

**KLASIFIKASI PENYAKIT DIABETES MELLITUS
MENGUNAKAN ALGORITMA DECISION TREE
C4.5**

TUGAS AKHIR

Disusun oleh:
Agam Juli Nur Susanto
3311501081

Disusun untuk memenuhi salah satu syarat kelulusan Program Diploma III



**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
POLITEKNIK NEGERI BATAM
BATAM
2018**

HALAMAN PENGESAHAN

**KLASIFIKASI PENYAKIT DIABETES MELLITUS MENGGUNAKAN
ALGORITMA DECISION TREE C4.5**

**Disusun oleh:
Agam Juli Nur Susanto
3311501081**

Telah diuji dan dipertahankan di depan Tim Penguji
dalam Sidang Tugas Akhir
pada tanggal 21 Agustus 2018
dan dinyatakan **LULUS**.

Batam, 12 September 2018

Disetujui dan disahkan oleh:

Pembimbing,

Mira Chandra Kirana, S.T., M.T.
NIP. 197905302014042002

Tim Penguji,

Ketua,

Anggota,

Yeni Rokhayati, S.Si., M.Sc.
NIP. 198602192014042001

Rina Yulius, S.Pd., M.Eng.
NIK. 118199

HALAMAN PERNYATAAN

Dengan ini, saya:

NIM : Agam Juli Nur Susanto

Nama : 3311501081

adalah mahasiswa Program Studi Teknik Informatika Politeknik Negeri Batam menyatakan bahwa Tugas Akhir dengan judul:

KLASIFIKASI PENYAKIT DIABETES MELLITUS MENGGUNAKAN ALGORITMA DECISION TREE C4.5

disusun dengan:

1. tidak melakukan plagiat terhadap naskah karya orang lain
2. tidak melakukan pemalsuan data
3. tidak menggunakan karya orang lain tanpa menyebut sumber asli atau tanpa ijin pemilik

Jika kemudian terbukti terjadi pelanggaran terhadap pernyataan di atas, maka saya bersedia menerima sanksi apapun termasuk pencabutan gelar akademik.

Lembar pernyataan ini juga memberikan hak kepada Politeknik Negeri Batam untuk mempergunakan, mendistribusikan ataupun memproduksi ulang seluruh hasil Tugas Akhir ini.

Batam, 16 Agustus 2018

Agam Juli Nur Susanto
3311501081

ABSTRAK

KLASIFIKASI PENYAKIT DIABETES MELLITUS MENGGUNAKAN ALGORITMA DECISION TREE C4.5

Berdasarkan Data Riset Kesehatan Dasar (Riskesdas) Kementerian Kesehatan Republik Indonesia, tahun 2013 Indonesia memiliki jumlah penderita penyakit diabetes mellitus mencapai angka 9,1 juta jiwa yang merupakan jumlah ke-empat terbanyak di Asia dan ke-tujuh di dunia, jumlah tersebut diprediksi akan semakin terus bertambah. Banyaknya masyarakat tidak mengetahui bahwa mereka menderita penyakit diabetes mellitus, penderita diabetes mellitus baru mengetahui diri mereka mengidap diabetes mellitus setelah mengalami komplikasi di berbagai organ tubuh. Oleh karena itu, diperlukan suatu sistem yang dapat menentukan seseorang apakah menderita diabetes atau tidak. Pada penelitian ini, dibuat suatu sistem untuk mendiagnosa penyakit diabetes militus dengan cara mengumpulkan data dari hasil rekam medik dan laboratorium rumah sakit BP Batam. Data tersebut akan diklasifikasikan menggunakan algoritma C4.5. Kemudian dari data tersebut akan dibuat pohon keputusan untuk menentukan diagnosa penyakit diabetes mellitus. Penelitian ini menghasilkan suatu sistem yang dapat memberikan informasi kepada masyarakat tentang diagnosa awal penyakit diabetes sehingga dapat membantu masyarakat melakukan tindakan awal atau pencegahan terhadap penyakit diabetes militus.

Kata Kunci : Algoritma C4.5, Diabetes Melitus, Pohon Keputusan

ABSTRACT

CLASSIFICATION OF DIABETES MELLITUS DISEASE USING DECISION TREE C4.5 ALGORITHM

Based on Basic Health Research Data (Riskesdas) Ministry of Health of the Republic of Indonesia, in 2013 Indonesia has the number of people with diabetes mellitus reaches 9.1 million people which is the fourth largest number in Asia and the seventh in the world, the number is predicted to continue to grow. Many people do not know that they have diabetes mellitus, people with diabetes mellitus just know themselves have diabetes mellitus after experiencing complications in various organs of the body. Therefore, a system is needed to determine whether a person has diabetes or not. In this research, a system to diagnose diabetes mellitus by collecting data from medical record and laboratory hospital of BP Batam. The data will be classified using C4.5 algorithm. Then from the data will be made a decision tree to determine the diagnosis of diabetes mellitus. This research resulted in a system that can provide information to the public about early diagnosis of diabetes so it can help the community take early action or prevention against diabetes mellitus disease.

Keywords : C4.5 Algorithm, Decision Tree, Diabetes Melitus.

KATA PENGANTAR

Puji syukur penulis ucapkan atas rahmat Tuhan Yang Maha Esa, karena dengan rahmat, hidayah dan nikmat-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul ” KLASIFIKASI PENYAKIT DIABETES MELLITUS MENGGUNAKAN ALGORITMA DECISION TREE C4.5 ”. Tugas Akhir ini disusun untuk memenuhi mata kuliah Workshop III pada semester 6 program studi Teknik Informatika Politeknik Negeri Batam.

Dalam penyusunan laporan Tugas Akhir ini penulis banyak mendapat bimbingan, penjelasan, dukungan, bantuan dan saran dari berbagai pihak baik secara langsung maupun tidak langsung terlibat dalam proses penyelesaiannya, sehingga dari hati yang terdalam penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Ibu Mira Chandra Kirana, S.T., M.T. selaku dosen pembimbing di Politeknik Negeri Batam
2. Ibu Nur Zahрати Janah, S.Kom., M.Sc selaku dosen wali di Politeknik Negeri Batam
3. Kedua orangtua yang telah memberikan dukungan, motivasi, dan doa selama proses pengerjaan Tugas Akhir ini.
4. Teman-teman IF 2015 yang telah membantu dan memberikan motivasi selama proses pengerjaan Tugas Akhir ini.

Dalam penyusunan laporan Tugas Akhir ini, penulis menyadari masih terdapat banyak kekurangan, karena masih terbatasnya ilmu pengetahuan yang dimiliki oleh penulis. Untuk itu saran dan kritik yang membangun dibutuhkan untuk menyempurnakan laporan Tugas Akhir ini. Penulis juga mengharapkan agar laporan Tugas Akhir ini dapat bermanfaat dan menambah pengetahuan bagi pembaca.

Batam, Agustus 2018

Penulis

DAFTAR ISI

HALAMAN PERNYATAAN.....	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	ix
DAFTAR TABEL	x
BAB I PENDAHULUAN.....	11
1.1 Latar Belakang.....	11
1.2 Rumusan Masalah.....	12
1.3 Batasan Masalah	12
1.4 Tujuan	12
1.5 Manfaat	12
1.6 Tinjauan Pustaka.....	13
1.7 Sistematika Penulisan	16
BAB II LANDASAN TEORI	17
2.1 Diabetes Mellitus	17
2.2 Glukosa Darah	17
2.3 Data Mining	19
2.4 <i>Preprocessing Data</i>	21
2.5 Algoritma <i>Decision Tree C4.5</i>	23
2.6 <i>Confusion Matrix</i>	25
2.7 MATLAB.....	26
BAB III ANALISIS DAN PERANCANGAN.....	28
3.1 Desain Penelitian	28
3.2 Lokasi dan Waktu Penelitian	31
3.3 Teknik Pengumpulan Data.....	31
3.4 Teknik Pengolahan dan Analisis Data	31

3.5 Alat dan Bahan.....	32
3.6 Penyajian Data	33
3.7 Deskripsi Umum Sistem	34
3.8 Kebutuhan Fungsional	35
3.9 Kebutuhan Non Fungsional	35
3.10 Use Case Diagram	35
3.11 Skenario Use Case	36
3.11.1 Training	36
3.11.2 Testing	37
3.11.3 Memeriksa Pohon Keputusan.....	37
3.11.4 Prediksi	38
3.12 Sequence Diagram	39
3.12.1 Training	39
3.12.2 Testing	39
3.12.3 Memeriksa Pohon Keputusan.....	40
3.12.4 Prediksi	40
3.13 Class Diagram.....	41
3.14 Perancangan Antarmuka	42
BAB IV HASIL DAN PEMBAHASAN	44
4.1 <i>Preprocessing Data</i>	44
4.2 Pengoperasian Sistem	45
4.2.1 Training	45
4.2.2 Testing	48
4.2.3 Memeriksa Pohon Keputusan.....	50
4.2.4 Prediksi	51
4.3 Pengujian	52
BAB V KESIMPULAN DAN SARAN	54
5.1 Kesimpulan	54
5.2 Saran	54
DAFTAR PUSTAKA	55
BIODATA PENULIS.....	56

DAFTAR GAMBAR

Gambar 1 Tahap Penelitian secara keseluruhan.....	28
Gambar 2 Tahap <i>Preprocessing Data</i>	29
Gambar 3 Tahap Pembuatan Aplikasi.....	30
Gambar 4 Deskripsi Umum Sistem.....	34
Gambar 5 Use Case Diagram.....	35
Gambar 6 Sequence Diagram (Training).....	39
Gambar 7 Sequence Diagram (Testing).....	39
Gambar 8 Sequence Diagram (Memeriksa Pohon Keputusan).....	40
Gambar 9 Sequence Diagram (Prediksi).....	40
Gambar 10 Class Diagram.....	41
Gambar 11 Halaman Utama.....	42
Gambar 12 Halaman Pengujian dan Prediksi.....	43
Gambar 13 Pengoperasian Sistem (Memasukkan Data Training).....	45
Gambar 14 Pengoperasian Sistem (Memproses Data Training).....	46
Gambar 15 Pengoperasian Sistem (Menampilkan Grafik Distribusi Data).....	47
Gambar 16 Pengoperasian Sistem (Pengujian Data Testing).....	48
Gambar 17 Pengoperasian Sistem (Pengujian Data Training).....	49
Gambar 18 Pengoperasian Sistem (Memeriksa Pohon Keputusan).....	50
Gambar 19 Pengoperasian Sistem (Prediksi).....	51

DAFTAR TABEL

Tabel 1 Perbandingan dengan Penelitian Sebelumnya	15
Tabel 2 Model <i>Confusion Matrix</i>	25
Tabel 3 Penyajian Data Asli.....	34
Tabel 4 Skenario Use Case (Training).....	36
Tabel 5 Skenario Use Case (Testing).....	37
Tabel 6 Skenario Use Case (Memeriksa pohon keputusan).....	37
Tabel 7 Skenario Use Case (Prediksi).....	38
Tabel 8 Data Training	44
Tabel 9 Data Testing	45
Tabel 10 Pengujian.....	52

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi saat ini sudah memasuki banyak bidang, salah satunya di bidang kesehatan. Penerapan teknologi informasi di bidang kesehatan memiliki peran penting dengan keanekaragaman bidang ilmunya seperti *artificial intelligence*, *data mining*, *database*, statistika, pemodelan matematika, pengolahan citra, dan sebagainya. Perkembangan teknologi informasi juga dapat mempermudah masyarakat untuk memperoleh informasi dengan cepat dan akurat.

Diabetes Mellitus adalah penyakit gangguan metabolik akibat pankreas tidak memproduksi cukup insulin atau tubuh tidak dapat menggunakan insulin yang diproduksi secara efektif. Klasifikasi Diabetes Mellitus menurut American Diabetes Association yaitu Diabetes Mellitus tipe 1, Diabetes Mellitus tipe 2, Diabetes Mellitus spesifik lainnya, dan Diabetes Mellitus Kehamilan. Faktor yang banyak mempengaruhi seseorang menderita penyakit diabetes mellitus adalah penambahan usia, nilai indeks massa tubuh, tingginya nilai kadar gula darah. Berdasarkan Data Riset Kesehatan Dasar (Riskesdas) Kementerian Kesehatan Republik Indonesia, tahun 2013 Indonesia memiliki jumlah penderita penyakit diabetes mellitus mencapai angka 9,1 juta jiwa yang merupakan jumlah ke-empat terbanyak di Asia dan ke-tujuh di dunia, jumlah tersebut diprediksi akan semakin terus bertambah. Peningkatan jumlah penderita diabetes mellitus disebabkan karena kurangnya pengetahuan masyarakat terhadap gejala penyakit diabetes mellitus, cara menangani penyakit diabetes mellitus serta keterlambatan penegakan diagnosis penyakit diabetes mellitus.

Berdasarkan masalah tersebut dibuatlah suatu sistem yang dapat menentukan klasifikasi penyakit diabetes mellitus dengan menggunakan algoritma C4.5. Sistem ini dibuat sebagai solusi alternatif bagi masyarakat untuk mendiagnosa awal penyakit diabetes mellitus sebelum dilakukan tindakan selanjutnya.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah diatas, serta melihat metode pengembangan perangkat lunak, maka penulis mencoba membuat perumusan masalah sebagai berikut :

1. Bagaimana melakukan klasifikasi penyakit diabetes mellitus?
2. Bagaimana menentukan akurasi dari klasifikasi yang dibangun?

1.3 Batasan Masalah

Pada Tugas Akhir ini terdapat batasan masalah untuk memfokuskan pengolahan data yaitu :

1. Data penderita penyakit diabetes mellitus yang dapat diolah adalah data yang didapatkan dari hasil rekam medik dan laboratorium rumah sakit BP Batam
2. Sistem yang dibangun hanya dapat mendiagnosa penyakit diabetes mellitus

1.4 Tujuan

Tujuan yang ingin dicapai dari penulisan Tugas Akhir adalah :

1. Mengklasifikasi dengan menggunakan algoritma C4.5 dari data pasien penyakit diabetes mellitus
2. Menguji hasil klasifikasi dengan menggunakan teknik *confusion matrix*

1.5 Manfaat

Berdasarkan dari latar belakang, maka didapatkan manfaat dalam pengelolaan data tersebut yaitu :

1. Mengetahui hubungan antara beberapa faktor yang ada dengan diabetes mellitus .
2. Membantu masyarakat untuk mendiagnosa awal penyakit diabetes mellitus sebelum dilakukan tindakan selanjutnya.

1.6 Tinjauan Pustaka

Penelitian dengan menggunakan metode algoritma C4.5 dan pohon keputusan sudah pernah dilakukan oleh Yusuf (2014) yang membahas tentang pengolahan data kelulusan mahasiswa fakultas komunikasi dan informatika universitas muhammadiyah surakarta menggunakan algoritma C4.5 dan pembuatan pohon keputusan dari data predikat kelulusan mahasiswa. Penelitian tersebut menggunakan atribut jurusan, gender, asal sekolah, jumlah sks, assiten, lama studi, dan predikat.

Penelitian mengenai pendukung keputusan penentu diabetes mellitus sudah pernah dilakukan oleh Fajar (2013) yang membahas tentang sistem pendukung keputusan penentuan penyakit Diabetes Mellitus menggunakan Metode Learning Vector Quantization (LVQ). Variabel-variabel pendukung penegakan diagnosis seperti kadar glukosa darah puasa, kadar glukosa plasma puasa, kadar glukosa plasma tidur, kadar insulin, kadar HbA1c, kadar kolesterol HDL, kadar trigliserida, dan umur digunakan sebagai data learning, dan hasil diagnosa data learning ditetapkan sebagai target awal. Hasil dari proses pembelajaran menggunakan LVQ berupa bobot yang akan digunakan untuk proses klasifikasi dengan menghitung jarak suatu data terhadap tiap bobot menggunakan Euclidean distance, selanjutnya dibandingkan, yang terkecilah pemenangnya.

Penelitian dengan menggunakan metode algoritma C4.5 juga sudah pernah dilakukan untuk memprediksi penerimaan calon pegawai di PT WISE (harryanto dan hansun, 2017). Penelitian dimulai dari studi pustaka, pengumpulan sampel data, analisis sampel data dan perancangan dan pembangunan aplikasi. Penelitian menggunakan metode pengukuran akurasi ten-fold cross validation telah didapatkan hasil pengukuran tingkat keberhasilan prediksi calon pegawai baru sebesar 71% dengan menggunakan aplikasi prediksi calon pegawai yang menerapkan algoritma C4.5.

Dalam membandingkan masalah yang diangkat pada kasus diatas maka dalam pembuatan Tugas Akhir ini mengangkat judul Klasifikasi Penyakit Diabetes Mellitus dengan Menggunakan Algoritma Decision Tree C4.5.

Penelitian ini menggunakan data pasien dari hasil rekam medik dan laboratorium rumah sakit BP Batam. Data tersebut nantinya akan diolah menggunakan algoritma C4.5 dan menghasilkan suatu keputusan untuk mendiagnosa awal penyakit diabetes mellitus.

Tabel 1 Perbandingan dengan Penelitian Sebelumnya

No	Judul	Peneliti (Tahun)	Objek yang diteliti	Metode	Pembuatan Aplikasi
1	Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Kelulusan Mahasiswa Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta	Nugroho (2014)	Data Predikat Kelulusan Mahasiswa	Algoritma C4.5	-
2	Implementasi Learning Vector Quantization Untuk Diagnosa Penyakit Diabetes Militus	Hariri (2013)	Data hasil laboratorium dan rekam medik dari pasien diabetes rumah sakit	Learning Vector Quantization	Desktop
3	Memprediksi Penerimaan Calon Pegawai Baru di PT WISE	Harryanto dan Hansun (2017)	Data calon pegawai baru	Algoritma C4.5	Desktop
4	Klasifikasi Penyakit Diabetes Mellitus dengan Menggunakan Algoritma Decision Tree C4.5	Susanto (2018)	Data hasil laboratorium dan rekam medik dari pasien diabetes rumah sakit BP Batam	Algoritma C4.5	Desktop

1.7 Sistematika Penulisan

Untuk mempermudah penulisan dan pembahasan selanjutnya, maka uraian pembahasan ditulis secara sistematis sebagai berikut :

BAB I PENDAHULUAN

Bab ini berisi mengenai pendeskripsian masalah yang terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat dan sistematika penulisan laporan tugas akhir.

BAB II LANDASAN TEORI

Bab ini berisi tentang penjelasan teori-teori yang berhubungan dengan penelitian

BAB III ANALISIS DAN PERANCANGAN

Bab ini berisi tentang desain penelitian, lokasi dan waktu penelitian, alat dan bahan penelitian, teknik pengumpulan data, teknik pengolahan dan analisis data, serta penyajian data.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tentang implementasi dan pembahasannya serta hasil pengujian.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan yang merupakan rangkuman dari hasil analisis kinerja pada bagian sebelumnya serta saran saran pengembangan dari penelitian yang dibuat dan aspek yang belum terselesaikan.

BAB II

LANDASAN TEORI

2.1 Diabetes Mellitus

Diabetes Mellitus (DM) merupakan suatu kelompok penyakit metabolik dengan karakteristik hiperglikemia yang terjadi karena kelainan sekresi insulin, kerja insulin atau kedua-duanya (*American Diabetes Association, 2010*). Insulin adalah hormon alami yang diproduksi oleh pankreas. Ketika kita makan, pankreas melepaskan hormon insulin yang memungkinkan tubuh mengubah glukosa menjadi energi dan disebarkan di seluruh tubuh. Hormon yang satu ini juga membantu tubuh menyimpan energi tersebut.

Klasifikasi Diabetes Mellitus menurut *American Diabetes Association* (2010) adalah sebagai berikut:

1. Diabetes Mellitus tipe 1 (Hasil dari kehancuran sel β pankreas, biasanya menyebabkan defisiensi insulin yang absolut)
2. Diabetes Mellitus tipe 2 (Hasil dari gangguan sekresi insulin yang progresif yang menjadi latar belakang terjadinya resistensi insulin)
3. Diabetes Mellitus spesifik lain
Misalnya : gangguan genetik pada fungsi sel β , gangguan genetik pada kerja insulin, penyakit eksokrin pankreas (seperti cystic fibrosis), dan yang dipicu oleh obat atau bahan kimia (seperti dalam pengobatan HIV/AIDS atau setelah transplantasi organ)
4. Diabetes Mellitus Kehamilan (Diabetes didiagnosis selama kehamilan)

2.2 Glukosa Darah

Glukosa darah atau kadar gula darah adalah istilah yang mengacu kepada tingkat glukosa di dalam darah. Konsentrasi gula darah, atau tingkat glukosa serum, diatur dengan ketat di dalam tubuh. Glukosa yang dialirkan melalui darah adalah sumber utama energi untuk sel-sel tubuh. Glukosa (kadar gula darah), suatu gula monosakarida, karbohidrat terpenting yang digunakan sebagai sumber tenaga utama dalam tubuh. Glukosa merupakan prekursor untuk sintesis

semua karbohidrat lain di dalam tubuh seperti glikogen, ribose dan deoxiribose dalam asam nukleat, galaktosa dalam laktosa susu, dalam glikolipid, dan dalam glikoprotein dan proteoglikan (Murray R. K. et al., 2003).

Berbagai macam pemeriksaan gula darah (Depkes RI, 1999), yaitu :

1. Tes gula darah sewaktu (GDS)

GDS merupakan tes gula darah yang dilakukan setiap waktu sepanjang hari tanpa memperhatikan makanan terakhir yang dimakan dan kondisi tubuh orang tersebut. Tes ini bisa Anda lakukan kapan saja dan di mana saja, sehingga biasa dilakukan oleh orang dengan diabetes. Tes GDS dilakukan secara acak dalam satu hari karena kadar glukosa darah bisa saja berubah setiap waktu pada orang dengan diabetes. Berbeda dengan orang sehat yang biasa mempunyai kadar gula darah yang tidak banyak berubah dalam satu hari. Jika pada orang sehat dilakukan tes gula darah sewaktu dan hasilnya menunjukkan hasil yang bervariasi, mungkin orang tersebut sedang mengalami masalah pada gula darahnya.

2. Tes gula darah 2 jam post-prandial (GD2PP)

Tes gula darah ini dilakukan 2 jam setelah Anda makan. Tes ini berguna untuk mengetahui apakah seseorang dengan diabetes sudah tepat dengan pola makannya. Jika hasilnya tinggi, kemungkinan makanan Anda yang Anda makan sebelumnya mengandung jumlah gula atau karbohidrat yang banyak, dan sebaliknya. Tes ini mungkin kurang tepat untuk mendiagnosis apakah Anda menderita diabetes atau tidak.

3. Tes gula darah puasa (GDP)

Tes gula darah ini dilakukan setelah Anda berpuasa selama 8-10 jam. Biasanya Anda disarankan untuk melakukan puasa pada malam hari dan pagi harinya Anda melakukan tes GDP ini. Tes GDP sering digunakan sebagai tes pertama untuk mengetahui apakah Anda menderita prediabetes atau diabetes.

4. Hemoglobin A1c (HbA1c) atau glikohemoglobin

Tes ini mengukur seberapa banyak glukosa (gula) yang menempel pada sel darah merah. Tes HbA1c biasanya dilakukan pada penderita diabetes

untuk mengetahui seberapa baik ia dapat mengontrol penyakitnya dalam dua sampai tiga bulan terakhir. Dari hasil tes tersebut, dokter juga dapat menentukan apakah obat diabetes Anda perlu diganti. Hasil tes HbA1c juga dapat memberi tahu berapa kadar rata-rata gula darah Anda. Tes HbA1c juga dapat digunakan pada orang sehat untuk mendiagnosis apakah ia menderita diabetes atau tidak.

2.3 Data Mining

Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk pengenalan permasalahan pengambilan informasi dari database yang besar (Larose,2005). Data Mining merujuk pada keseluruhan proses yang terdiri dari pengumpulan dan analisis data, pengembangan model pembelajaran induktif dan adopsi praktek pengambilan keputusan serta konsekuensi dari tindakan berdasarkan pengetahuan yang diperoleh. Kegiatan data mining dapat dibagi menjadi dua bagian utama penelitian, sesuai dengan tujuan utama dari analisisnya yaitu interpretasi dan prediksi.

Data Mining merupakan bagian yg terintegrasi dari Knowledge Discovery in Databases (KDD). KDD terbagi atas 3 tahap secara global yakni :

1. Data preprocessing
2. Data mining
3. Postprocessing

Data preprocessing bertujuan mentransformasikan data mentah ke format yg sesuai untuk analisis. Terdiri atas proses seleksi fitur, reduksi dimensionalitas, normalisasi dan subsetting data. Data preprocessing merupakan salah satu tahapan yang digunakan didalam penelitian untuk menghasilkan data yang sesuai untuk dapat dilakukan klasifikasi. Postprocessing bertujuan untuk menjamin bahwa hasil proses data mining yang diintegrasikan pada sistem penunjang keputusan, benar-benar hasil yang valid. Terdiri atas proses penapisan pola, visualisasi dan interpretasi pola (Larose,2005).

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

1. Deskripsi

Menggambarkan sekumpulan data secara ringkas. Data yang digambarkan berupa:

- a. Deskripsi grafis : diagram titik, histogram.
- b. Deskripsi lokasi : mean(rata-rata), median (nilai tengah), modus, kuartil, persentil.
- c. Deskripsi keberagaman : range(rentang), varians dan standar deviasi.

2. Estimasi

Memperkirakan suatu hal dari sejumlah sample yang kita miliki (yg tidak kita ketahui). Estimasi hampir sama dengan klasifikasi, kecuali variable target. Estimasi lebih kearah numeric dari pada kearah kategori.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa datang (memperkirakan hal yang belum terjadi). Kita bisa menunggu hingga hal itu terjadi untuk membuktikan seberapa tepat prediksi kita.

4. Klasifikasi

Kegiatan menggolongkan, dengan menggunakan data historis (sebagai data yang digunakan untuk latihan dan sebagai pengalaman). Dalam klasifikasi terdapat variabel prediktor dan target variable. Metode ini digunakan didalam penelitian untuk melakukan klasifikasi dengan menggunakan algoritma pemrograman C4.5.

5. Pengklusteran

Pengklusteran merupakan pengelompokan record, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainya dan memiliki ketidak miripan dengan record-record dalam cluster.

6. Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang biasa.

Data mining di dalam penelitian ini berfungsi untuk proses pengolahan data menjadi data yang sesuai kebutuhan agar dapat digunakan untuk klasifikasi dengan menggunakan algoritma C4.5.

2.4 *Preprocessing Data*

Preprocessing data merupakan proses pembersihan data agar menghasilkan data yang berkualitas atau data yang baik untuk dapat dilakukan proses data mining. Proses *Preprocessing* secara umum terbagi 4, (Han, Kamber, Pei, 2012) yaitu :

a. Data Cleaning

Data cleaning digunakan untuk melengkapi atau menghilangkan data yang tidak lengkap, menghaluskan atau meniadakan data noisy (data yang menyimpang dari biasanya) serta memperbaiki data yang tidak konsisten.

- Cara Menangani data yang tidak lengkap

1. Mengabaikan atau menghilangkan.
2. Mengisi dengan konstanta, misalnya “unknown” atau “null”
3. Mengisi nilainya secara manual.
4. Mengisi dengan mean/median.
5. Mengisi dengan nilai yang paling mungkin, misalnya dari fungsi regresi.

- Cara menangani data noisy (data yang menyimpang dari biasanya)

1. **Pemeriksaan** : data outlier oleh komputer dan manusia, setelah diketahui ada data outlier, maka hapus data outlier tersebut.
2. **Binning** : data dikelompokkan ke dalam bin-bin. Kemudian , data outlier dihaluskan dengan rata-rata bin atau median bin atau batas

bin

3. **Klastering** : Mengelompokkan data, kemudian akan terdeteksi data outlier dan kemudian menghapusnya.

4. **Regresi** : Menghaluskan data dengan fungsi regresi

b. Data Integration

Data integration digunakan untuk penggabungan data dari berbagai database kedalam satu database baru. Masalah yang sering terjadi dalam integrasi data adalah terjadinya konflik nilai data seperti :

- Perbedaan nilai atribut (perlu diidentifikasi apakah memang berbeda atau sebenarnya sama) contoh : "id user" dengan "NIK"
- Perbedaan representasi ukuran atau skala. Misal: cm dan m, kg dan g.

c. Data Transformation

Data transformation digunakan untuk merubah data menjadi bentuk yang sesuai untuk proses mining. Cara transformasi data dibagi menjadi 5 yaitu :

1. **Penghalusan** : misalnya dengan metode binning, klastering, fungsi regresi, dll. Metode binning pada data cleaning hanya untuk menggantikan data noisy, sedangkan pada transformasi untuk menggantikan data per atribut secara keseluruhan.
2. **Agregasi** : menerapkan operasi rangkuman pada data.
3. **Normalisasi** : data diskalakan agar jatuh ke dalam suatu range kecil tertentu. Mis: -1 sampai 1, 0 sampai 1. Jenis Normalisasi:
 - Normalisasi min-max
 - Normalisasi z-score
 - Normalisasi decimal scaling
4. **Generalisasi** : data mentah(raw) diubah ke konsep level yang lebih tinggi melalui penggunaan konsep hirarki.
5. **Konstruksi atribut** : Atribut-atribut baru dibangun dari atribut-atribut yang ada.

d. Data reduction

Data reduction digunakan untuk mengurangi ukuran data tetapi menghasilkan hasil analisis yang sama.

2.5 Algoritma *Decision Tree* C4.5

Algoritma C4.5 merupakan salah satu algoritma pemrograman yang digunakan untuk melakukan klasifikasi. *Decision Tree* 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 (Berry&Linoff,2004). Simpul pada sebuah pohon keputusan dibedakan menjadi tiga, akar simpul, simpul percabangan, dan simpul akhir. Pohon keputusan ini sendiri juga sudah banyak digunakan pada berbagai bidang ilmu pengetahuan, salah satunya yaitu bidang kesehatan untuk diagnosa penyakit pasien. Pohon Keputusan dengan Algoritma C4.5 sangat berhubungan, karena dasar dari algoritma C4.5 adalah pohon keputusan. Algoritma C4.5 merupakan kelompok algoritma *Decision Tree*. Algoritma ini mempunyai input berupa training samples dan samples. Training samples berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan samples merupakan field-field data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data (Sunjana.2010).

Ada beberapa tahapan dalam membuat sebuah *decision tree* dalam algoritma C4.5 (Larose, 2005) yaitu :

1. Mempersiapkan data training.

Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.

2. Menghitung akar dari pohon.

Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang

akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung dahulu nilai entropi.

Untuk menghitung nilai entropi digunakan rumus :

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

Keterangan :

S= Himpunan kasus

k = jumlah partisi S

P_j = proporsi S_i terhadap S

Kemudian hitung nilai gain menggunakan rumus :

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

Keterangan :

S = Himpunan Kasus

A = Atribut

k = jumlah partisi atribut A

|S_i| = Proporsi S_i terhadap S

|S| = jumlah kasus dalam S

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

3. Ulangi langkah ke 2 hingga semua record terpartisi
4. Proses partisi decision tree akan berhenti saat :
 - a. semua record dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut didalam record yang dipartisi lagi
 - c. Tidak ada record didalam cabang yang kosong

Algoritma *decision tree* C4.5 di dalam penelitian ini merupakan metode yang digunakan untuk pengolahan data yang dapat menghasilkan pohon keputusan untuk memprediksi diabetes mellitus.

2.6 Confusion Matrix

Confusion Matrix digunakan untuk pengujian hasil klasifikasi yang dibuat menggunakan algoritma C4.5 agar mengetahui berapa persen keberhasilan dalam melakukan klasifikasi data. *Confusion Matrix* adalah tool yang digunakan untuk evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah. Sebuah matrix dari prediksi yang akan dibandingkan dengan kelas yang berisi informasi nilai aktual dan prediksi pada klasifikasi. Sebuah *rule* hasil klasifikasi dengan *decision tree* jika diterapkan untuk prediksi perlu dilakukan evaluasi dan validasi hasil sehingga diketahui seberapa akurat hasil prediksi. Untuk evaluasi dan validasi hasil *rule* klasifikasi dapat menggunakan *confusion matrix*.

Metode ini menggunakan tabel matriks seperti pada Tabel 2, jika dataset hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif (Bramer, 2007). Evaluasi dengan confusion matrix menghasilkan nilai accuracy, precision, dan recall. Accuracy dalam klasifikasi adalah persentase ketepatan record data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi (Han & Kamber, 2006). Sedangkan precision atau confidence adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. Recall atau sensitivity adalah proporsi kasus positif yang sebenarnya diprediksi positif secara benar (Powers, 2011).

Tabel 2 Model Confusion Matrix

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	TP (<i>True Positive</i>)	FN (<i>False Negative</i>)
Negatif	FP (<i>False Positive</i>)	TN (<i>True Negative</i>)

Berdasarkan nilai (TN), (FP), (FN) dan (TP) dapat diperoleh nilai akurasi, presisi dan *recall*. Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh dengan Persamaan 1. Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan

secara benar dibagi dengan total data yang diklasifikasi positif. Presisi dapat diperoleh dengan Persamaan 2. Sementara itu, *recall* menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar oleh sistem. Nilai *recall* diperoleh dengan Persamaan 3.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (1)$$

$$Presisi = \frac{TP}{FP+TP} * 100\% \quad (2)$$

$$Recall = \frac{TP}{FN+TP} * 100\% \quad (3)$$

Keterangan

- TP adalah True Positive, yaitu jumlah data positif yang terklasifikasi dengan benar oleh sistem.
- TN adalah True Negative, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
- FN adalah False Negative, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.
- FP adalah False Positive, yaitu jumlah data positif namun terklasifikasi salah oleh sistem

2.7 MATLAB

MATLAB (*Matrix Laboratory*) adalah suatu program untuk analisis dan komputasi numerik dan merupakan suatu bahasa pemrograman matematika lanjutan yang dibentuk dengan dasar pemikiran menggunakan sifat dan bentuk matriks (Cahyono,2013). MATLAB merupakan produk komersial dari perusahaan Mathworks, Inc. yang dalam perkembangan selanjutnya dikembangkan menggunakan bahasa C++ dan assembler.

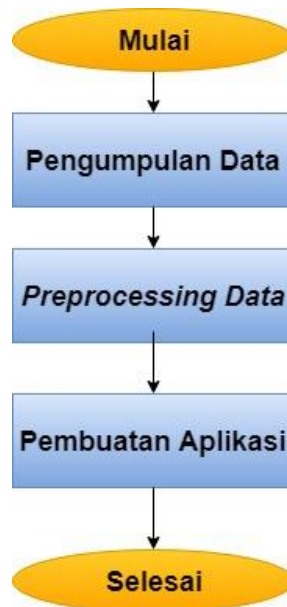
Dengan demikian jika di dalam perhitungan kita dapat menformulasikan masalah ke dalam format matriks maka MATLAB merupakan software terbaik untuk penyelesaian numeriknya. MATLAB yang merupakan bahasa pemrograman tingkat tinggi berbasis pada matriks sering digunakan untuk teknik komputasi numerik, untuk menyelesaikan masalah-masalah yang

melibatkan operasi matematika elemen, matrik, optimasi, aproksimasi dan lain-lain. Sehingga MATLAB banyak digunakan pada : (1) Matematika dan Komputansi, (2) Pengembangan dan Algoritma, (3) Pemrograman modeling, simulasi, dan pembuatan prototype, (4) Analisa Data , eksplorasi dan visualisasi, (5) Analisis numerik dan statistic, dan (6) Pengembangan aplikasi teknik.

MATLAB di penelitian ini digunakan untuk pembuatan algoritma C4.5 yang berfungsi untuk klasifikasi data yang telah di dapatkan dari proses data mining dan pembuatan aplikasi desktop

BAB III ANALISIS DAN PERANCANGAN

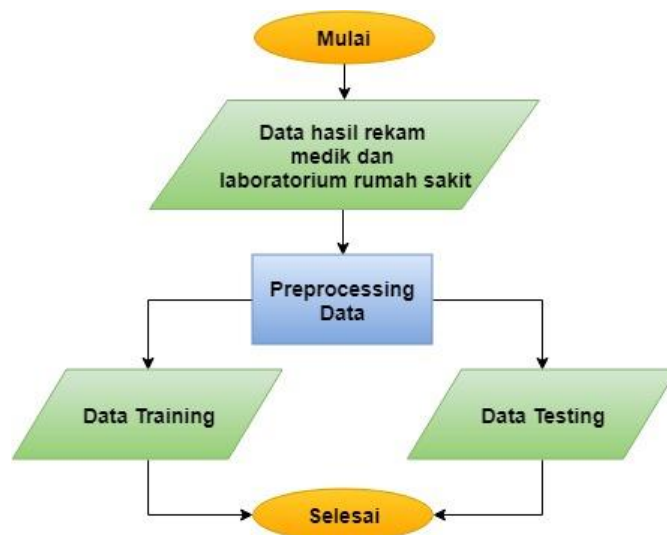
3.1 Desain Penelitian



Gambar 1 Tahap Penelitian secara keseluruhan

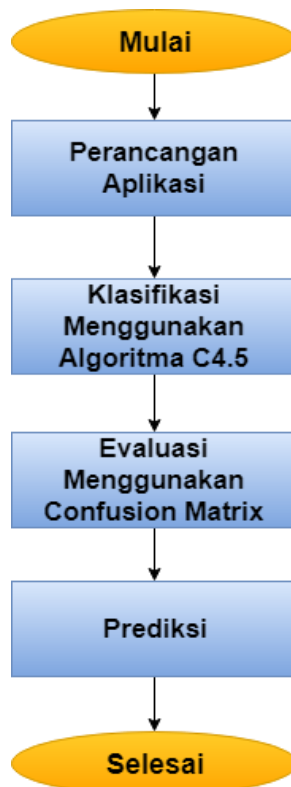
Gambar 1 menunjukkan keseluruhan tahap yang dilakukan dalam proses penelitian. Penelitian yang dilakukan dibagi menjadi 3 tahap yaitu pengumpulan data, preprocessing data, dan pembuatan aplikasi. Pada tahap pengumpulan data langkah yang dilakukan yaitu mengambil data dari hasil rekam medik dan laboratorium rumah sakit BP Batam. Pada tahap preprocessing data langkah yang dilakukan adalah melakukan preprocessing data agar menghasilkan data yang berkualitas atau data yang baik untuk data mining serta mengelompokkan data yang digunakan untuk training dan testing. Data training digunakan oleh algoritma *decision tree* C4.5 untuk membentuk sebuah model *classifier*. Model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum pernah ada. Data Testing digunakan untuk mengukur sejauh mana *classifier* berhasil melakukan klasifikasi dengan benar. *Preprocessing data* yang dilakukan adalah data cleaning, data integration, data transformation, dan data reduction. Data cleaning digunakan untuk melengkapi atau menghilangkan data yang tidak lengkap,

menghaluskan atau meniadakan data noisy serta memperbaiki data yang tidak konsisten. Data integration digunakan untuk penggabungan data dari berbagai database kedalam satu database baru. Data transformation digunakan untuk merubah data menjadi bentuk yang sesuai untuk proses mining. Data reduction digunakan untuk mengurangi ukuran data tetapi menghasilkan hasil analisis yang sama. Tahap pembuatan aplikasi merupakan tahap untuk membuat sebuah produk perangkat lunak (aplikasi) yang dapat melakukan klasifikasi menggunakan algoritma C4.5 serta dapat melakukan prediksi dari penyakit diabetes mellitus.



Gambar 2 Tahap *Preprocessing Data*

Gambar 2 menunjukkan tahap *preprocessing data* yaitu dengan melakukan *preprocessing* terhadap data hasil rekam medik dan laboratorium rumah sakit dan setelah itu mengelompokkan data menjadi data training dan data testing. Data training digunakan oleh algoritma *decision tree* C4.5 untuk membentuk sebuah model *classifier*. Model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum pernah ada. Data Testing digunakan untuk mengukur sejauh mana *classifier* berhasil melakukan klasifikasi dengan benar.



Gambar 3 Tahap Pembuatan Aplikasi

Gambar 3 menunjukkan tahap pembuatan aplikasi yang dimulai dari tahap perancangan yaitu perancangan deskripsi umum sistem, kebutuhan fungsional dan non fungsional sistem, use case diagram, skenario use case, activity diagram, class diagram, dan perancangan tampilan antarmuka. Setelah dilakukan perancangan, tahap selanjutnya adalah melakukan pembuatan aplikasi berdasarkan perancangan-perancangan yang telah dibuat sebelumnya. Pembuatan aplikasi ini memiliki beberapa tahapan yaitu tahap klasifikasi yang dimulai dari melakukan klasifikasi menggunakan algoritma C4.5 dari data training yang didapat, setelah itu akan menghasilkan pohon keputusan yang digunakan untuk pembuatan rules hasil klasifikasi. Tahap selanjutnya yaitu tahap untuk mengevaluasi hasil klasifikasi data untuk memperkirakan apakah objek yang dilakukan klasifikasi benar atau salah dengan menggunakan *confusion matrix*. Tahap yang terakhir yaitu membuat suatu sistem agar dapat memprediksi penyakit diabetes dari hasil klasifikasi tersebut.

3.2 Lokasi dan Waktu Penelitian

Lokasi penelitian dilakukan di Rumah Sakit BP Batam pada tanggal 5 Februari 2018 pukul 09:00-14:00 WIB, tanggal 12 Februari 2018 pukul 09:00-14:00 WIB dan tanggal 19 Februari 2018 pukul 09:00-12:00 WIB. Penelitian ini dilakukan untuk pengambilan data pasien penyakit diabetes mellitus.

3.3 Teknik Pengumpulan Data

Dalam penelitian ini dilakukan pengumpulan data dari data pasien diabetes mellitus yang diperoleh dari hasil data rekam medik dan laboratorium pasien tahun 2017 di RSBP Batam. Survei dilakukan untuk mendapatkan gambaran secara umum tentang data pasien diabetes mellitus dan pengelompokan penyakit diabetes mellitus di Rumah Sakit. Hal ini akan memudahkan identifikasi permasalahan dalam perancangan sistem. Pengamatan dilakukan dengan survei dan wawancara kepada pihak Rumah Sakit BP Batam.

Teknik pengumpulan data dilakukan dengan melihat hasil rekam medik dan hasil laboratorium pasien. Untuk atribut-atribut dari hasil diagnosa yang digunakan dalam penelitian ini didasarkan atas data yang tersedia di rumah sakit, wawancara dengan perawat dan pihak laboratorium.

3.4 Teknik Pengolahan dan Analisis Data

Tahap-tahap yang dilakukan untuk pengolahan data dan analisis data adalah sebagai berikut :

1. Mengumpulkan data dari hasil rekam medik dan laboratorium rumah sakit BP Batam.
2. Tahap *preprocesssing data* agar menghasilkan data yang berkualitas atau data yang baik untuk data mining serta mengelompokkan data yang digunakan untuk training dan testing. Proses *Preprocessing* yang diterapkan, yaitu :
 - a. Data Cleaning
Data cleaning digunakan untuk melengkapi atau menghilangkan data yang tidak lengkap, menghaluskan atau meniadakan data noisy (data yang menyimpang dari biasanya) serta memperbaiki data yang tidak

konsisten.

- Cara Menangani data yang tidak lengkap

1. Mengabaikan atau menghilangkan.

b. Data Transformation

Data transformation digunakan untuk merubah data menjadi bentuk yang sesuai untuk proses mining.

Mengelompokkan data yang digunakan untuk training dan testing. Data training digunakan oleh algoritma *decision tree* C4.5 untuk membentuk sebuah model *classifier*. Model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum pernah ada. Data Testing digunakan untuk mengukur sejauh mana *classifier* berhasil melakukan klasifikasi dengan benar.

3. Data yang sudah dilakukan *preprocessing* kemudian dilakukan proses klasifikasi dengan menggunakan metode algoritma C4.5 untuk menghasilkan pohon keputusan dan rule hasil klasifikasi.
4. Setelah didapatkan hasil klasifikasi, langkah berikutnya adalah melakukan evaluasi terhadap hasil klasifikasi tersebut dengan menggunakan *confusion matrix* untuk memperkirakan apakah objek yang dilakukan klasifikasi benar atau salah.

3.5 Alat dan Bahan

Alat

1. Laptop

Merk : Asus

Model : Asus A455L

OS : Windows 10

Digunakan untuk penelitian seperti pencatatan data, pembuatan aplikasi, penyajian data serta pembuatan laporan

Software

1. Microsoft Excel

Digunakan untuk pencatatan data yang diambil dari data hasil rekam medik dan laboratorium rumah sakit serta untuk proses preprocessing data.

2. Matlab

Digunakan untuk pembuatan aplikasi desktop yang berfungsi untuk klasifikasi diabetes menggunakan algoritma C4.5 yang menghasilkan pohon keputusan dan rule hasil klasifikasi.

Bahan-bahan yang digunakan untuk penelitian yaitu :

1. Data hasil rekam medik dan laboratorium rumah sakit

Merupakan data yang akan diolah dalam penelitian. Data-data tersebut mencakup Umur, Gula Darah Puasa (GDP), Gula Darah 2 Jam *Postrandial* (GD2PP), Gula Darah Sewaktu (GDS), Hemoglobin A1c (HBA1c).

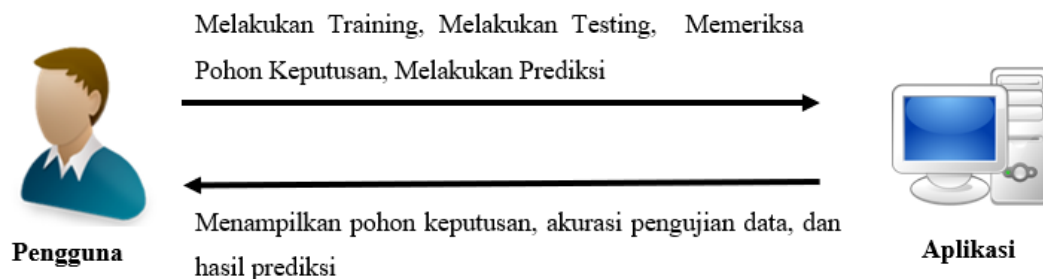
3.6 Penyajian Data

Tabel 3 merupakan data yang dikumpulkan, setelah itu di letakkan di dalam sebuah tabel yang nantinya akan dilakukan *preprocessing* agar menghasilkan data yang berguna untuk proses training dan testing. Data tersebut memiliki 6 variabel yaitu jenis kelamin, Umur, Gula Darah Puasa (GDP), Gula Darah 2 Jam *Postrandial* (GD2PP), Gula Darah Sewaktu (GDS), Hemoglobin A1c (HBA1c) serta memiliki 1 variabel target yaitu keterangan Diabetes 2, Diabetes 1 dan Normal.

Tabel 3 Penyajian Data Asli

No	Jenis Kelamin	Umur	GDP	G2JPP	GDS	HBA1c (%)	Keterangan
1	L	53	131	228	241	6,6	Diabetes 2
2	P	49	149	232	259	6,6	Diabetes 2
3	L	56	162	257	274	6,8	Diabetes 2
4	P	12	153	237	254	6,7	Diabetes 1
5	L	18	162	248	262	6,6	Diabetes 1
6	L	14	156	241	258	6,8	Diabetes 1
7	P	37	82	92	99	5,1	Normal
8	P	31	92	93	95	5,4	Normal
9	P	57	96	106	115	5,7	Normal

3.7 Deskripsi Umum Sistem



Gambar 4 Deskripsi Umum Sistem

Gambar 4 merupakan gambaran umum sistem yang akan dibangun, dari gambar tersebut bahwa pengguna dapat melakukan training dengan memasukkan data training berformat .xlsx, data training tersebut digunakan untuk mencari pohon keputusan dengan menggunakan algoritma C4.5. Pengguna dapat melakukan testing dengan memasukkan data testing berformat .xlsx, data testing tersebut digunakan untuk menghitung akurasi atau ketepatan klasifikasi dengan menggunakan *confusion matrix*. Pengguna dapat memeriksa pohon keputusan serta pengguna dapat melakukan prediksi dengan memasukkan data yang dibutuhkan.

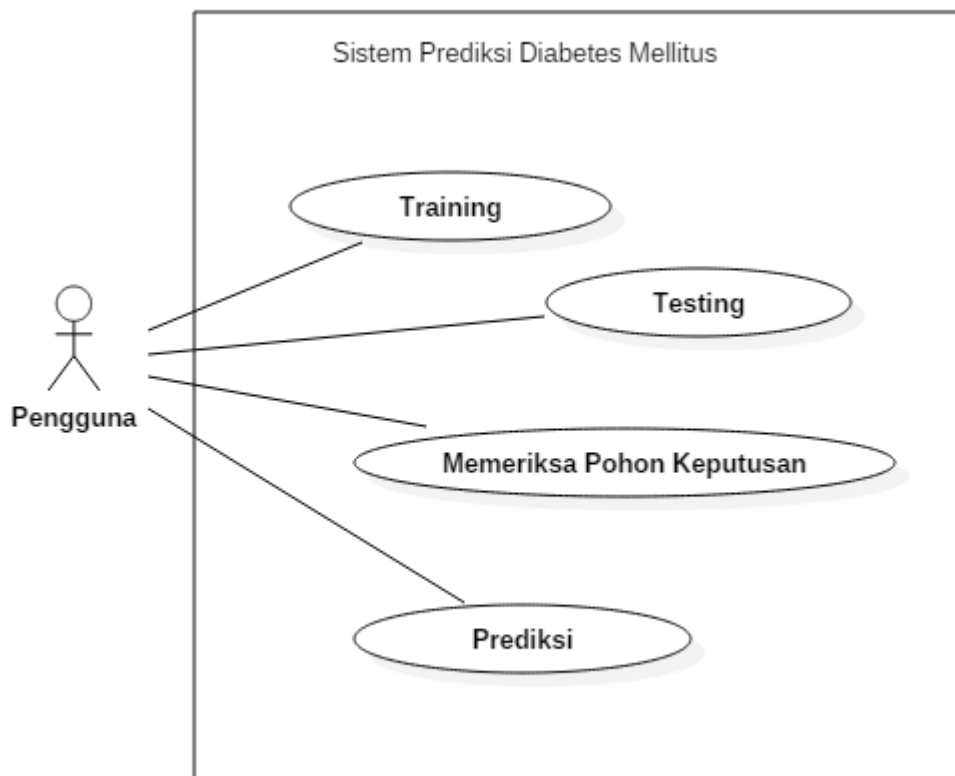
3.8 Kebutuhan Fungsional

- F001. Pengguna melakukan training
- F002. Pengguna melakukan testing
- F003. Pengguna memeriksa pohon keputusan
- F004. Pengguna melakukan prediksi

3.9 Kebutuhan Non Fungsional

- NF001. Sistem harus mudah digunakan

3.10 Use Case Diagram



Gambar 5 Use Case Diagram

Gambar 5 merupakan use case diagram dengan memiliki 4 use case. Pengguna dapat melakukan training dengan memasukkan data training berformat .xlsx, data training tersebut digunakan untuk mencari pohon keputusan dengan menggunakan algoritma C4.5. Pengguna dapat melakukan testing dengan memasukkan data testing berformat .xlsx, data

testing tersebut digunakan untuk menghitung akurasi atau ketepatan klasifikasi dengan menggunakan *confusion matrix*. Pengguna dapat memeriksa pohon keputusan serta pengguna dapat melakukan prediksi dengan memasukkan data yang dibutuhkan.

3.11 Skenario Use Case

3.11.1 Training

Tabel 4 Skenario Use Case (Training)

Nama Use Case	Training
Deskripsi	Untuk melakukan training
Aktor	Pengguna
Kondisi Awal	Form untuk memasukkan data masih kosong
Kondisi Akhir	Menampilkan hasil klasifikasi
Aktor	Sistem
Skenario Normal	
	1. Menampilkan halaman utama
2. Memilih data untuk training	
	3. Memeriksa valid tidaknya data masukan
4. Mengkonfirmasi dimulainya proses training	
	5. Melakukan training menggunakan algoritma C4.5
	6. Menyimpan nilai gain
	7. Menampilkan nilai gain
Skenario Alternatif	
3a. [data training tidak sesuai] memasukkan kembali data training	

3.11.2 Testing

Tabel 5 Skenario Use Case (Testing)

Nama Use Case	Testing
Deskripsi	Untuk melakukan testing
Aktor	Pengguna
Kondisi Awal	Nilai gain sudah tersimpan
Kondisi Akhir	Menampilkan hasil pengujian data
Aktor	Sistem
Skenario Normal	
	1. Menampilkan halaman pengujian dan prediksi
2. Memilih data untuk testing	
	3. Memeriksa valid tidaknya data masukan
	4. Mengambil nilai gain
	5. Melakukan pengujian data menggunakan confusion matrix
	6. Menampilkan hasil pengujian data
Skenario Alternatif	
3a. [data testing tidak sesuai] memasukkan kembali data training	

3.11.3 Memeriksa Pohon Keputusan

Tabel 6 Skenario Use Case (Memeriksa pohon keputusan)

Nama Use Case	Memeriksa pohon keputusan
Deskripsi	Untuk menampilkan pohon keputusan
Aktor	Pengguna
Kondisi Awal	Nilai gain sudah tersimpan
Kondisi Akhir	Menampilkan pohon keputusan
Aktor	Sistem
Skenario Normal	
	1. Menampilkan halaman pengujian dan prediksi
	2. Membaca nilai gain
	3. Menampilkan pohon keputusan
Skenario Alternatif	

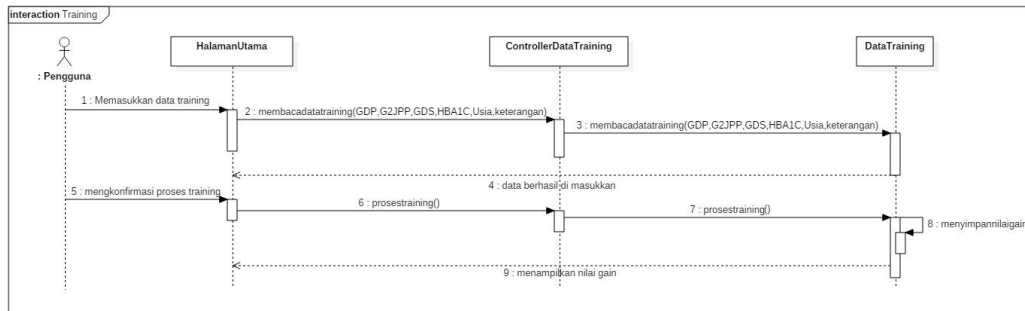
3.11.4 Prediksi

Tabel 7 Skenario Use Case (Prediksi)

Nama Use Case	Prediksi
Deskripsi	Untuk melakukan prediksi
Aktor	Pengguna
Kondisi Awal	Nilai gain sudah tersimpan
Kondisi Akhir	Menampilkan hasil prediksi
Aktor	Sistem
Skenario Normal	
	1. Menampilkan halaman pengujian dan prediksi
2. Memasukkan data yang diperlukan meliputi nama, usia, GDP, G2JPP, GDS, HBA1c	
	3. Memeriksa valid tidaknya data masukan
4. Mengkonfirmasi proses prediksi	
	5. Mengambil nilai gain
	6. Menyusun rule pohon keputusan berdasarkan nilai gain
	7. Melakukan klasifikasi berdasarkan data masukan dan rule pohon keputusan
	8. Menampilkan hasil prediksi
Skenario Alternatif	
3a. [data yang dimasukkan tidak valid] memasukkan kembali data sesuai kolom yang tersedia	

3.12 Sequence Diagram

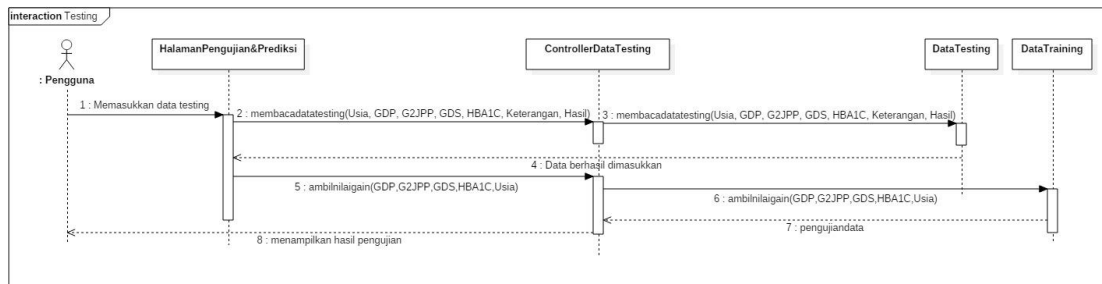
3.12.1 Training



Gambar 6 Sequence Diagram (Training)

Gambar 6 merupakan sequence diagram untuk melakukan training. Pengguna memasukkan data training. Data training yang dimasukkan merupakan data yang telah melalui proses pengolahan data sehingga layak untuk dilakukan klasifikasi dan data training tersebut berformat .xlsx. Setelah itu sistem akan mencari nilai gain menggunakan algoritma C4.5 dan nilai gain akan ditampilkan.

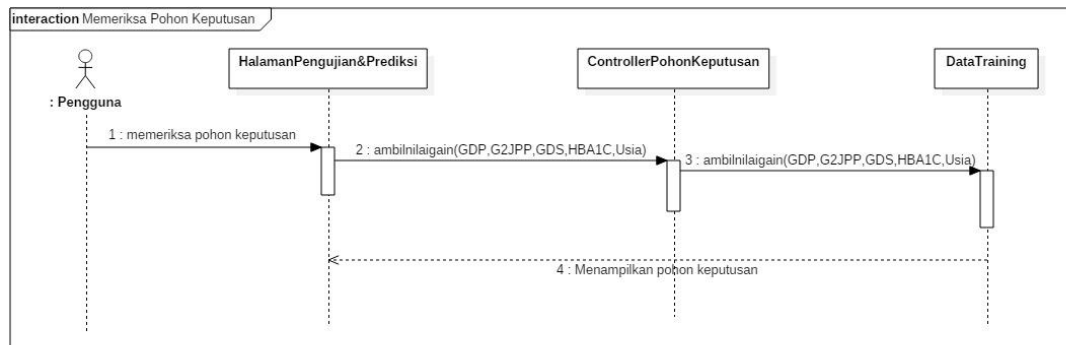
3.12.2 Testing



Gambar 7 Sequence Diagram (Testing)

Gambar 7 merupakan sequence diagram untuk melakukan testing. Pengguna memasukkan data testing. Data testing yang dimasukkan merupakan data yang telah melalui proses pengolahan data sehingga layak untuk dilakukan pengujian dengan menggunakan confusion matrix dan data testing tersebut berformat .xlsx.

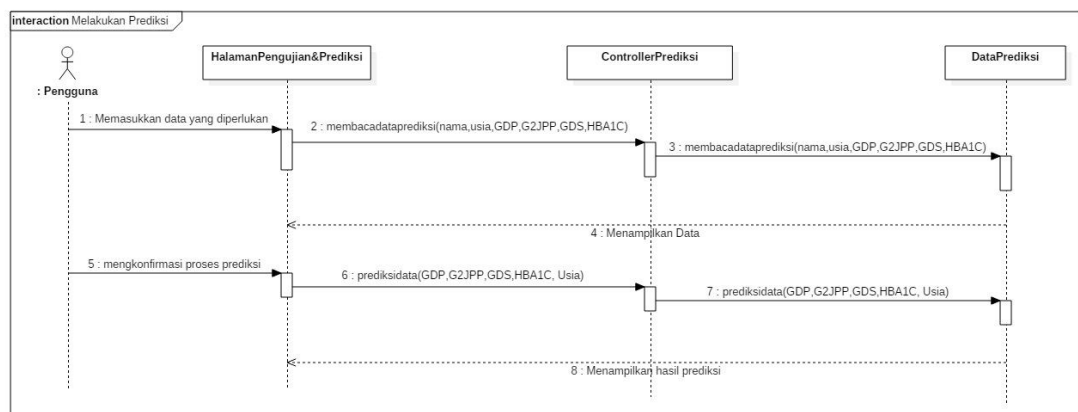
3.12.3 Memeriksa Pohon Keputusan



Gambar 8 Sequence Diagram (Memeriksa Pohon Keputusan)

Gambar 8 merupakan sequence diagram untuk memeriksa pohon keputusan. Pengguna memeriksa pohon keputusan, selanjutnya sistem akan mengambil nilai gain yang telah dilakukan training menggunakan algoritma C4.5 dan setelah itu sistem akan menampilkan pohon keputusan dari hasil training tersebut.

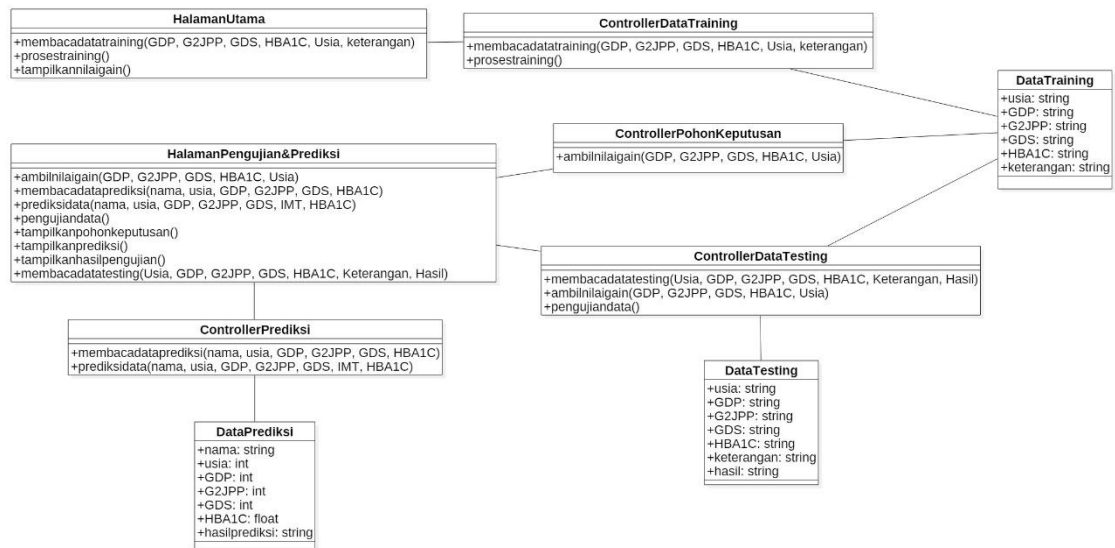
3.12.4 Prediksi



Gambar 9 Sequence Diagram (Prediksi)

Gambar 9 merupakan sequence diagram untuk melakukan prediksi. Pengguna memasukkan data pasien berupa nama, usia, GDP, G2JPP, GDS, HBA1c lalu sistem akan memproses data masukkan tersebut dan akan menampilkan hasil prediksi.

3.13 Class Diagram



Gambar 10 Class Diagram

Gambar 10 merupakan class diagram untuk aplikasi prediksi diabetes mellitus. Pada Class Diagram terdapat 3 objek yaitu Data Training, Data Testing dan Data Prediksi. Controller data training akan membaca data training yang dimasukkan. Controller pohon keputusan akan mengambil nilai gain yang sudah didapatkan. Pada controller data testing akan membaca data testing yang dimasukkan. Pada controller prediksi akan membaca data prediksi kemudian menampilkan data tersebut ke halaman pengujian dan prediksi. Setelah data dimasukkan maka controller prediksi akan melakukan prediksi dari data tersebut dan akan menampilkan hasil prediksinya ke halaman pengujian dan prediksi.

3.14 Perancangan Antarmuka

Pengolahan Data

GDP

G2JPP

GDS

HBA1C

Usia

GDP ▼

Grafik Distribusi Data

Gambar 11 Halaman Utama

Gambar 11 merupakan perancangan antarmuka halaman utama yang digunakan sebagai pengelolaan data training untuk melakukan klasifikasi menggunakan algoritma C4.5. Halaman utama ini digunakan untuk memasukkan data training, menampilkan hasil perhitungan gain dengan menggunakan algoritma C4.5, serta dapat menampilkan distribusi jumlah data dari masing-masing variabel misalnya variabel GDP dengan kelas >125 dan ≤ 125 .

Aplikasi Prediksi

Pohon Keputusan

Pengujian Data

Pilih Data

Cari

TP

FP

Akurasi

TN

FN

Prediksi

Nama
 Usia
 GDP

G2JPP
 GDS
 HBA1C

Prediksi

Hasil Prediksi

Nama

Hasil Prediksi

Gambar 12 Halaman Pengujian dan Prediksi

Gambar 12 merupakan perancangan antarmuka halaman pengujian dan prediksi yang digunakan untuk pengujian data menggunakan *confusion matrix*, Menampilkan pohon keputusan serta melakukan prediksi. Di bagian pengujian, pengguna dapat memasukkan data testing untuk dilakukan pengujian terhadap klasifikasi yang didapat dari data training dan pengguna juga dapat memeriksa pohon keputusan dari hasil klasifikasi data training. Di bagian prediksi, pengguna dapat melakukan prediksi dengan memasukkan data sesuai kolom yang tersedia dan setelah itu akan ditampilkan hasil prediksi sesuai data yang dimasukkan.

BAB IV HASIL DAN PEMBAHASAN

4.1 *Preprocessing Data*

Pada proses *preprocessing* bertujuan untuk mempersiapkan data yang layak digunakan untuk proses training menggunakan algoritma C4.5 dan proses testing dengan menggunakan *confusion matrix*.

Tabel 8 merupakan data training yang telah dilakukan *preprocessing* dari data asli. Data training ini akan digunakan untuk proses training dengan menggunakan algoritma C4.5. Dari proses *preprocessing* didapatkan beberapa kelas dari masing-masing variabel yaitu :

1. Umur = <19, 19-35, >35
2. GDP = >125 dan <=125
3. G2JPP = >205 dan <=205
4. GDS = >227 dan <=227
5. HBA1c = >6,4 % dan <=6,4%

Tabel 8 Data Training

No	Umur	GDP	G2JPP	GDS	HBA1c (%)	Keterangan
1	>35	>125	>205	>227	>6,4	Diabetes 2
2	>35	>125	>205	>227	>6,4	Diabetes 2
3	>35	>125	>205	>227	>6,4	Diabetes 2
4	<19	>125	>205	>227	>6,4	Diabetes 1
5	<19	>125	>205	>227	>6,4	Diabetes 1
6	<19	>125	>205	>227	>6,4	Diabetes 1
7	>35	<=125	<=205	<=227	<=6,4	Normal
8	19-35	<=125	<=205	<=227	<=6,4	Normal
9	>35	<=125	<=205	<=227	<=6,4	Normal

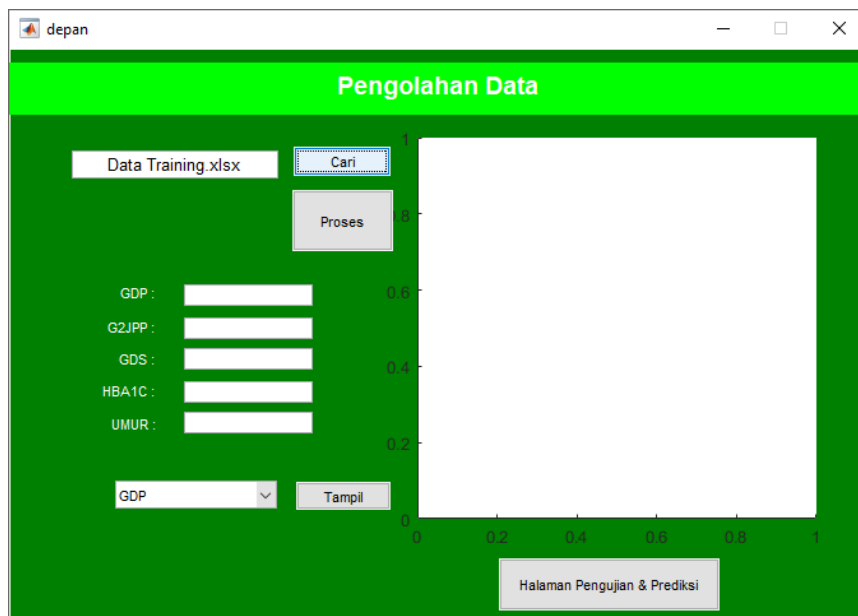
Tabel 9 merupakan data testing yang telah dilakukan *preprocessing* dari data asli. Data testing memiliki tambahan variabel hasil yang digunakan sebagai target untuk pengujian dengan menggunakan *confusion matrix*. Hasil 1 jika keterangannya diabetes 2, hasil 2 jika keterangannya diabetes 1, dan hasil 0 jika keterangannya normal. Hasil ini menentukan apakah hasil dari data sebelum dilakukan klasifikasi sama atau tidak dengan hasil setelah dilakukan klasifikasi.

Tabel 9 Data Testing

Umur	GDP (mg/dl)	2JPP (mg/dl)	GDS (mg/dl)	HBA1c (%)	Keterangan	Hasil
66	445	572	587	8,0	Diabetes 2	1
60	277	429	445	7,2	Diabetes 2	1
62	277	423	256	6,8	Diabetes 2	1
14	156	241	258	6,8	Diabetes 1	2
16	148	225	243	6,6	Diabetes 1	2
43	85	88	93	5,2	Normal	0
30	80	86	89	5,8	Normal	0
38	98	104	110	5,2	Normal	0

4.2 Pengoperasian Sistem

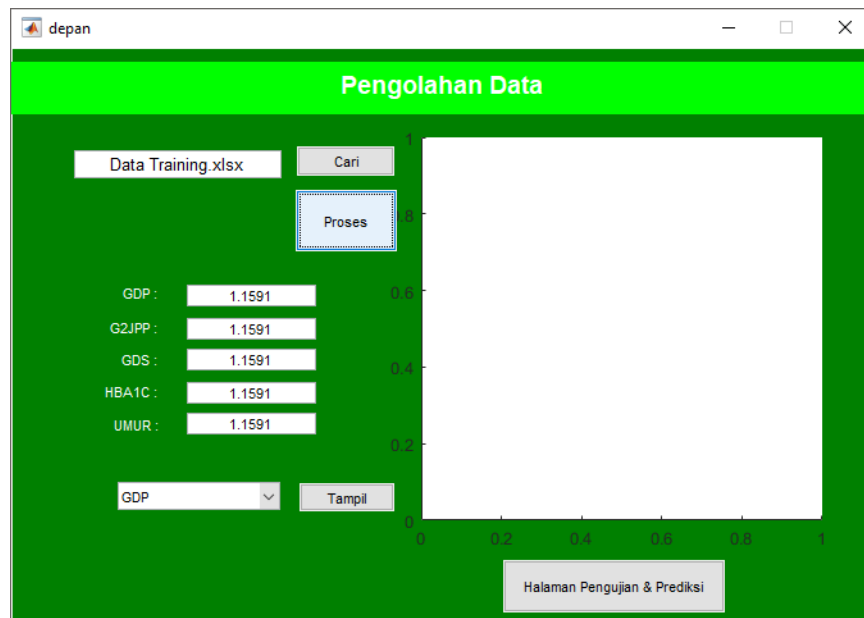
4.2.1 Training



Gambar 13 Pengoperasian Sistem (Memasukkan Data Training)

Gambar 13 merupakan pengoperasian sistem untuk melakukan training. File yang digunakan adalah “Data training” yang berekstensi “.xlsx” yaitu data pasien penyakit diabetes mellitus setelah dilakukan *preprocessing data*, data tersebut memiliki lima variabel pendukung yaitu usia, Gula Darah Puasa (GDP), Gula Darah 2 Jam Setelah Makan (G2JPP), Gula Darah Sementara, dan HemoglobinA1c. Serta mempunyai satu variabel target sebagai klasifikasi keputusan Diabetes Mellitus 1,

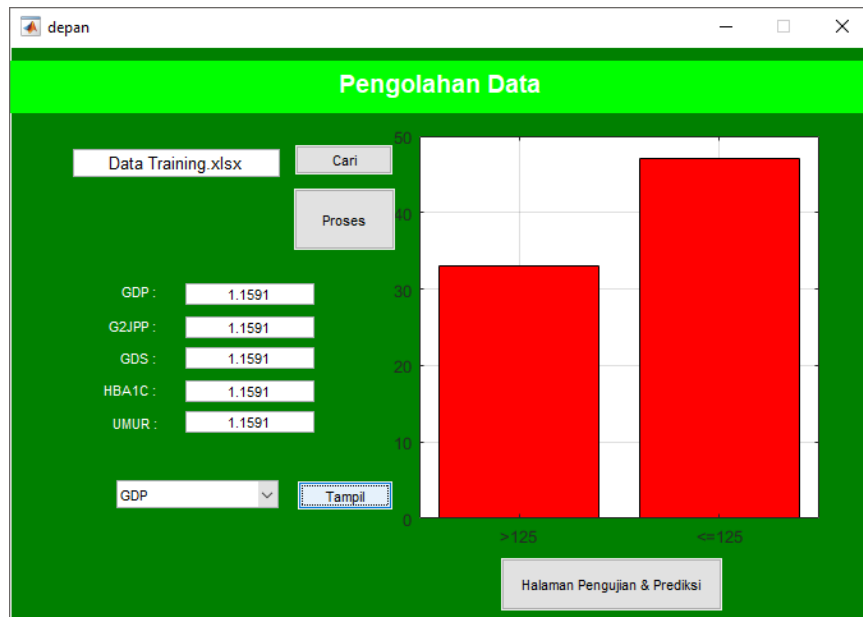
Diabetes Mellitus 2, atau Normal. Data yang dimasukkan tersebut selanjutnya akan dilakukan klasifikasi menggunakan algoritma C4.5.



Gambar 14 Pengoperasian Sistem (Memproses Data Training)

Gambar 14 merupakan pengoperasian sistem untuk memproses data training. Data training yang sudah dimasukkan akan di proses menggunakan algoritma C4.5 untuk mencari nilai gain dari setiap variabel seperti nilai gain GDP, G2JPP, GDS, HBA1c, dan Umur. Nilai gain tersebut akan ditampilkan dan disimpan dengan nama file sip.mat. Hasil training didapatkan nilai gain dari masing-masing variabel yaitu :

GDP	: 1.1591
G2JPP	: 1.1591
GDS	: 1.1591
HBA1c	: 1.1591
Umur	: 1.1591



Gambar 15 Pengoperasian Sistem (Menampilkan Grafik Distribusi Data)

Gambar 15 merupakan pengoperasian sistem untuk menampilkan grafik distribusi data dari setiap variabel. Data distribusi tersebut didapatkan dari perhitungan jumlah masing-masing kelas dari setiap variabel yang ada. Gambar tersebut menampilkan grafik distribusi data pada variabel GDP dengan kelas >125 berjumlah 33 data dan kelas ≤ 125 berjumlah 47.

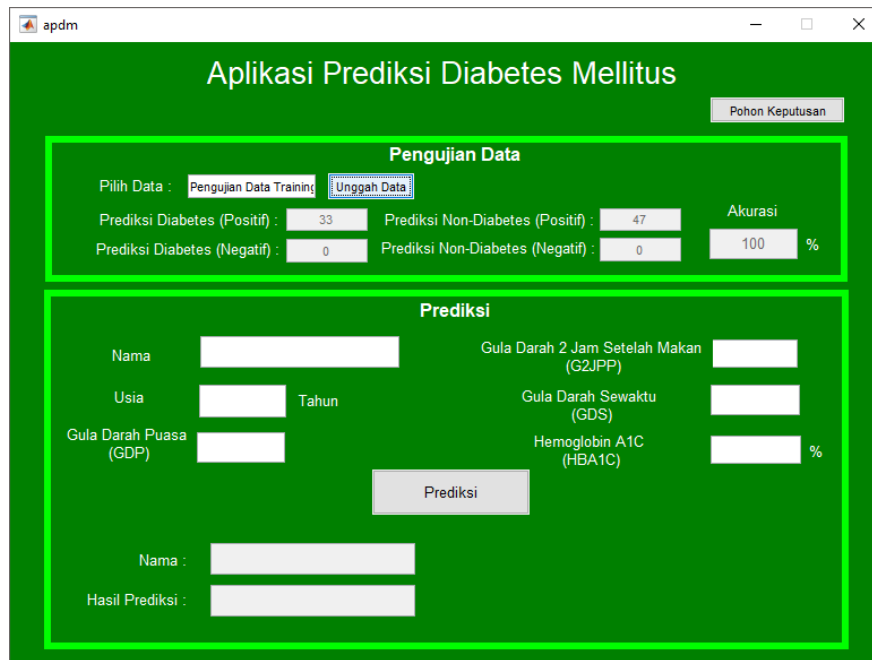
4.2.2 Testing

The screenshot shows a web application window titled 'Aplikasi Prediksi Diabetes Mellitus'. The interface is divided into two main sections: 'Pengujian Data' and 'Prediksi'. In the 'Pengujian Data' section, there is a dropdown menu for 'Pilih Data' set to 'Pengujian Data Testing' and a button 'Unggah Data'. Below this, there are four input fields for prediction counts: 'Prediksi Diabetes (Positif)' with value 12, 'Prediksi Diabetes (Negatif)' with value 0, 'Prediksi Non-Diabetes (Positif)' with value 8, and 'Prediksi Non-Diabetes (Negatif)' with value 0. An 'Akurasi' field shows '100 %'. The 'Prediksi' section contains input fields for 'Nama', 'Usia' (with 'Tahun' label), 'Gula Darah Puasa (GDP)', 'Gula Darah 2 Jam Setelah Makan (G2JPP)', 'Gula Darah Sewaktu (GDS)', and 'Hemoglobin A1C (HBA1C)'. A 'Prediksi' button is located below these fields. At the bottom, there are fields for 'Nama :', 'Hasil Prediksi :', and a 'Pohon Keputusan' button in the top right corner.

Gambar 16 Pengoperasian Sistem (Pengujian Data Testing)

Gambar 16 merupakan pengoperasian sistem untuk pengujian data testing. Data yang dilakukan pengujian merupakan data yang telah dilakukan *preprocessing data* serta file berekstensi “.xlsx”. Pengujian data testing tersebut berjumlah 20 data yang dilakukan pengujian. Gambar 17 menjelaskan bahwa jumlah Prediksi Diabetes (Positif) atau (a) yaitu 12, jumlah Prediksi Diabetes (Negatif) atau (b) yaitu 0, jumlah Prediksi Non-Diabetes (Positif) atau (c) yaitu 8, dan jumlah Prediksi Non-Diabetes (Negatif) atau (d) yaitu 0. Dari keempat data tersebut akan dilakukan pengujian data menggunakan akurasi pada *confusion matrix*.

$$\begin{aligned} \text{Akurasi} &= (a+c) / (a+b+c+d) \\ &= (12+8) / (12+0+8+0) \times 100 \% \\ &= 100 \% \end{aligned}$$

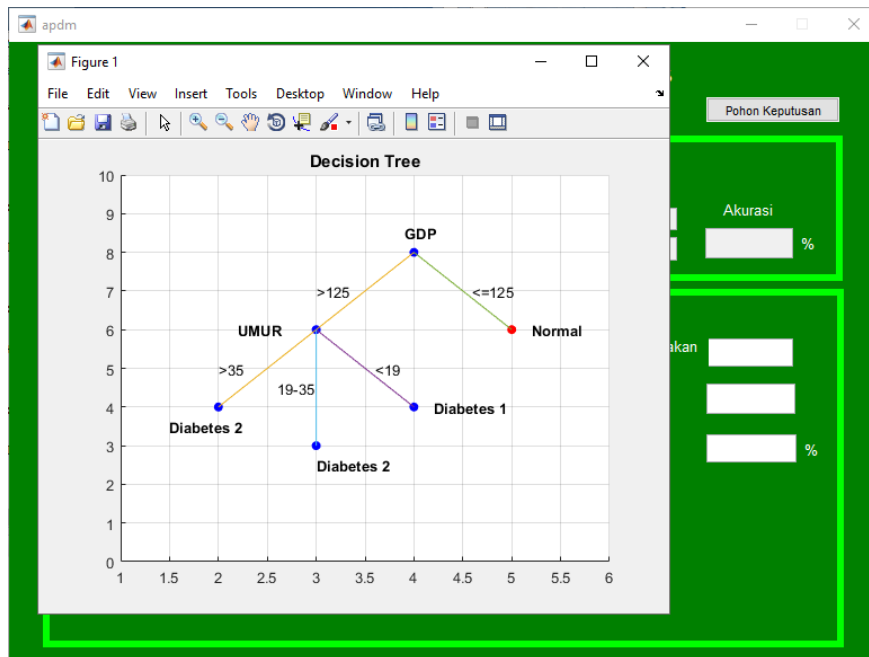


Gambar 17 Pengoperasian Sistem (Pengujian Data Training)

Gambar 17 merupakan pengoperasian sistem untuk pengujian data training. Data yang dilakukan pengujian merupakan data yang telah dilakukan *preprocessing data* serta file berekstensi “.xlsx”. Pengujian data training tersebut berjumlah 80 data yang dilakukan pengujian. Gambar 16 menjelaskan bahwa jumlah Prediksi Diabetes (Positif) atau (a) yaitu 33, jumlah Prediksi Diabetes (Negatif) atau (b) yaitu 0, jumlah Prediksi Non-Diabetes (Positif) atau (c) yaitu 47, dan jumlah Prediksi Non-Diabetes (Negatif) atau (d) yaitu 0. Dari keempat data tersebut akan dilakukan pengujian data menggunakan akurasi pada *confusion matrix*.

$$\begin{aligned}
 \text{Akurasi} &= (a+c) / (a+b+c+d) \\
 &= (33+47) / (33+0+47+0) \times 100\% \\
 &= 100\%
 \end{aligned}$$

4.2.3 Memeriksa Pohon Keputusan



Gambar 18 Pengoperasian Sistem (Memeriksa Pohon Keputusan)

Gambar 18 merupakan pengoperasian sistem untuk memeriksa pohon keputusan. Pohon keputusan tersebut didapatkan dari klasifikasi data training yang dimasukkan sebelumnya dengan melihat perhitungan nilai gain menggunakan algoritma C4.5. Hasil klasifikasi yang didapatkan dari data training yang dimasukkan sebagai berikut :

1. IF GDP “<=125” THEN KETERANGAN = NORMAL
2. IF GDP “>125” AND USIA “>35” THEN KETERANGAN = DIABETES 2
3. IF GDP “>125” AND USIA “19-35” THEN KETERANGAN = DIABETES 2
4. IF GDP “>125” AND USIA “<19” THEN KETERANGAN = DIABETES 1

4.2.4 Prediksi

The screenshot shows a web application window titled 'Aplikasi Prediksi Diabetes Mellitus'. The interface is divided into two main sections: 'Pengujian Data' and 'Prediksi'. In the 'Prediksi' section, the user has entered the following data:

Parameter	Value
Nama	Hadyka
Usia	21 Tahun
Gula Darah Puasa (GDP)	100
Gula Darah 2 Jam Setelah Makan (G2JPP)	134
Gula Darah Sewaktu (GDS)	176
Hemoglobin A1C (HBA1C)	6.2 %

The predicted result is 'Normal'.

Gambar 19 Pengoperasian Sistem (Prediksi)

Gambar 19 merupakan pengoperasian sistem untuk melakukan prediksi. Pengguna diminta untuk memasukkan data nama, usia, GDP, G2JPP, GDS serta HBA1c. Setelah data sudah dimasukkan semua dan menekan tombol prediksi maka sistem akan melakukan prediksi dan menampilkan hasil prediksi.

4.3 Pengujian

Tabel 10 merupakan pengujian dari aplikasi sesuai dengan kebutuhan fungsional seperti melakukan training, melakukan testing, memeriksa pohon keputusan dan melakukan prediksi.

Tabel 10 Pengujian

No	Skenario Pengujian	Data Uji	Hasil yang diharapkan	Hasil yang didapatkan	Validasi
1	Memasukkan data training yang berformat .xlsx	Data training yang sudah di <i>preprocessing data</i>	Data berhasil dimasukkan	Data berhasil dimasukkan	Sukses
2	Melakukan proses perhitungan menggunakan algoritma C4.5	Data training yang sudah dimasukkan	Data berhasil dilakukan perhitungan	Data berhasil dilakukan perhitungan	Sukses
3	Memasukkan data testing yang berformat .xlsx	Data testing yang sudah di <i>preprocessing data</i>	Data berhasil dimasukkan	Data berhasil dimasukkan	Sukses
4	Menampilkan pohon keputusan	Data training yang sudah dilakukan perhitungan menggunakan algoritma C4.5	Berhasil menampilkan pohon keputusan	Berhasil menampilkan pohon keputusan	Sukses

No	Skenario Pengujian	Data Uji	Hasil yang diharapkan	Hasil yang didapatkan	Validasi
5	Memasukkan data prediksi	Nama, Usia, GDP, G2JPP, GDS dan HBA1c	Berhasil dilakukan prediksi	Berhasil dilakukan prediksi	Sukses
	Tidak memasukkan data prediksi	-	Muncul Pesan kesalahan	Muncul Pesan kesalahan	Sukses

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan dari tujuan yang dibuat dan hasil implementasi pada klasifikasi penyakit diabetes mellitus menggunakan algoritma C4.5

1. Hasil klasifikasi data mining bahwa algoritma C4.5 dapat digunakan untuk mengklasifikasi penyakit diabetes mellitus menjadi diabetes 1, diabetes 2 atau normal.
2. Dari metode klasifikasi data mining dengan algoritma C4.5 dan pengaplikasian pohon keputusan yang membentuk aturan tersebut terdapat akurasi pada data training yang berjumlah 80 dari 100 data pasien sebesar 100% sedangkan akurasi pada data testing yang berjumlah 20 dari 100 data pasien sebesar 100%. Perhitungan keduanya menggunakan *confusion matrix*.

5.2 Saran

Saran yang dapat diberikan untuk penelitian selanjutnya adalah dapat dilakukan klasifikasi dengan menggunakan metode *Support Vector Machine (SVM)*, *Nearest Neighbor*, Algoritma *CART (Classification And Regreesion Trees)* maupun metode klasifikasi yang lainnya.

DAFTAR PUSTAKA

- American Diabetes Association*. 2010. *Standards of Medical Care in Diabetes-2010*. *Diabetes Care* 33 (1)
- Cahyono, B.2013. Penggunaan *Software Matrix Laboratory* (MATLAB) Dalam Pembelajaran Aljabar Linier. *Jurnal PHENOMENON* 1 (1). 45-62
- Han, Jiawei, Micheline Kamber, dan Jian Pei. 2012. *Data Mining Concept and Techniques Third Edition*. USA : Elsevier
- Harryanto, F.F., dan Hansun, S. 2017. Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE. *Jatani* 3 (2), 95-103
- Hariri, F. R. 2013. Implementasi *Learning Vector Quantization* Untuk Diagnosa Penyakit Diabetes Militus. Seminar Nasional Teknologi Informasi dan Multimedia 2013
- Larose, D. T. 2005. *Discovering Knowledge in Data An Introduction to Data Mining*. New Jersey: John Willey and Sons.
- Murray, R.K. et al.. *Biokimia Harper*. Edisi 25. Jakarta. Kedokteran. EGC. 2003
- Nugroho, Y.S. 2014. Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Kelulusan Mahasiswa Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta. Yogyakarta : Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST).
- Prasetyo, E. 2014. *Data Mining - Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi.
- Quinlan, J.R. 1986. *Introduction of Decision Trees*. *Machine Learning* 1:81-106.

BIODATA PENULIS



Nama : Agam Juli Nur Susanto
NIM : 3311501081
Tempat,Tanggal Lahir : Batam, 21 Juli 1997
Jenis Kelamin : Laki-Laki
Alamat : Perumahan Bandara Mas
Blok C3 No 02 Batam Kota
Hobby : SepakBola, Membaca

Agam Juli Nur Susanto yang biasa dipanggil agam merupakan anak keempat dari Pasangan Bapak Maryanto dan Ibu Ni Ketut Suardani. Penulis lulusan dari SDN 008 Nongsa pada tahun 2009, SMP N 28 Batam pada tahun 2012, dan SMA N 15 Batam pada tahun 2015. Setelah lulus Sekolah Menengah Atas, penulis melanjutkan pendidikan ke perguruan tinggi negeri di Politeknik Negeri Batam jurusan D3 Teknik Informatika dan sekarang berada di semester 6. Selain fokus dalam kegiatan perkuliahan penulis juga aktif di sebuah unit kegiatan mahasiswa di Politeknik Negeri Batam yaitu penulis tergabung di Lembaga Dakwah Kampus Ikatan Mahasiswa Muslim Politeknik Negeri Batam dari periode 2016-2017. Penulis juga memiliki berbagai sertifikasi keahlian diantaranya Sertifikasi MTA:Software Developmet Fundamental dan Junior Web Programming.