

BAB II

LANDASAN TEORI

2.1 Diabetes Mellitus

Diabetes mellitus adalah adalah penyakit yang disebabkan oleh tingginya kadar gula dalam darah akibat gangguan sekresi insulin. Diabetes mellitus di sebut juga penyakit kencing manis. dalam bahasa Yunani memiliki arti tembus atau pancuran air, dan dari bahasa latin memiliki arti rasa manis, sedang di Indonesia diebetes mellitus lebih dikenal dengan penyakit kencing manis, di mana kadar glukosa (gula sederhana) di dalam darah menjadi tinggi karena tubuh tidak dapat memproduksi atau mengeluarkan insulin secara cukup. Dan dari beberapa tes secara langsung, pada umumnya air seni pengidap diabetes rasanya manis karena mengandung banyak gula.

Tanda awal yang dapat diketahui bahwa seseorang menderita diabetes mellitus atau kencing manis yaitu dilihat langsung dari efek peningkatan kadar gula darah, dimana peningkatan kadar gula dalam darah mencapai nilai 160 - 180 mg/dL dan air seni (urine) penderita kencing manis yang mengandung gula (glucose), sehingga urine sering dilebung atau dikerubuti semut .Penderita kencing manis umumnya menampakkan tanda dan gejala dibawah ini meskipun tidak semua dialami oleh penderita :

1. Jumlah urine yang dikeluarkan lebih banyak (Polyuria)
2. Sering atau cepat merasa haus/dahaga (Polydipsia)
3. Lapar yang berlebihan atau makan banyak (Polyphagia)
4. Frekwensi urine meningkat/kencing terus (Glycosuria)
5. Kehilangan berat badan yang tidak jelas sebabnya
6. Kesemutan/mati rasa pada ujung syaraf ditelapak tangan & kaki
7. Cepat lelah dan lemah setiap waktu
8. Mengalami rabun penglihatan secara tiba-tiba
9. Apabila luka/tergores (korengan) lambat penyembuhannya
10. Mudah terkena infeksi terutama pada kulit.

2.2 Data Mining

Data Mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Patut diingat bahwa kata mining sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu Data Mining sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan database. Data mining adalah proses menerapkan metode ini untuk data dengan maksud untuk mengungkap pola-pola tersembunyi. Dengan arti lain Data mining adalah proses untuk penggalian pola-pola dari data. Data mining menjadi alat yang semakin penting untuk mengubah data tersebut menjadi informasi.

2.2.1 Pengelompokan Data Mining

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Daniel T. Larose, 2005):

1. Deskripsi

Deskripsi adalah menggambarkan pola dan kecenderungan yang terdapat dalam data secara sederhana. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Klasifikasi

Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Klasifikasi menggunakan *supervised learning*.

3. Estimasi

Estimasi hampir sama dengan klasifikasi, perbedaannya adalah variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model

dibangun dengan menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi.

4. Prediksi

Prediksi memiliki kesamaan dengan klasifikasi dan estimasi, perbedaannya adalah hasil dari prediksi akan ada dimasa mendatang. Beberapa teknik yang digunakan dalam klasifikasi dan estimasi dapat juga digunakan (untuk keadaan yang tepat) untuk prediksi.

5. Klustering

Klustering merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain. Klustering menggunakan *unsupervised learning*.

6. Asosiasi

Tugas asosiasi atau sering disebut juga sebagai *market basket analysis* dalam data mining adalah menemukan relasi atau korelasi diantara himpunan item-item dan menemukan atribut yang muncul dalam satu waktu. Asosiasi menggunakan *unsupervised learning*. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, *support* dan *confidence*.

Metode yang akan digunakan pada penelitian ini termasuk kedalam kelompok prediksi, karena menggunakan teknik klasifikasi yang hasilnya akan ada dimasa mendatang.

2.2.2 Knowledge Discovery in Databases (KDD)

Knowledge Discovery in Databases (KDD) adalah keseluruhan proses untuk mengkonversi data mentah menjadi suatu pengetahuan yang bermanfaat. Istilah data mining dan *Knowledge Discovery in Databases* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar.

Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Salah satu tahapan dalam keseluruhan proses KDD adalah data mining.

Proses KDD secara garis besar dapat dijelaskan sebagai berikut (Fayyad, 1996).

1. Data Selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. Pre-processing/Cleaning

Sebelum proses data mining, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak.

3. Transformation

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. Data mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

5. Interpretation Evaluation

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*.

Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.3 Metode Data Mining Klasifikasi

Klasifikasi adalah proses penemuan pola atau fungsi yang menjelaskan dan membedakan konsep atau kelas data dengan tujuan untuk dapat memprediksi kelas dari suatu objek yang labelnya tidak diketahui. Konsep klasifikasi dengan pengawasan (*supervised classification*) adalah untuk membangun sebuah model dari data yang telah diketahui, atau sering disebut sebagai *classifier*. Model atau fungsi ini kemudian dapat digunakan untuk memetakan data didalam suatu basis data kepada suatu atribut target, selanjutnya dapat memperkirakan suatu kelas dari data yang baru. Tiap record berisi banyak atribut dimana masing-masing atribut memiliki satu dari beberapa kemungkinan nilai. Di dalam klasifikasi diberikan sejumlah record yang dinamakan sekumpulan data latih yang terdiri dari beberapa atribut, dimana salah satu atribut menunjukkan kelas untuk record.

2.4 Decision Tree (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. Selain itu dapat diekspresikan dalam bentuk bahasa basis data seperti *Structure Query Language* untuk mencari *record* pada kategori tertentu (Kusrini dan Emha, 2009).

Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan variabel target. Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan

serangkaian aturan keputusan, dengan masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain.

2.4.1 Model Decision Tree

Decision tree adalah *flow-chart* seperti *struktur tree*, dimana tiap *internal node* menunjukkan sebuah test pada sebuah atribut, tiap cabang menunjukkan hasil dari test, dan *leaf node* menunjukkan *class-class* atau *class distribution*.

Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun mudah untuk dipahami. Pada *decision tree* terdapat 3 jenis *node*, yaitu:

- a. *Root Node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
- b. *Internal Node*, merupakan *node* percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.
- c. *Leaf node* atau *terminal node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.

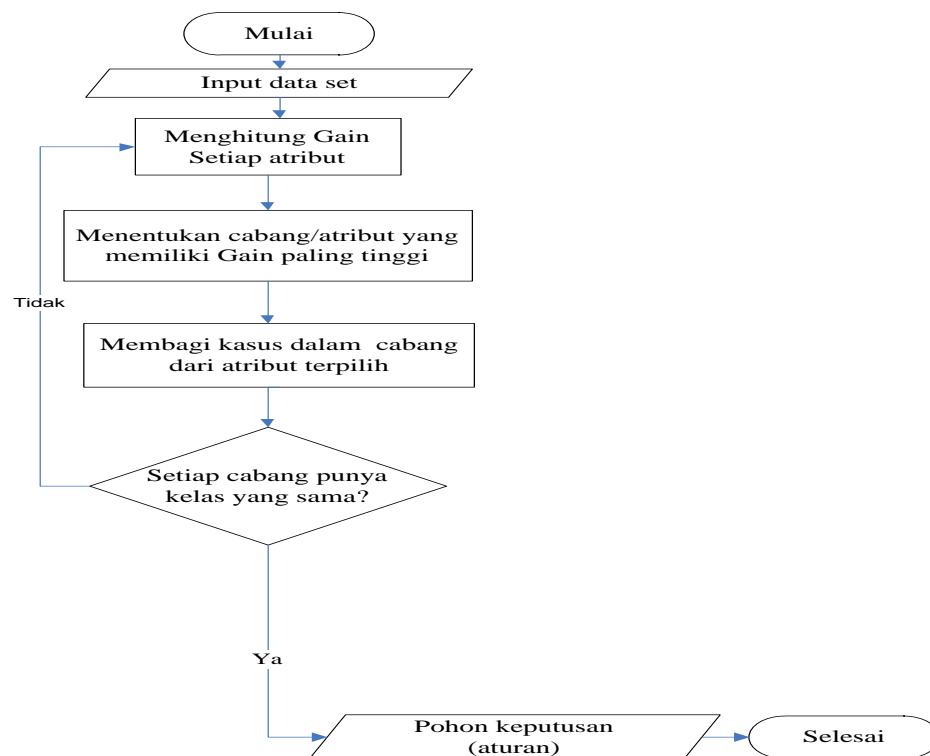
2.4.2 Decision Tree ID3

ID3 singkatan dari Iterative Dichotomiser Three. Ada juga yang menyebut Induction of Decision Tree. ID3 adalah suatu algoritma matematika yang digunakan untuk menghasilkan suatu pohon keputusan yang mampu mengklasifikasi suatu obyek. ID3 diperkenalkan pertama kali oleh Ross Quinlan (1979). ID3 merepresentasi konsep-konsep dalam bentuk pohon keputusan. Aturan-aturan yang dihasilkan oleh ID3 mempunyai relasi yang hirarkis seperti suatu pohon (mempunyai akar, titik, cabang, dan daun). Beberapa peneliti menyebut struktur model yang dihasilkan ID3 sebagai pohon keputusan (*decision tree*) sementara peneliti yang lain menyebutnya pohon aturan (*rule tree*).

2.4.3 Algoritma

Secara umum algoritma id3 untuk membangun pohon keputusan adalah sebagai berikut (F.Romansyah, 2007) :

1. Di mulai dari node akar.
 2. Untuk semua fitur, hitung nilai entropy untuk semua sampel (data latih) pada node.
 3. Pilih fitur dengan informasi gain yang maksimal
 4. Gunakan fitur tersebut sebagai node pemecahan sebagai cabang.
 5. Lakukan secara rekursif pada setiap cabang yang dibuat dengan mengulangi langkah 2 sampai 4 hingga semua data dalam setiap node hanya memberikan satu label kelas. Node yang tidak dapat di pecah lagi merupakan daun yang berisi keputusan (label kelas).
1. Berikut ini akan dijelaskan secara lebih detail algoritma ID3 menggunakan *flowcart* yang disajikan pada **gambar 2.1**



Gambar 2.1 Flowchart algoritma Decision Tree ID3

Untuk memilih atribut sebagai simpul akar (*root node*) atau simpul dalam (*internal node*), didasarkan pada nilai *information gain* tertinggi dari atribut-atribut yang ada. Sebelum perhitungan *information gain*, akan dilakukan perhitungan *entropy*. *Entropy* merupakan distribusi probabilitas dalam teori informasi dan diadopsi kedalam algoritma *ID3* untuk mengukur tingkat homogenitas distribusi kelas dari sebuah himpunan data (*data set*). Semakin tinggi tingkat *entropy* dari sebuah data maka semakin homogen distribusi kelas pada data tersebut. Perhitungan *information gain* menggunakan rumus 2.2, sedangkan *entropy* menggunakan rumus 2.3.

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

dimana,

S : Himpunan kasus

A : Atribut

n : Jumlah partisi atribut A

|S_i| : Jumlah kasus pada partisi ke i

|S| : Jumlah kasus dalam S

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

dimana,

S : Himpunan kasus

A : Fitur

n : Jumlah partisi S

p_i : Proporsi dari S_i terhadap S

2.4.4 Contoh Perhitungan

Berikut ini akan dijelaskan ilustrasi dari alur proses perhitungan algoritma *Decision Tree ID3*. Data set yang digunakan pada contoh ini adalah data untuk menentukan bermain atau tidak bermain dengan beberapa atribut yaitu atribut cuaca, suhu, kelembaban, angin. Dimana semua atribut bertipe kategorikal. Sedangkan kolom bermain adalah kelas tujuannya atau label kelas-nya.

Tabel 2.1 Contoh data set

Cuaca	Suhu	Kelembaban	Angin	Bermain
Cerah	Panas	Tinggi	Pelan	Tidak
Cerah	Panas	Tinggi	Kencang	Tidak
Mendung	Panas	Tinggi	Pelan	Ya
Hujan	Lembut	Tinggi	Pelan	Ya
Hujan	Dingin	Normal	Pelan	Ya
Hujan	Dingin	Normal	Kencang	Tidak
Mendung	Dingin	Normal	Kencang	Ya
Cerah	Lembut	Tinggi	Pelan	Tidak
Cerah	Dingin	Normal	Pelan	Ya
Hujan	Lembut	Normal	Pelan	Ya
Cerah	Lembut	Normal	Kencang	Ya
Mendung	Lembut	Tinggi	Kencang	Ya
Mendung	Panas	Normal	Pelan	Ya
Hujan	Lembut	Tinggi	Kencang	Tidak

Pada contoh ini rumus yang digunakan untuk memilih atribut sebagai *node* adalah rumus *information gain*. Proses pertama adalah menghitung *entropy* untuk semua data.

Jumlah ya = 9

Jumlah tidak = 5

Berikut adalah perhitungan *entropy* untuk semua data:

$$\begin{aligned}
 Entropy(S) &= -\frac{9}{14} * \log_2\left(\frac{9}{14}\right) - \frac{5}{14} * \log_2\left(\frac{5}{14}\right) \\
 &= 0.940
 \end{aligned}$$

Data Cuaca :

Cerah = 5

Mendung = 4

Hujan = 5

Berdasarkan data cuaca, maka nilai *information gain* untuk atribut *outlook* adalah sebagai berikut:

$$\begin{aligned} \text{Gain}(\text{outlook}) &= 0.940 - \left(\frac{5}{14} * \left(-\frac{2}{5} * \log_2 \left(\frac{2}{5} \right) - \frac{3}{5} * \log_2 \left(\frac{3}{5} \right) \right) \right. \\ &\quad + \frac{4}{14} * \left(-\frac{4}{4} * \log_2 \left(\frac{4}{4} \right) - \frac{0}{4} * \log_2 \left(\frac{0}{4} \right) \right) \\ &\quad \left. + \frac{5}{14} * \left(-\frac{3}{5} * \log_2 \left(\frac{3}{5} \right) - \frac{2}{5} * \log_2 \left(\frac{2}{5} \right) \right) \right) \\ &= 0.940 - 0.694 \end{aligned}$$

Selanjutnya dihitung entropy untuk setiap nilai fitur terhadap kelas, untuk entropy nilai dalam Cuaca, didapat:

$$E(\text{Semua}_{\text{cerah}}) = - \left(\begin{array}{l} (p(\text{ya}|\text{cerah}) \times \log_2 p(\text{ya}|\text{cerah})) \\ + (p(\text{tidak}|\text{cerah}) \times \log_2 p(\text{tidak}|\text{cerah})) \end{array} \right)$$

$$\begin{aligned} &= -(((2/5) \times \log_2(2/5)) + ((3/5) \times \log_2((3/5))) \\ &= 0.9403 \end{aligned}$$

$$E(\text{Semua}_{\text{mendung}}) = - \left(\begin{array}{l} (p(\text{ya}|\text{mendung}) \times \log_2 p(\text{ya}|\text{mendung})) \\ + (p(\text{tidak}|\text{mendung}) \times \log_2 p(\text{tidak}|\text{mendung})) \end{array} \right)$$

$$= -(((4/4) \times \log_2(4/4)) + ((0/4) \times \log_2((0/4))))$$

$$= 0$$

$$E(\text{Semua}_{\text{hujan}}) = - \left(\begin{array}{l} (p(\text{ya}|\text{hujan}) \times \log_2 p(\text{ya}|\text{hujan})) \\ + (p(\text{tidak}|\text{hujan}) \times \log_2 p(\text{tidak}|\text{hujan})) \end{array} \right)$$

$$= -(((3/5) \times \log_2(3/5)) + ((2/5) \times \log_2((2/5))))$$

$$= 0.9710$$

Entropy tersebut hitung pada setiap fitur. Selanjutnya dihitung Gain untuk setiap fitur. Dari perhitungan, didapat Gain seperti dibawah ini.

$$\begin{aligned}
 G(\text{semua}, \text{cuaca}) &= E(\text{Semua}_{\text{hujan}}) = \sum_{i=1}^n p(v_i | \text{semua}) \times E(\text{semua}_{\text{cuaca}}) \\
 &= E(\text{semua}) \\
 &\quad - \left((p(\text{cerah} | \text{semua}) \times E(\text{semua}_{\text{cerah}})) + (p(\text{mendung} | \text{semua}) \times E(\text{semua}_{\text{mendung}})) \right) \\
 &\quad \quad + (p(\text{hujan} | \text{semua}) \times E(\text{semua}_{\text{hujan}})) \\
 &= 0.9403 - ((5/14) \times 0.9710) + ((4/14) \times 0) + ((5/14) \times 0.9710) \\
 &= 0.2467
 \end{aligned}$$

$$G(\text{semua}, \text{suhu}) = 0.0292$$

$$G(\text{semua}, \text{kelembaban}) = 0.0292$$

$$G(\text{semua}, \text{angin}) = 0.0292$$

Jadi, nilai gain yang tertinggi adalah atribut cuaca maka dilakukan percabangan pada atribut cuaca.

Tabel 2.2 Pembagian atribut cuaca cerah

Cuaca	Suhu	Kelembaban	Angin	Bermain
Cerah	Panas	Tinggi	Pelan	Tidak
Cerah	Panas	Tinggi	Kencang	Tidak
Cerah	Lembut	Tinggi	Pelan	Tidak
Cerah	Dingin	Normal	Pelan	Ya
Cerah	Lembut	Normal	Kencang	Ya

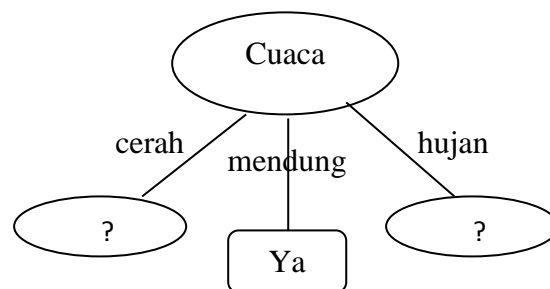
Tabel 2.3 Pembagian atribut mendung

Cuaca	Suhu	Kelembaban	Angin	Bermain
Mendung	Panas	Tinggi	Pelan	Ya
Mendung	Dingin	Normal	Kencang	Ya
Mendung	Lembut	Tinggi	Kencang	Ya
Mendung	Panas	Normal	Pelan	Ya

Tabel 2.4 Pembagian atribut hujan

Cuaca	Suhu	Kelembaban	Angin	Bermain
Hujan	Lembut	Tinggi	Pelan	Ya
Hujan	Dingin	Normal	Pelan	Ya
Hujan	Dingin	Normal	Kencang	Tidak
Hujan	Lembut	Normal	Pelan	Ya
Hujan	Lembut	Tinggi	Kencang	Tidak

Dari hasil data diatas dapat di lihat pohon keputusan pada **gambar 2.2**

**Gambar 2.2** Contoh pohon keputusan

Dari gambar diatas dapat disimpulkan cerah dan hujan belum bisa diambil keputusan karena data masih heterogen (masih ada yang beda) maka dari itu harus menghitung nilai gain kembali untuk menentukan atribut yang dijadikan pemecah cabang selanjutnya. Proses ini berlangsung sampai semua kondisi mempunyai hasil keputusan.

2.5 Penelitian Sebelumnya

Penelitian sebelumnya yang menggunakan metode pohon keputusan ID3 adalah penelitian yang berjudul “Penentuan Jenis Fumigasi Dengan Menggunakan Metode Decision Tree ID3”. Di buat oleh Irina Amalia Nastiti. Institut Teknologi Sepuluh Nopember (ITS Tahun 2011), Tujuan penelitian ini adalah untuk membantu fumigator agar dapat menentukan treatment fumigasi yang tepat, Pembuatan aplikasi ini, ada delapan atribut yang harus diperhatikan yaitu Jenis Komoditi, Negara tujuan, Shipper, Volume, Alat Angkut, Bahan Pengepak, Dosis, dan Tempat Stuffing. Metode decision tree ID3 mampu menyelesaikan permasalahan penentuan treatment fumigasi terbukti dengan tingkat akurasi yang dihasilkan cukup

baik yaitu tingkat akurasi 83.2% yang dilakukan pada 50% data training dan 50% data testing.

Penelitian yang lain yang berjudul “FUZZY DECISION TREE DENGAN ALGORITMA ITERATIVE DICHOTOMISER 3 (ID3) UNTUK KLASIFIKASI GANGGUAN HATI”. Di buat oleh Kingkin Bangkit Rinarto, Universitas Brawijaya Malang Tahun 2011. Untuk mengklasifikasikan gangguan hati yang mungkin timbul akibat konsumsi alkohol berlebih berdasarkan hasil pemeriksaan darah. Data yang digunakan adalah dataset Liver Disorder dari UCI Machine Learning yang berjumlah 345 data dengan 6 atribut parameter dan kelas klasifikasi. Hasil Penelitian ini adalah tingkat akurasi tertinggi yang diperoleh sebesar 64,44% dengan nilai FCT 70% sampai 90%, dan nilai LDT 3% sampai 15%. Jumlah aturan terbanyak yang dihasilkan adalah 61 aturan pada nilai FCT sebesar 90% dengan nilai LDT sebesar 3%.

Ada lagi penelitian yang berjudul “Penerapan Metode Decision Tree Dengan Menggunakan Algoritma ID3 Untuk Pembuatan Sistem Kinerja Guru”. Yang di buat oleh Lilik Shofiyatin, Universitas Islam Negeri Malang tahun 2011. Tujuan penelitian ini adalah untuk pembuatan sistem penilaian kinerja guru. Hasil penelitian ini akurasi sebesar 85% untuk data pelatihan sebanyak 40 dan 65% untuk data pelatihan sebanyak 30.

Kemudian penelitian dari F.Romansyah, I.S. Sitanggang, S. Nurditi. dengan judul “*Fuzzy Decision tree* dengan algoritme ID3 pada data diabetes. Tujuan penelitian ini adalah penerapkan salah satu teknik klasifikasi yaitu *Fuzzy ID*, menemukan aturan klasifikasi pada data diabetes yang membedakan kelas-kelas atau konsep sehingga dapat digunakan untuk memprediksi penyakit diabetes berdasarkan nilai dari atribut lain yang diketahui. Rata-rata akurasi yang didapat 94%.

Yang terakhir penelitian dari mahasiswa Universitas Muhammadiyah Gresik yaitu tugas akhir Azhuardhi Azhro dengan judul “Aplikasi Perhitungan Tingkat Resiko Penyakit Diabetes Dengan Metode Sugeno”. perancangan dan pembuatan aplikasi penentuan tingkat resiko penyakit

diabetes dengan menggunakan metode fuzzy Sugeno, serta pengujian sistem hasil penelitiannya, maka dapat diambil kesimpulan sebagai berikut:

- a. Aplikasi yang dibangun dapat menghasilkan penentuan tingkat resiko penyakit diabetes dengan kriteria umur, index masa tubuh, keturunan diabetes, intensitas olahraga sehingga pasien dapat mengetahui dengan cepat dan mudah untuk tingkat resiko penyakit dalam dirinya.
- b. Dari pengujian hasil aplikasi didapatkan dari data yang tidak cocok yaitu ada 2 orang dari 20 data uji dan presentase hasil perhitungan tingkat kesesuaian 90%.
- c. Metode Fuzzy Sugeno memiliki kekurangan pada penentuan kuisisioner, jika data kuisisioner berubah maka hasil yang didapatkan pada proses penentuan tingkat resiko penyakit diabetes juga berubah, dan semakin banyak uji responden yang dilakukan maka tingkat kevalidan data lebih tinggi.