

## **BAB II**

### **TINJAUAN PUSTAKA**

Dalam bab ini berisi tentang penjelasan mengenai referensi penelitian terdahulu yang menguraikan landasan-landasan teori yang mendukung judul, dan mendasari pembahasan secara detail. Pada bab ini dijelaskan juga mengenai metode, teknik, dan *tools* (komponen) yang akan digunakan dalam pembuatan aplikasi atau tujuan penelitian terkait dengan judul yang dibuat.

#### **2.1 Penelitian Sebelumnya**

Berikut beberapa penelitian terdahulu yang terkait dengan metode maupun sistem yang digunakan peneliti dalam membuat sistem pendukung keputusan.

**1. Herlina W, Jayawardanu, Hansun (2015). “Rancang Bangun Sistem Pakar Untuk Deteksi Dini Katarak Menggunakan Algoritma C4.5”.**

51% dari 39 juta orang mengalami kebutaan disebabkan oleh katarak. Pada 2013, ada 1,8% 1.027.763 orang Indonesia yang menderita katarak. Setengah dari mereka belum dirawat karena ketidaktahuan mereka tentang penyakit katarak. Karena itu, Dalam penelitian ini, kami mencoba membangun sistem yang dapat mendeteksi penyakit katarak secara dini sebagai dokter mata akan dilakukan. Sistem akan menggunakan algoritma C4.5 itu menerima 150 set data pelatihan sebagai input, yang dihasilkan dalam seperangkat aturan yang dapat digunakan sebagai keputusan faktor. Untuk menguji sistem, validasi k-fold cross teknik digunakan dengan k sama dengan 10. Dari Hasil analisis akurasi sistem 93.2% untuk mendeteksi penyakit katarak dan 80,5% untuk mendeteksi jenis penyakit katarak yang mungkin diderita seseorang.

**2. Sugara, Widyatmoko, Prakoso, Saputro (2017). “Penerapan Algoritma C4.5 Untuk Deteksi Dini Gangguan Autisme Pada Anak”.**

Deteksi dini gangguan autisme merupakan hal yang terpenting dalam proses tumbuh kembang anak. Namun, kebanyakan orang tua masih belum memahami

tentang gangguan autisme yang dialami oleh anaknya dan bagaimana cara penanganannya. Para orang tua lebih memilih untuk mencari dan mendatangi seorang dokter/pakar untuk berkonsultasi. Tetapi, ketersediaan dokter ataupun dalam bidang tumbuh kembang anak sekarang tidaklah terlalu banyak sehingga para orang tua harus mengeluarkan biaya yang relatif besar hanya untuk berkonsultasi kepada seorang dokter ataupun pakar. Oleh karena itu diperlukan cara atau metode untuk memudahkan para orang tua dalam mendeteksi dini gangguan autisme yang dialami oleh anaknya. Algoritma *C4.5* merupakan salah satu metode data mining untuk memprediksi kemampuan untuk mengetahui tingkat akurasi deteksi dini gangguan autisme pada anak. Dengan menggunakan aplikasi RipedMiner dan menggunakan metode decision tree *C4.5*, hasil penelitian ini menunjukkan nilai akurasi sebanyak 72%.

**3. Tyasti, Inspriyanti, Hoyyi (2018). “Algoritma *Iterative Ditochomiser 3(ID3)* Untuk Mengidentifikasi Data Rekam Medis(Studi Kasus Penyakit Diabetes Mellitus Di Balai Kesehatan Kementerian Perindustrian, Jakarta)”.**

Algoritma *Iterative Dichotomiser 3 (ID3)* merupakan algoritma pembelajaran pohon keputusan dasar. Ini algoritma melakukan pencarian menyeluruh (serakah) di semua pohon keputusan yang mungkin. Algoritma *ID3* bisa diimplementasikan menggunakan fungsi rekursif, (fungsi yang memanggil dirinya sendiri). Salah satu masalah itu dapat diatasi dengan menggunakan algoritma *ID3* yaitu klasifikasi pasien diabetes. Diabetes adalah penyakit karena tubuh tidak mampu mengontrol jumlah gula atau glukosa dalam tubuh aliran darah. Klasifikasi menggunakan *ID3* pada kasus penderita diabetes menghasilkan pohon dengan banyak simpul hingga 32 simpul dimana 21 di antaranya adalah simpul daun dan atribut puasa glukosa postprandial dua jam terpilih sebagai simpul akar dalam pohon pengambilan keputusan. Berdasarkan kinerja klasifikasi hasil pengukuran menunjukkan akurasi klasifikasi atau akurasi pengukuran mencapai 89,75%. Sedangkan akurasi pengukuran algoritma klasifikasi *ID3* menggunakan total sampel uji 84 sampel menunjukkan akurasi 72.619%.

**4. Meiyanti, Komarudin, (2020). “Klasifikasi Diagnosa Penyakit Paru-paru Pada Klinik Raditya Medical Center Dengan Metode Algoritma C4.5”.**

Dalam pelayanan pemeriksaan penyakit paru-paru pada klinik masih menggunakan cara manual yaitu menganalisis dan mengelompokan gejala-gejala yang dialami oleh pasien di anggap belum efektif. Terlebih belum adanya validasi data dalam pendiagnosaan penyakit paru-paru. Paru-paru merupakan salah satu organ terpenting yang ada pada tubuh manusia, dan yang paling utama mempunyai fungsi sebagai tempat bergantinya suatu oksigen yang mengandung sebuah zat karbondioksida pada darah. Banyak terjadi kelalayan dalam mengatasi salah satu penyakit yang berbahaya ini. Sebagian dari beberapa masyarakat yang kurang memahami dampak negatif dari penyakit paru-paru. Fasilitas dan peyanaan terhadap penderita diharapkan dapat diberikan secara maksimal. Sehingga tidak ada lagi masyarkat sekitar yang tidak mengetahui dampak dan gejala penyakit tersebut. Klasifikasi Diagnosa Penyakit Paru -Paru Dengan Menggunakan Algoritma C4.5 memungkinkan dapat membantu pihak klinik dalam pengelompokan suatu gejala berdasarkan kriteria untuk mendiagnosa penyakit pada paru-paru. Aplikasi yang digunakan untuk mengklasifikasikan penyakit berdasarkan kriteria pemeriksaan klinis penyakit pada paru-paru yaitu Weka 3.9 menggunakan pohon j48.

**5. Setiawati, Wibowo, Hermawan (2019).“Implementasi Decision Tree Untuk Mendiagnosa Penyakit Liver”.**

Hati merupakan salah satu organ manusia yang paling penting. UCI Machine Learning Repository mempunyai banyak dataset, salah satunya adalah dataset ILPD (Indian Liver Patient Dataset). Penelitian ini membahas tentang klasifikasi penyakit liver pada dataset ILPD menggunakan Algoritma *Decision Tree C4.5*. Berdasarkan hasil pengolahan yang dilakukan, didapatkan bahwa Algoritma *Decision Tree C4.5* menghasilkan nilai akurasi sebesar 72.67% dan juga membuktikan bahwa dari 11 variabel penyakit liver yang ada pada dataset ILPD, hanya 2 variabel (Almine Alminotransferase) yang menjadi pokok dalam penentuan penyakit liver.

**6. Srimenganti, taufil, Mulyana (2018).“Implementasi Algoritma Decision Tree Untuk Penyakit Campak”.**

Pusat Kesehatan Masyarakat (Puskesmas) merupakan tempat rawat jalan. Seringkali, Komunitas Puskesmas tidak menyediakan layanan kesehatan yang lengkap. Para pekerja medis – khususnya dokter sering kali merasa kesulitan untuk menentukan penyakit pasien. Salah satu tantangannya adalah mendiagnosis campak. Iterative Dichotomizer (*ID3*) merupakan salah satu algoritma untuk membantu diagnosa penyakit campak dengan cara mendapatkan data. dari beberapa atribut. Atribut algoritme dapat bervariasi bergantung pada kebutuhan pasien. Itu algoritma bekerja dengan menghitung entropi dari setiap atribut dan Information Gain (IG) dari setiap atribut untuk membangun pohon keputusan. Pohon keputusan membantu dokter mendiagnosis campak. Ada 118 Data pada penelitian ini, sebagian besar responden terdiagnosis campak. Uji akurasi *ID3* Aplikasi berbasis algoritma menunjukkan angka 89,83% termasuk kategori sangat baik. Ini menyimpulkan itu Algoritma *ID3* efektif untuk mendiagnosis campak.

Perhitungan Algoritma *ID3* Yang Dilakukan :

**Tabel 2.1 Tabel Kasus Campak**

Nama	Umur	Jk	Panas	Rash	Sesak nafas	Pilek	Igm
Kabir	Muda	L	Tinggi	Ya	Pernah	Sakit	Positif
Bagas	Tua	L	Sedang	Tidak	Tidak Pernah	Sakit	Negatif
Siti	Muda	P	Tinggi	Ya	Pernah	Tidak Sakit	Positif
Rani	Muda	P	Tinggi	Ya	Pernah	Tidak Sakit	Positif
Edi	Tua	L	Sedang	Tidak	Pernah	Sakit	negatif

**1. Pembuatan node**

Pembuatan tabel ini untuk menentukan nilai terbesar dari setiap kelas dan dijadikan akar dalam setiap keputusan, dengan cara menghitung jumlah kasus sesuai dengan jumlah kasus yang sudah ditentukan. Kemudian setelah itu, menghitung dan mencari nilai *Entropi* setiap atribut dan nilai *Gain* untuk setiap kelas. Sehingga dapat ditemukan nilai terbesar dan menjadi akar (*root*) pohon. Dapat dilihat pada Tabel 2.2

**Tabel 2.2 Perhitungan Entropy Status Igm Campak**

Hasil	Jumlah Kasus	$P_j \cdot \log_2 P_j$
Igm Positif	86	0,33
Igm Negatif	32	0,50
	Entropy (Total)	0,83

$$\text{Entropi (Total)} = (-86/118 \cdot \log_2(86/118)) + (-(32/118) \cdot \log_2(32/118)) = 0,83$$

## 2. Menentukan Gain

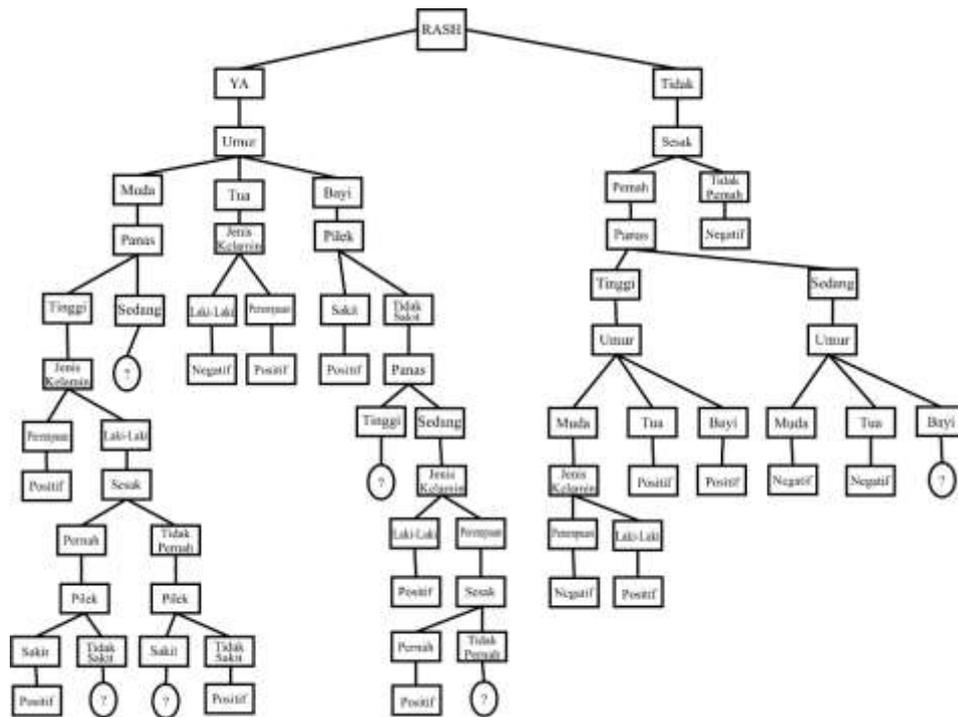
$$\text{Gain (Total Umur)} = 0,8432 - ((46/188 \times 0,950) + 6/188 \times 0,918) + 66/188 \times 0,65)) = 0,63$$

**Tabel 2.3 Perhitungan Gain**

Node	Atribut	Value	Jumlah Kasus	Campak	Tidak Campak	Entropy	Gain
1	Umur						0.063
		Muda	46	29	17	0.9503	
		Tua	6	2	4	0.918	
		Bayi	66	55	11	0.65	
	JK						0.0064
		L	61	42	19	0.894	
		P	57	44	13	0.774	
	Panas						0.0594
		Tinggi	68	42	11	0.638	
		Sedang	50	44	21	0.981	
	Sesak						0.022
		Pernah	53	43	10	0.698	
		Tidak Pernah	65	43	22	0.923	
	Pilek						0.009
		Sakit	43	34	9	0.740	
		Tidak Sakit	75	52	23	0.889	

## 3. Membuat pohon keputusan

Pada Gambar 5 Dapat dilihat dan diuraikan pohon keputusan dengan hasil nilai *Gain* tertinggi yang akan dijadikan akar (*root*) sehingga dapat membuat pohon keputusan yang dapat menemukan keputusan akhir sebuah kasus.



**Gambar 2.1** Pohon Keputusan

**7. Aziz, Karpen (2018).“Diagnosa Penyakit Kulit Wajah Menggunakan Metode Decision Tree Dan Algoritma C4.5”.**

Wajah merupakan hal yang sangat diperhatikan baik oleh perempuan maupun laki-laki. Hal ini disebabkan karena wajah merupakan hal pertama yang dilihat ketika bertemu dengan seseorang. Namun demikian ada yang sering mengganggu pada wajah, salah satunya adalah penyakit kulit yang sangat beragam mulai dari jerawat, kusam, komedo hingga kanker. Pada kenyataannya untuk mengatasi hal ini, maka banyak orang selalu berkonsultasi dengan dokter terutama masalah wajah. Salah satu yang bisa dilakukan adalah dengan melakukan diagnose penyakit kulit wajah menggunakan metode decision tree dan algoritma C4.5. Metode decision tree digunakan untuk menemukan hubungan antara calon sejumlah variabel (data mining), sehingga menjadi sebuah variabel target dengan mengklasifikasinya dalam bentuk pohon keputusan. Hasil dari klasifikasi akan dimasukkan dalam algoritma C4.5 untuk dilakukan pemilihan variabel sebagai akar dan cabang, dimana setiap cabang akan diberikan nilai. Variabel menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon keputusan. Selanjutnya

pohon keputusan adalah mengubah bentuk data menjadi model rule (pedoman) dan menyederhanakan dalam bentuk sistem pakar. Penelitian ini menggunakan sejumlah variabel gejala penyakit kulit, seperti bintik-bintik dan benjolan yang ada pada kulit. Adanya sistem diagnosa penyakit kulit ini bisa dijadikan pedoman dalam melakukan perawatan wajah, sehingga bisa terhindar dari penyakit kulit khususnya pada wajah.

Perhitungan Algoritma C45 Yang Dilakukan :

**Tabel 2.4 Tabel kasus penyakit kulit wajah**

No	Gejala	Efek	tampilan	usia	Penyakit
1	Benjolan	sakit disentuh	merah	16	Jerawat
2	bintik-bintik	muka berminyak, kusam	putih	18	Jerawat
3	Benjolan	muka berminyak, kusam	hitam	43	Skin Tag
4	Kemerahan di kulit wajah	sakit disentuh	merah	29	Rosacea
5	bintik-bintik	sakit disentuh	hitam	16	Skin Tag
6	bintik-bintik	sakit disentuh	merah	10	Jerawat

### 1. Diagnosa penyakit kulit wajah

Dari rumus yang sudah ditentukan sebelumnya pada algoritma C 4.5, dalam diagnosis ini usia penderita sengaja diambil berdasarkan data-data yang ada, yaitu usia yang bervariasi antara umur 10 tahun sampai dengan usia 43 tahun, maka dapat dilakukan perhitungan dan dihasilkan tabel decision tree sebagai berikut:

Proses perhitungan nilai gain dan entropy:

$$Entropy [Total] = (-8/15) * (\log_2 (8/15)) + (-2/15) * (\log_2 (2/15)) + (-3/15) * (\log_2 (3/15)) + (-2/15) * (\log_2 (2/15)) = 1.723231$$

$$Entropy [gejala - bintik bintik] = (-2/5) * (\log_2 (2/5)) + (-2/5) * (\log_2 (2/5)) + (-1/5) * (\log_2 (1/5)) + (-0/5) * (\log_2 (0/5)) = \text{Tak Terhingga, Apabila hasil tidak ditemukan maka diganti dengan 0}$$

$$Entropy [gejala - benjolan] = (-5/7) * (\log_2 (5/7)) + (-0/7) * (\log_2 (0/7)) + (-2/7) * (\log_2 (2/7)) + (-0/7) * (\log_2 (0/7)) = \text{Tak Terhingga, Apabila hasil tidak ditemukan maka diganti dengan 0}$$

$$Entropy [gejala - kemerahan diwajah] = (-1/3) * (\log_2 (1/3)) + (-0/3) * (\log_2 (0/3)) + (-0/3) * (\log_2 (0/3)) + (-2/3) * (\log_2 (2/3)) = \text{Tak Terhingga, Apabila hasil}$$

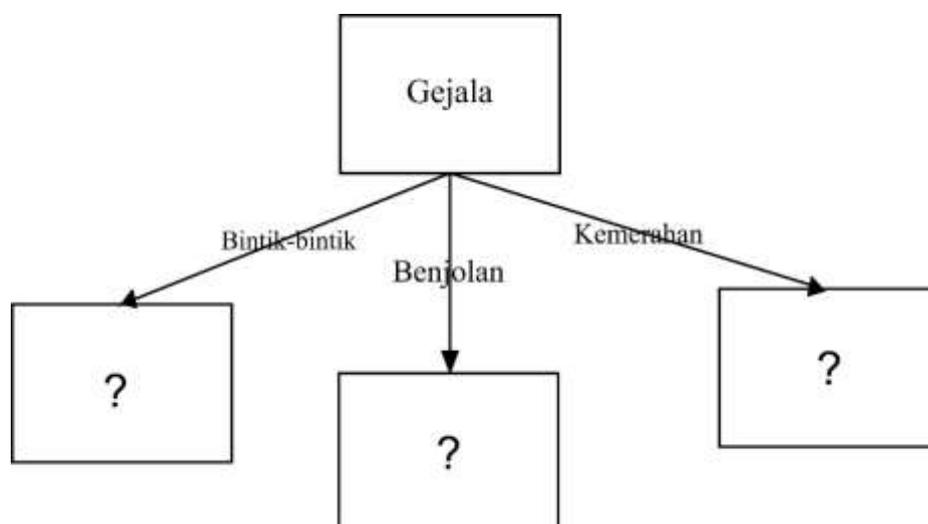
tidak ditemukan maka diganti dengan 0

Gain [Pelatih] =  $1.723231 - ((5/15) * 0) + ((7/15) * 0) + ((3/15) * 0) = 1,723231$   
dan seterusnya.

**Tabel 2.5 Tabel Decision Tree dari perhitungan algoritma C4.5**

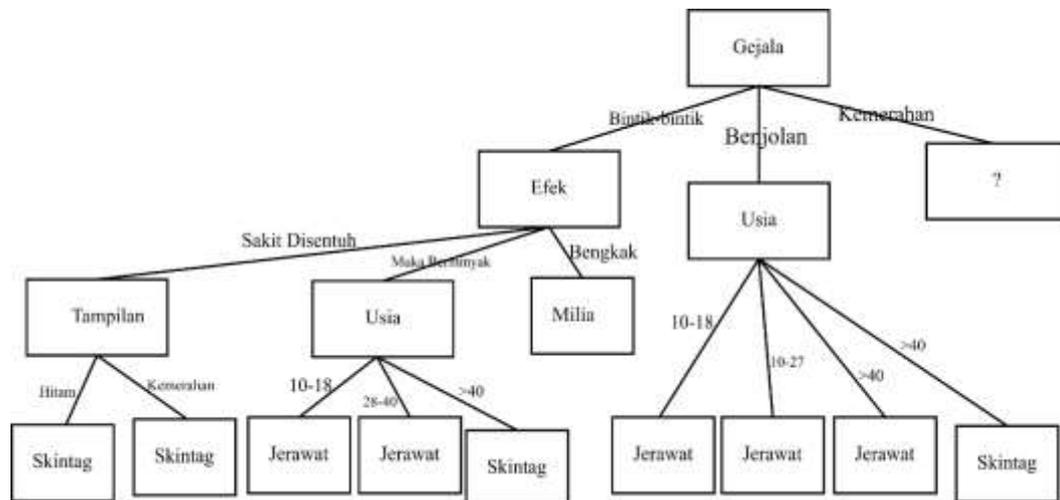
		jumlah kasus	Jerawat	Milia	Skin Tag	Rosacea	Entropy	Gain
Total		15	8	2	3	2	1,723231	
Gejala								1,723231
	bintik-bintik	5	2	2	1	0	0	
	Benjolan	7	5	0	2	0	0	
Efek								1,189
	sakit disentuh	6	4	0	1	1	0	
Tampilan								1,7232
	Warna hitam	4	1	0	2	1	0	
	warna putih	3	1	1	1	0	0	
	Warna merah	8	6	1	0	1	0	
usia penderita								1,7232
	10 sd 18	3	2	0	1	0	0	
	>40	2	0	0	2	0	0	

Dari hasil tabel tersebut, selanjutnya dapat dilakukan analisis untuk kemudian dimasukkan ke dalam algoritma *Decision tree C4.5* yang dapat digambarkan sebagai berikut:



**Gambar 2.2 Decision Tree dari data hasil perhitungan algoritma C4.5**

Dari gambar yang di atas belum ditemukan gejala-gejala penyakit wajah yang ditemukan, untuk itu harus dilakukan perhitungan selanjutnya. Pencarian berikutnya dilakukan pada hasil tree yang masih kosong atau memiliki kemungkinan penyakit yang lebih dari satu maka dilakukan pencarian pada Efek Muka berminyak.



**Gambar 2.3 Hasil Algoritma 4.5 dengan Decision Tree**

Dari data gejala sudah semua ditemukan dan pohon keputusan yang akan dijadikan basis pengetahuan pada sistem pakar sudah ditemukan maka pencarian selesai. Berdasarkan hasil akhir dari pohon keputusan dapat di berikan contoh mengenai proses diagnosa sistem pakar. Berikut contohnya : Apabila di ketahui gejala yang di timbulkan adalah :G01 (bintik-bintik), G05 (muka berminyak), G10 (usia 10-18).

**8. Pribadi, Athiry, Saputra, Supiandi, Prayudi (2018). “Sistem Pakar Diagnosa Penyakit Demam Berdarah Dengue Menggunakan Algoritma Iterative Dichotomiser 3 (ID3)”.**

Penyakit demam berdarah merupakan salah satu penyakit menular yang sering menimbulkan wabah dan menyebabkan kematian yang disebabkan oleh virus dengue dan ditularkan melalui gigitan nyamuk *Aedes aegypti* dan *Aedes albopictus*. Kurangnya pengetahuan masyarakat terhadap penyakit DBD serta keterbatasan penanganan penyakit DBD seringkali terlambat didiagnosa. Oleh sebab itu perlu adanya tindakan/penanganan secara dini untuk mencegah masyarakat terjangkit

penyakit DBD. *ID3* termasuk algoritma *decision tree learning* (algoritma pembelajaran pohon) yang paling dasar yang melakukan pencarian secara menyeluruh pada semua kemungkinan pohon keputusan. Dalam penelitian ini akan dilakukan analisa data penyakit demam berdarah *dengue* menggunakan klasifikasi data mining yakni algoritma *ID3* agar tingkat akurasi lebih diskrit. Berdasarkan uraian tersebut, dibutuhkan sebuah sistem yang dapat mewakili seorang pakar yang memiliki basis pengetahuan dan pengalaman tentang penyakit DBD, yaitu sebuah sistem pakar. Agar mendapatkan nilai informasi yang lebih cepat dan akurat, sistem pakar ini akan diaplikasikan berbasis *website*. Dari 198 jumlah kasus yang terdiri dari 103 kasus pasien yang terjangkit penyakit DBD dan 95 kasus pasien yang tidak terjangkit penyakit DBD yang didapat dari UPTD Puskesmas Sukaraja, maka didapatkan 12 *rule* yang dihasilkan dari pohon keputusan algoritma *ID3* dengan jumlah *class* tidak sebanyak 7 *rule* dan jumlah *class* ya sebanyak 5 *rule*, dan diperoleh tingkat akurasi sebesar 75,253% dan dengan kesalahan memprediksi (*error rate*) sebesar 24,747% sehingga dapat disimpulkan bahwa penelitian yang diimplementasikan berbasis *website* ini cukup akurat dan dapat membantu para pengguna dan masyarakat dalam mendiagnosa penyakit demam berdarah dengue (DBD).

**9. Yusuf, Wibowo (2020). “Sistem Pakar Diagnosa Penyakit Mata Pada Manusia Menggunakan Algoritma C4.5”.**

Mata merupakan panca indra yang sangat penting dalam kehidupan manusia. Dari mata kita bisa melihat isi dunia. Akan tetapi mata juga seperti organ tubuh lain yang bisa terkena penyakit dan fungsinya mengalami penurunan. Penyakit mata adalah penyakit yang berbahaya, bila tidak segera didiagnosa dan diobati maka dapat menimbulkan kebutaan pada mata. Maka dari itu dibutuhkan sebuah sistem dalam bidang kesehatan. Adapun tujuan yang akan dicapai adalah memberi kemudahan gejala yang dipilih dengan menerapkan algoritma *C4.5*. Algoritma *C4.5* merupakan algoritma klasifikasi pohon keputusan yang banyak digunakan karena memiliki kelebihan utama yaitu dapat menghasilkan pohon keputusan yang mudah diterapkan, memiliki tingkat akurasi yang dapat diterima, efisien dalam menangani

atribut bertipe diskret dan numerik. Berdasarkan permasalahan diatas, maka penulis mengambil judul “Sistem Pakar Diagnosa Penyakit Mata pada Manusia Menggunakan Algoritma C4.5”

**10. Abdillah (2020). “Penerapan Algoritma Decision Tree C4.5 Untuk Diagnosa Penyakit Stroke Dengan Klasifikasi Data Mining Pada Rumah Sakit Santa Maria Pematang”.**

Data Mining adalah proses ekstraksi sebelumnya tidak dikenal dan dipahami dari database berukuran besar dan digunakan untuk membuat keputusan bisnis yang penting. Studi kasus yang digunakan dan diterapkan dalam tugas akhir ini adalah data pasien penyakit saraf khususnya penyakit Stroke untuk dikelola menggunakan algoritma C4.5 dengan metode klasifikasi data mining.

Stroke termasuk penyakit pembuluh darah otak ditandai dengan kematian jaringan otak yang terjadi karena berkurangnya aliran darah dan oksigen ke otak. Salah satu cara untuk mempelajari Stroke yaitu dengan ilmu data mining tepatnya menggunakan algoritma C4.5, hasil laporan ini menentukan pasien penyakit stroke dengan variabel yang diketahui kemudian diolah menggunakan data mining algoritma C4.5.

**11. Mauladi (2017). “Sistem Penerimaan Pegawai Di Fif Pt Astra Kabupaten Lamongan Dengan Metode Fuzzy Logic Tsukamoto”.**

Permasalahan yang terdapat dalam suatu perusahaan yaitu sistem rekrutmen. Rekrutmen karyawan sistem yang digunakan untuk mendapatkan karyawan yang sesuai dengan karakteristik perusahaan ASTRA OTOPARTS berlokasi di Lamongan. Oleh karena itu, komputer merupakan salah satu alat untuk memenuhi kebutuhan perusahaan dalam mempercepat proses penentuan kandidat yang tepat untuk menjadi karyawan layanan pelanggan atau sebagai ahli sehingga akan memudahkan pelatih untuk menginformasikan jumlah calon yang lulus sertifikat dan memperoleh sertifikat pelatihan untuk manusia pengembangan sumber daya (SDM). Metode *fuzzy tsukamoto* sebagai pengambilan keputusan pelatihan outcome, yang akan memberikan solusi kepada trainer untuk mengatasi

permasalahan yang ada dengan hasil akurasi data sebesar 95% dan menghilangkan proses rekrutmen secara manual dimana ditentukan oleh manajer dan dapat meminimalisir kecurangan dalam rekrutmen di FIF ASTRA OTOPART di Kabupaten Lamongan.

**12. Fuad (2019). “Pemanfaatan Algoritma Fuzzy Naive Bayes Dalam Pemilihan Bidang Keahlian Mahasiswa Teknik Informatika Universitas Islam Lamongan”.**

Salah satu faktor yang dapat menentukan kelancaran mahasiswa dalam mengikuti perkuliahan dengan baik adalah pemilihan bidang minat yang sesuai dengan kemampuan siswa, artinya jika siswa menentukan pilihan jurusan sesuai dengan bidang keahliannya maka dipastikan tidak akan ada kesulitan untuk mengembangkan diri dalam proses perkuliahan. penelitian ini dengan memanfaatkan dua algoritma yaitu *Fuzzy* dan *Naive Bayes* untuk menentukan luas daerah keahlian siswa. Mahasiswa Teknik Informatika memiliki kesempatan untuk memilih bidang keahlian, yaitu: bidang keahlian jaringan, pemrograman, dan desain. Tidak semua siswa akan menguasai bidang keahlian, sehingga mereka harus memilih sesuai dengan keterampilan, hobi dan mata pelajaran favorit, dan nilai mereka. Penelitian ini akan membuat sistem klasifikasi keahlian mahasiswa teknik informatika UNISLA. Algoritmanya adalah diimplementasikan yaitu *Fuzzy* untuk mengkategorikan nilai atau input angka, sedangkan *Naive Bayes* digunakan untuk klasifikasi.

## **2.2 Landasan Teori**

Pada bagian ini dipaparkan kajian teori yang berkaitan dengan masalah yang dibahas, menjelaskan konsep-konsep dasar yang digunakan dalam penelitian ini.

### **2.2.1 Sistem Pakar**

Menurut Yuniarto, (2013:2) Sistem pakar adalah suatu sistem yang mengadopsi sebuah pengetahuan seorang pakar atau ahli yang selanjutnya di implementasikan kedalam sebuah sistem komputer, dengan tujuan sistem tersebut dapat menyelesaikan sebuah permasalahan seperti biasa yang di lakukan oleh para

pakar atau ahli. Seperti contohnya seorang dokter, dokter merupakan seorang yang ahli dalam mendiagnosis sebuah penyakit pada pasien dan bisa membantu memberikan solusi terhadap penyakit tersebut.

Sistem Pakar dapat membantu menyelesaikan sebuah permasalahan yang hanya bisa diselesaikan oleh sang pakar atau ahlinya saja, dikatakan dapat menyelesaikan masalah atau mampu mengambil sebuah keputusanlayaknya yang para pakar atau ahli lakukan maka sang pakar dari saat memproses dalam mengambil keputusan ataupun hasil dari sebuah keputusan.

Sistem pakar sebenarnya punya komponen yang utama contohnya dasar mesin inferensi dan pengetahuan. Dasar dari pengetahuan sebagai tempat menyimpan pemahaman dalam memori yang diambil dari ilmu para pakar.

#### **2.2.1.1. Ciri-Ciri Sistem Pakar**

Menurut Yunianto, (2013:3) Sistem pakar memiliki ciri-ciri sebagai berikut:

1. Pengetahuan sistem pakar merupakan sebuah konsep bukan numerik.
2. Dapat memberikan penalaran untuk data yang tidak pasti.
3. Dapat mengemukakan rangkaian alasan-alasan yang diberikannya dengan cara yang dapat dipahami.
4. Berdasarkan pada kaidah/ketentuan/ *rule* tertentu.
5. Dirancang untuk dapat dikembangkan secara bertahap.
6. Pengetahuan & mekanisme penalaran (*inference*) jelas terpisah.
7. Keluarannya bersifat anjuran.
8. Sistem dapat mengaktifkan kaidah secara searah yang sesuai dituntun oleh dialog dengan user.

Secara garis besar, banyak keuntungan yang dapat diperoleh dengan mengembangkan sistem pakar, antara lain :

1. Membuat seorang yang awam bekerja seperti layaknya seorang pakar.
2. Meningkatkan produktivitas akibat meningkatnya kualitas hasil pekerjaan, disebabkan meningkatnya efisiensi kerja.
3. Menghemat waktu kerja.
4. Menyederhanakan pekerjaan.

5. Merupakan arsip yang terpercaya dari sebuah keahlian.
6. Memperluas jangkauan, dari keahlian secara pakar.

### 2.2.2. *Coronavirus Disease(COVID-19)*

*Coronavirus* merupakan satu anggota keluarga besar virus yang mengjangkiti manusia dan hewan. Pada manusia penyakit ini biasanya menyebabkan infeksi pada saluran pernapasan, mulai dari flu biasa hingga penyakit yang serius seperti Middle East Respiratory Syndrome (MERS) dan Sindrom Pernafasan Akut Berat/ Severe Acute Respiratory Syndrome (SARS). Pada Desember 2019 ditemukan salah satu penyakit jenis baru yang disebabkan oleh *Coronavirus*, kemudian diberi nama *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-COV2), dan menyebabkan penyakit *Coronavirus Disease-2019* (COVID-19) (KEMENKES RI, 2020 : 10).

#### 2.2.2.1 **Penyebab Infeksi *Coronavirus Disease (COVID-19)***

Menurut MENKES, (2020:12-13) Infeksi coronavirus disebabkan oleh virus corona itu sendiri. Virus corona menyebar seperti virus lain pada umumnya, seperti:

1. Percikan air liur penderita (batuk dan bersin).
2. Menyentuh tangan atau wajah penderita.
3. Menyentuh mata, hidung, atau mulut setelah memegang barang yang terkena percikan air liur penderita coronavirus.
4. Feses atau kotoran, tapi ini biasanya jarang terjadi.

#### 2.2.2.2 **Gejala Infeksi *Coronavirus Disease (COVID-19)***

Menurut Sugihantono, (2020: / 12) *Coronavirus* bisa menimbulkan beragam gejala pada penderitanya seperti :

**Tabel 2.6 Tabel Gejala *Coronavirus***

NO	Nama Gejala
1.	Demam lebih dari 38°C
2.	Batuk
3.	Pilek
4.	Sakit pada tenggorokan

**Tabel 2.6 Tabel Lanjutan Gejala *Coronavirus***

NO	Nama Gejala
5.	Sesak nafas
6.	Menggigil
7.	Sakit kepala

Sumber : Kemenkes RI

### 2.2.3 Algoritma *Iterative Ditochomiser 3(ID3)*

*Iterative Dichotomizes 3 (ID3)* dikembangkan oleh J. Ross Quinlan. Algoritma ini merupakan pembelajaran pohon keputusan yang paling dasar. *ID3* melakukan pencarian secara rakus/menyeluruh pada semua kemungkinan pohon keputusan. Salah satu algoritma induksi pohon keputusan yaitu *ID3* (*Iterative Dichotomizes 3*). Algoritma *ID3* dapat diimplementasikan menggunakan fungsi rekursif (fungsi yang memanggil dirinya sendiri). Algoritma *ID3* berusaha membangun decision tree (pohon keputusan) dari atas ke bawah, mulai dengan atribut mana yang pertama kali harus dicek dan diletakkan pada root, ini bisa di selesaikan dengan mengevaluasi semua atribut yang ada dengan menggunakan suatu ukuran statistic (yang banyak digunakan adalah information gain) untuk mengukur efektivitas suatu atribut dalam mengklasifikasikan kumpulan sampel data(Safii,2018:82-86).

Secara ringkas, cara kerja Algoritma *ID3* dapat digambarkan sebagai berikut. Pemilihan atribut dengan menggunakan Information Gain

- a. Pilih atribut dimana nilai gainnya terbesar.
- b. Buat simpul yang berisi atribut tersebut.
- c. Proses perhitungan information gain akan terus dilaksanakan sampai semua data telah termasuk dalam kelas yang sama. Atribut yang telah dipilih tidak diikutkan lagi dalam perhitungan nilai information gain.

Menurut Hikmatullah, (2008:120) Pemilihan atribut pada *ID3* dilakukan menggunakan properti statistik, yang disebut dengan information gain. Gain mengukur seberapa baik suatu atribut memisahkan training example ke dalam kelas target. Atribut dengan informasi tertinggi akan dipilih. Dengan tujuan untuk mendefinisikan gain, pertama-tama digunakanlah ide dari teori

informasi yang disebut entropi. Entropi mengukur jumlah dari informasi yang ada pada atribut dengan rumus :

$$Entropi(S) = \sum_{j=1}^k - p_j \log_2 p_j \dots \dots \dots \text{Rumus 2.1}$$

Keterangan:

$S$  adalah himpunan (dataset) kasus

$k$  adalah banyaknya partisi  $S$

$p_j$  adalah probabilitas yang di dapat dari Sum(Ya) dibagi Total Kasus.

Setelah mendapat nilai entropi, pemilihan atribut dilakukan dengan nilai information gain terbesar.

Entropi(S) = 0, jika semua contoh pada  $S$  berada dalam kelas yang sama.

Entropi(S) = 1, jika jumlah contoh positif dan jumlah contoh negatif dalam  $S$  adalah sama.

$0 < \text{Entropi}(S) < 1$ , jika jumlah contoh positif dan negatif dalam  $S$  tidak sama.

*Information Gain* merupakan perubahan dari *entropy* yang sudah dibagi atributnya pada sebuah dataset, menjadi subset terkecil. Perhitungan information gain merupakan selisih antara entropy dataset sebelum dan sesudah pembagian. Pembagian terbaik akan menghasilkan entropy yang paling kecil sehingga berdampak pada information gain yang terbesar. Pemilihan atribut sebagai root berdasarkan nilai gain tertinggi dari atribut yang ada. Perhitungan information gain menggunakan persamaan dibawah ini

$$\text{Gain}(A) = \text{Entropy}(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times \text{Entropi}(S_i) \dots \dots \dots \text{Rumus 2.2}$$

Keterangan:

$S$  = ruang (data) sample yang digunakan untuk training.

$A$  = atribut.

$|S_i|$  = jumlah sample untuk nilai  $V$ .

$|S|$  = jumlah seluruh sample data.

Entropi( $S_i$ ) = entropy untuk sample-sample yang memiliki nilai.

### 2.2.4 Algoritma C4.5

Algoritma C4.5 adalah program yang memberi kontribusi satu set data berlabel dan menghasilkan pohon keputusan sebagai keluaran. Pohon keputusan tindak lanjut ini kemudian diverifikasi terhadap data uji berlabel yang tidak terlihat untuk menghitung generalisasinya. C4.5 adalah program yang digunakan untuk menghasilkan peraturan taksonomi dengan menggunakan pohon keputusan dari sekumpulan data yang diberikan. Algoritma C4.5 merupakan perpanjangan dari algoritma ID3 dasar dan dirancang oleh Quinlan. C4.5 adalah salah satu algoritma pembelajaran yang banyak digunakan. Algoritma C4.5 membangun pohon keputusan dari serangkaian data pelatihan yang mirip dengan Algoritma ID3, dengan menggunakan konsep entropi informasi. C4.5 juga dikenal sebagai klasifikasi statistik. (Manikantan dan Latha, 2013:2).

Algoritma ini dapat menyelesaikan masalah secara sistematis dengan membentuk suatu pohon keputusan dengan langkah-langkah sebagai berikut:

1. Pilih atribut sebagai akar.
2. Buat cabang untuk masing-masing record dari atribut.
3. Membagi kasus ke dalam cabang,
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang menghasilkan suatu keputusan yang sesuai.

Dalam memilih sebuah atribut menjadi akar, dilakukan perhitungan nilai dari atribut yang ada. Nilai gain yang paling tinggi dijadikan root pada pohon keputusan. Untuk menghitung nilai gain digunakan rumus:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \dots \dots \dots \text{Rumus 2.3}$$

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi S

pi : proporsi dari Si terhadap S

1. Kemudian hitung nilai *gain* yang menggunakan rumus:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots \dots \dots \text{Rumus 2.4}$$

Keterangan:

$S$  = himpunan kasus

$A$  = fitur

$n$  = jumlah partisi atribut  $A$

$|S_i|$  = proporsi  $S_i$  terhadap  $S$

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

2. Ulangi langkah ke-2 hingga semua record terpartisi.
3. Proses partisi pohon keputusan akan berhenti saat:
  - a. Semua record dalam simpul  $N$  mendapat kelas yang sama.
  - b. Tidak ada atribut di dalam record yang dipartisi lagi.
  - c. Tidak ada record di dalam cabang yang kosong.