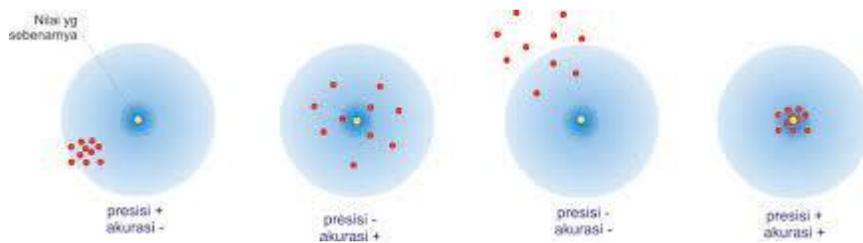
Tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi.

Sedangkan di “dunia lain” seperti dunia statistika dikenal juga istilah *accuracy*.

Akurasi didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual. Ilustrasi berikut ini memberikan gambaran perbedaan antara akurasi dan presisi.

Tabel 2.2 Presisi, akurasi dan recall

Nilai Sebenarnya			
		TRUE	FALSE
Nilai Prediksi	TRUE	TP (True Positive) Corect result	FP (False Positive) Unexpected result
	FALSE	FN (False Negative) Missing result	TN (Trues Negative) Corect result



Gambar 2.1 Perbedaan Akurasi, Presisi dan Recall

Secara umum presisi, recall dan akurasi dapat dirumuskan sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Presisi} = \frac{TP}{TP+FP}$$

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