**DAFTAR ISI**

HALAMAN JUDUL i

HALAMAN PERSETUJUAN ii

KATA PENGANTAR iii

DAFTAR ISI v

DAFTAR GAMBAR vii

DAFTAR TABEL viii

DAFTAR ARTI SIMBOL DAN SINGKATAN ix

ABSTRAK xi

ABSTRACT xii

[BAB I 1](#_Toc489969202)

[1.1. Latar Belakang 1](#_Toc489969203)

[1.2 Rumusan Masalah 3](#_Toc489969204)

[1.3 Batasan Masalah 3](#_Toc489969205)

[1.4 Tujuan Penelitian 4](#_Toc489969206)

[1.5 Manfaat Penelitian 4](#_Toc489969207)

[1.6 Sistematika Penulisan 5](#_Toc489969208)

[BAB II 7](#_Toc489969209)

[2.1 Tinjauan Pustaka 7](#_Toc489969210)

[2.2 Landasan Teori 9](#_Toc489969211)

[2.2.1 *Data Mining* 9](#_Toc489969212)

[2.2.2 Fungsi *Data Mining* 10](#_Toc489969213)

[2.2.3 Fase *Cross-Industry Standart Process for Data Mining* (CRISP-DM) 14](#_Toc489969214)

[2.2.4 *Star Schema* 17](#_Toc489969215)

[2.2.5 *Data Warehouse* 18](#_Toc489969216)

[2.2.6 *Naive Bayes Clasifier* 19](#_Toc489969217)

[2.2.7. Penentuan Tingkat Keberhasilan Sistem 28](#_Toc489969218)

[2.2.8 Penyakit *Stroke* 29](#_Toc489969219)

[2.2.8 Rekam Medis 32](#_Toc489969220)

[BAB III 33](#_Toc489969221)

[3.1. Bahan dan Alat Penelitian 33](#_Toc489969222)

[3.1.1 Bahan Penelitian 33](#_Toc489969223)

[3.1.2 Alat Penelitian 33](#_Toc489969224)

[3.2. Desain Penelitian 33](#_Toc489969225)

[3.2.1. Jenis Penelitian 33](#_Toc489969226)

[3.2.2. Tipe Penelitian 34](#_Toc489969227)

[3.4. Jenis dan Sumber Data 34](#_Toc489969228)

[3.5. Teknik Pengumpulan Data 34](#_Toc489969229)

[3.6. Metode Analisis Data 36](#_Toc489969230)

[3.7. Metode Pengembangan Sistem 36](#_Toc489969231)

[3.8 Tahapan dan Diagram Alir Penelitian 38](#_Toc489969232)

[BAB IV 44](#_Toc489969233)

[4.1. Hasil Penelitian 44](#_Toc489969234)

[4.1.1. Analisa Sistem 44](#_Toc489969235)

[4.1.2. Implementasi Sistem 49](#_Toc489969236)

[4.1.3. Pengujian Sistem 58](#_Toc489969237)

[4.2. Pembahasan 73](#_Toc489969238)

[4.2.1. Fungsi Sistem 73](#_Toc489969239)

[4.2.2. Proses Algoritma Naive Bayes Classifier 75](#_Toc489969240)

[4.2.3. Tahapan Alir Sistem 76](#_Toc489969241)

[BAB V 78](#_Toc489969242)

[5.1. Kesimpulan 78](#_Toc489969243)

[5.2. Saran 79](#_Toc489969244)

[DAFTAR PUSTAKA 80](#_Toc489969245)

# **BAB I**

**PENDAHULUAN**

## **Latar Belakang**

Penyakit *stroke* merupakan salah satu penyakit yang dapat menyerang dan menyebabkan kematian secara tiba-tiba pada penderitanya tanpa memandang usia, ras dan jenis kelamin. *Stroke* atau *Cerebrovascular disease* menurut *World Health Organization* (*WHO*)adalah tanda-tanda klinis yang berkembang cepat akibat gangguan fungsi otak fokal atau global karena adanya sumbatan atau pecahnya pembuluh darah di otak dengan gejala-gejala yang berlangsung selama 24 jam atau lebih.

Berdasarkan data dari hasil Riset Kesehatan Dasar 2013 (Riskesdas, 2013), *stroke* merupakan penyebab kematian utama di Indonesia. Prevalensi *stroke* di Indonesia berdasarkan diagnosis tenaga kesehatan sebesar 7,0/mil telah menjadi penderita penyakit *stroke* dan berdasarkan diagnosis tenaga kesehatan sebesar 12,1/mil memiliki gejala menderita penyakit *stroke*.

Oleh karena itu untuk meningkatkan upaya menurunkan angka kematian penyakit *stroke* serta tindakan pencegahannya maka dapat dilakukan upaya preventif agar penderita penyakit *stroke* semakin menurun bahkan dapat dicegah sebelum terjadi, sehingga dampak terjadinya dapat diminimalisir.

Untuk itu, dengan memanfaatkan perkembangan peradaban manusia dengan kemajuan di bidang ilmu pengetahuan dan teknologi maka teknologi yang ada saat ini diharapkan mampu digunakan dalam dunia kesehatan. Yaitu dengan melakukan kolaborasi antara pengaplikasian teknologi dalam dunia medis untuk menekan angka kematian akibat penyakit *stroke*. Salah satu pemanfaatan teknologi dalam bidang kesehatan adalah disiplin ilmu yang berkembang pesat dewasa ini yaitu *Data Mining*.

Secara umum *data mining* merupakan suatu disiplin ilmu yang bertujuan untuk mendapatkan informasi yang bermanfaat dari sekumpulan data yang tersedia, dimana data tersebut biasanya tersedia dalam jumlah yang cukup besar. Informasi yang ingin didapatkan melalui *data mining* ini seringkali didasarkan pada sejumlah fitur yang sudah ditentukan sebelumnya. Fungsi yang dapat dilakukan pada *data mining* antara lain melakukan deskripsi, estimasi, prediksi, klasifikasi, pengklusteran dan asosiasi. Proses penentuan informasi yang didasarkan pada nilai-nilai tertentu dari sejumlah fitur inilah yang mendorong para ahli untuk menciptakan beberapa metode yang disarankan dalam proses *data mining* (penggalian data) sehingga pada akhirnya dapat ditentukan informasi yang bermanfaat. Beberapa metode yang sering disarankan dalam proses *data mining* antara lain, yaitu *Naive Bayes Classifier*, *Decision Tree*, *Artificial Neural Network* (ANN), *Super Vector Machine* (SVM) dan K-*Nearest Neighbor*.

Dalam kaitannya dengan pengklasifikasian penyakit *stroke* proses *data mining* bisa melakukan penelusuran pada data historis untuk mengidentifikasi pola dan memprediksi *trend* didasarkan pada sifat-sifat yang teridentifikasi sebelumnya, kemudian memberikan alternatif pencegahan bila ditemukan indikasi yang mengarah pada timbulnya penyakit *stroke*. Informasi yang dihasilkan untuk selanjutnya bisa digunakan oleh edukator *stroke* maupun dokter sebagai dasar untuk melakukan tindakan-tindakan yang diperlukan.

Oleh karena itu, pada penelitian yang berbasis pada aplikasi metode *data mining* ini, penulis merumuskan judul penelitian “Klasifikasi Penyakit *Stroke* Menggunakan Metode *Naive Bayes Classifier*”.

## **Rumusan Masalah**

Berdasarkan latar belakang masalah yang telah dipaparkan sebelumnya, maka penulis merumuskan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana membuat sistem klasifikasi penyakit *stroke* menggunakan metode *Naive Bayes Classifier*?.
2. Bagaimana tingkat keberhasilan metode *Naive Bayes Classifier* dalam melakukan klasifikasi penyakit *stroke*?.

## **Batasan Masalah**

Berdasarkan permasalahan yang telah dirumuskan sebelumnya, maka penulis membatasi masalah dalam penelitian ini, sebagai berikut:

1. Klasifikasi *stroke* dilakukan berdasarkan kelainan patologis yaitu *ischemic* dan *hemorrhagic*.
2. Data pasien yang diambil antara tahun 2014-2015 pada Rumah Sakit Umum Daerah Undata Palu.
3. Untuk menjaga kode etik pasien maka identitas tertentu pasien tidak dimasukan ke dalam data.

## **Tujuan Penelitian**

Adapun tujuan yang ingin dicapai oleh penulis dalam penelitian ini adalah membuat sistem yang dapat mengimplementasikan metode *Naive Bayes Classifier* dalam mengklasifikasikan penyakit *stroke* dengan *data mining* dan mengukur tingkat keberhasilan metode *Naive Bayes Classifier* dalam melakukan klasifikasi penyakit *stroke*.

## **Manfaat Penelitian**

Manfaat yang dapat diambil dari penelitian ini adalah:

1. Manfaat Akademis
2. Penulis dapat menerapkan ilmu pengetahuan di bidang informatika pada penerapan praktis.
3. Dapat dijadikan bahan referensi bagi penelitian-penelitian berikutnya yang memiliki topik yang sama dengan penelitian ini.
4. Manfaat Praktis
5. Mempermudah pihak rumah sakit untuk mengolah data yang besar dan memilih data yang dapat bermanfaat yang berkaitan dengan penyakit *stroke*.
6. Melakukan penanganan penyakit *stroke* dengan cepat dan tepat agar tingkat kematian penderita *stroke* dapat dikurangi jumlahnya.

## **Sistematika Penulisan**

Adapun sistematika penulisan dalam proposal ini adalah sebagai berikut:

BAB I: PENDAHULUAN

Bab ini terdiri dari latar belakang, rumusan masalah, tujuan, manfaat dan sistematika penulisan.

BAB II: TINJAUAN PUSTAKA DAN LANDASAN TEORI

Bab ini menjelaskan tentang perbandingan antara penelitian ini dengan penelitian lain yang pernah ada sebelumnya dan teori-teori yang terkait dengan penyelesaian masalah yang diambil.

BAB III: METODE PENELITIAN

Bab ini menjelaskan tentang alat dan bahan yang akan digunakan dalam penelitian, tahapan penelitian dan hipotesis yang dibuat berdasarkan teori dasar serta rumusan masalah yang ada.

BAB IV: HASIL DAN PEMBAHASAN

Bab ini menjelaskan tentang hasil dari penelitian dan perancangan yang telah dilakukan selama beberapa bulan ini.

BAB V: KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang diambil berdasarkan analisa dan penelitian yang telah dilakukan serta saran bagi penelitian berikutnya di kemudian hari.

DAFTAR PUSTAKA

Berisi sumber bacaan dan kutipan yang digunakan oleh peneliti dalam melakukan penelitian.

# **BAB II**

**TINJAUAN PUSTAKA DAN LANDASAN TEORI**

## **2.1 Tinjauan Pustaka**

Penelitian yang sebelumnya pernah dilakukan sehubungan dengan klasifikasi *data mining* antara lain adalah sebagai berikut:

1. Dalam penelitiannya, Abdilah (2015) membuat diagnosa penyakit *stroke* dengan algoritma *Decision Tree c4.5* dengan klasifikasi *data mining* pada rumah sakit Santa Maria Pemalang. Penelitian ini bertujuan untuk mengklasifikasikan penyakit *stroke* dan *non-stroke*. Persamaan penelitian ini dengan penelitian yang penulis lakukan adalah keduanya melakukan klasifkasi penyakit *stroke* sedangkan perbedaanya terdapat pada metode yang digunakan yaitu penulis menggunakan *Naive Bayes Classifier* sedangkan dalam penelitian sebelumnya menggunakan metode *decision tree*.
2. Dalam penelitiannya, Arifianto, dkk (2014) membuat klasifikasi penyakit *stroke* berdasarkan kelainan patologis dengan *Learning Vector Quantization.* Diagnosis *stroke* harus dilakukan dengan cepat dan tepat agar segera mengetahui tipe klasifikasi patologisnya termasuk dalam *stroke* *infark* atau *hemorrhagic* guna pemberian tindakan medis dan obat yang tepat pula. Prosedur wajib atau *Gold Standart Procedure* untuk klasifikasi *stroke* menggunakan *Computed Tomograph Scan* atau *Magnetic Resonance Imaging*, permasalahannya di Indonesia terkendala biaya yang mahal dan tidak semua rumah sakit memilikinya. Jika prosedur standar tidak dapat dilakukan maka diagnosis *stroke* dapat dilakukan melalui analisis terhadap data klinis pasien. Persamaan penelitian ini dengan penelitian yang dilakukan adalah keduanya melakukan klasifkasi penyakit *stroke* berdasarkan kelainan patologis sedangkan perbedaanya terdapat pada metode yang digunakan yaitu penulis menggunakan *Naive Bayes Classifier* sedangkan dalam penelitian sebelumnya menggunakan metode *Learning Vector Quantization*.
3. Dalam penelitiannya, Mukhlis (2011) membuat Diagnosa Kemungkinan Pasien Terkena Stroke Dengan Menggunakan Metode Naive Bayes Clasifier dan Metode Jaringan Syaraf Tiruan Berbasis Web. Terdapat banyak faktor resiko yang dapat mengarah pada *stroke* seperti umur, jenis kelamin, penyakit diabetes, penyakit darah tinggi, dll. Berdasarkan banyaknya variabel dari faktor resiko tersebut maka dapat digunakan metode *Naive Bayes* dan metode jaringan syaraf tiruan yang digunakan untuk mengolah data-data tentang faktor resiko penderita *stroke*. Data tersebut digunakan sebagai data training untuk proses pembelajaran dari sistem yang akan dibuat untuk menentukan *suspect* *Stroke* atau tidak. Persamaan penelitian ini dengan penelitian yang dilakukan adalah menggunakan metode *Naive Bayes Classifier* sedangkan perbedaan dengan penelitian sebelumnya adalah peneliti mengklasifikasi penyakit *stroke* berdasarkan kelainan patologis dan penelitian sebelumnya mendiagnosa kemungkinan pasien terkena *stroke* atau tidak.

Pada penelitian ini akan dibuat sebuah aplikasi *data mining* untuk mengklasifikasikan penderita *stroke* berdasarkan fitur-fitur dari tes kesehatan pasien. Dengan adanya penelitian ini maka diharapkan agar penanganan penyakit *stroke* dapat dilakukan secepatnya serta resiko penyakit *stroke*  dapat diturunkan jika telah positif memiliki gejala-gejala tertentu.

## **2.2 Landasan Teori**

### **2.2.1 *Data Mining***

*Data mining* merupakan proses pencarian pola dan relasi-relasi yang tersembunyi dalam sejumlah data yang besar dengan tujuan untuk melakukan klasifikasi, estimasi, prediksi, *association rule*, *clustering*, dan deskripsi. Berdasarkan aktifitasnya *data mining* dikelompokkan menjadi dua jenis, yaitu *directed* *data mining* dan *undirected data mining*. *Directed* *data mining* digunakan jika sudah diketahui secara pasti apa yang akan diprediksi, sehingga bisa secara langsung menambang data untuk diarahkan pada tujuan tertentu. Misalnya model prediktif yang digunakan untuk membuat prediksi tentang diagnosa penyakit yang belum diketahui. Model prediktif menggunakan pengalaman untuk menentukan bobot dan tingkat kepercayaan. Salah satu kunci keberhasilan model prediktif adalah adanya data yang cukup dengan hasil yang sudah diketahui untuk mengarahkan/melatih model.

*Undirected data mining* berkaitan dengan menelusuri pola-pola baru dalam data. Tidak seperti *directed* *data mining*, yang sudah mengetahui apa yang akan diprediksi. Pada *undirected* *data mining*, ingin diketahui bagaimana model mengusulkan jawaban. Dalam prakteknya *data mining* sering berisi kombinasi dari keduanya. Misalnya saat membangun *predictive model*, sering berguna untuk mencari pola dalam data menggunakan teknik *undirected*.

*Data mining* merupakan proses menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidenfikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai *database* besar/*Data Warehouse* (Turban, dkk. 2005)

### **2.2.2 Fungsi *Data Mining***

*Data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Kusrini dan Luthfi, 2009):

1. Deskripsi

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari data untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menentukan keterangan atau fakta bahwa siapa yang tidak cukup professional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik dari pada kearah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel predikasi. Sebagai contoh akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam predikasi nilai dari hasil akan ada di masa mendatang. Contoh prediksi bisnis dan penelitian adalah:

a. Prediksi harga beras dalam tiga bulan yang akan datang.

b. Prediksi persentasi kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikkan. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Contoh lain klasifikasi dalam bisnis dan penelitian adalah:

a. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau tidak.

b. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.

c. Mendiagnosis penyakit seorang pasien untuk mendapatkan termasuk kategori penyakit apa.

5. Pengklusteran (*Clustering*)

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 lainnya dan memiliki ketidakmiripan dengan *record-record* dalam kluster lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (*homogeny*), yang mana kemiripan dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal. Contoh pengklusteran dalam bisnis dan penelitian adalah:

* 1. Mendapatkan kelompok-kelompok konsumen untuk target pemasaran dari satu suatu produk bagi perusahaan yang tidak memiliki dana pemesaran yang besar.
  2. Untuk tujuan audit akuntansi, yaitu melakukan pemisahan terhadap perilaku financial dalam baik dan mencurigakan.

c. Melakukan pengklusteran terhadap ekspresi dari gen, untuk mendapatkan kemiripan perilaku dari gen dalam jumlah besar.

6. Asosiasi

Tugas asosiasi dalam *data mining* adalah menemukan *attribut* yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja. Contoh asosiasi dalam bisnis dan penelitian adalah:

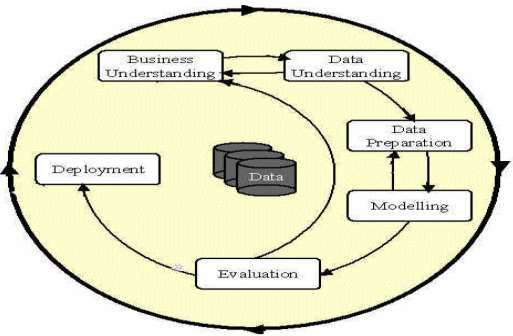
* 1. Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respon positif terhadap penawaran *upgrade* layanan yang diberikan.
  2. Menentukan barang dalam supermarket yang dibeli secara bersamaan dan yang tidak pernah dibeli secara bersamaan.

### **2.2.3 Fase *Cross-Industry Standart Process for Data Mining* (CRISP-DM)**

*Cross-Industry Standart Process for* *Data Mining* (CRISP-DM) yang dikembangkan tahun 1996 oleh analisis dari beberapa industri seperti Daimler Chrysler, SPSS dan NCR. CRISP-DM menyediakan standar proses *data mining* sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian.

Menurut Larose (Kusrini dan Luthfi, 2009), *Cross Industry Standard Process Model for Data Mining* (CRISP-DM) menyediakan proses *data mining* sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian.

Dalam CRISP-DM sebuah proyek *data mining* memiliki siklus hidup yang terbagi dalam enam fase yang dapat dilihat pada Gambar 2.1. Keseluruhan fase berurutan yang ada tersebut bersifat adaptif. Fase berikutnya dalam urutan bergantung kepada keluaran dari fase sebelumnya. Hubungan penting antar fase digambarkan dengan panah. Sebagai contoh, jika proses berada pada fase *modeling*. Berdasar pada perilaku dan karakteristik model, proses mungkin kembali kepada fase *data preparation* untuk perbaikan lebih lanjut terhadap data atau berpindah maju kepada fase *evaluation*.

Gambar 2.1 Fase CRISP-DM

(Sumber: Kusrini dan Luthfi, 2009)

Enam fase dalam CRISP-DM (*Cross Industry Standard Process for* *Data Mining*) (Larose, 2005) :

1. Fase Pemahaman Bisnis (*Business Understanding Phase*)
2. Penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan.
3. Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan *data mining*.
4. Menyiapkan strategi awal untuk mencapai tujuan.
5. Fase Pemahaman Data (*Data Understanding Phase*)
6. Mengumpulkan data.
7. Menggunakan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal.
8. Mengevaluasi kualitas data.
9. Jika diinginkan, pilih sebagian kecil kelompok data yang mungkin mengandung pola dari permasalahan
10. Fase Pengolahan Data (*Data Preparation Phase*)
11. Siapkan dari data awal, kumpulan data yang akan digunakan untuk keseluruhan fase berikutnya. Fase ini merupakan pekerjaan berat yang perlu dilaksanakan secara intensif.
12. Pilih kasus dan variabel yang ingin dianalisis dan yang sesuai analisis yang akan dilakukan.
13. Lakukan perubahan pada beberapa variabel jika dibutuhkan.
14. Siapkan data awal sehingga siap untuk perangkat pemodelan.
15. Fase Pemodelan (*Modeling Phase*)
16. Pilih dan aplikasikan teknik pemodelan yang sesuai.
17. Kalibrasi aturan model untuk mengoptimalkan hasil.
18. Perlu diperhatikan bahwa beberapa teknik mungkin untuk digunakan pada permasalahan data mining yang sama.
19. Jika diperlukan, proses dapat kembali ke fase pengolahan data untuk menjadikan data ke dalam bentuk yang sesuai dengan spesifikasi kebutuhan teknik *data mining* tertentu.
20. Fase Evaluasi (*Evaluation Phase*)
21. Mengevaluasi satu atau lebih model yang digunakan dalam fase pemodelan untuk mendapatkan kualitas dan efektivitas sebelum disebarkan untuk digunakan.
22. Menetapkan apakah terdapat model yang memenuhi tujuan pada fase awal.
23. Menentukan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik.
24. Mengambil keputusan berkaitan dengan penggunaan hasil dari data mining.
25. Fase Penyebaran (Deployment Phase)
26. Menggunakan model yang dihasilkan. Terbentuknya model tidak menandakan telah terselesaikannya proyek.
27. Contoh sederhana penyebaran: Pembuatan laporan.
28. Contoh kompleks Penyebaran: Penerapan proses *data mining* secara paralel pada departemen lain.

### **2.2.4 *Star Schema***

Menurut (Connolly dan Begg, 2010), *star schema* adalah model data dimensional yang mempunyai *fact table* di bagian tengah, dikelilingi oleh tabel dimensi yang terdiri dari data *reference* (yang bisa di-*denormalized*). *Star schema* mengambil karakteristik dari *factual* data yang di-*generate* oleh *event* yang terjadi di masa lampau.

Menurut (Ponniah, 2001), skema bintang (*star schema*) adalah teknik dasar perancangan data untuk *data warehouse*. Struktur skema bintang adalah suatu struktur yang dapat dengan mudah dipahami dan digunakan oleh pengguna. Struktur tersebut mencerminkan bagaimana pengguna biasanya memandang ukuran-ukuran kritis mengikuti dimensi-dimensi bisnis yang ada. Dalam skema bintang tergambar dua jenis tabel, yaitu tabel dimensi dan tabel fakta.

### **2.2.5 *Data Warehouse***

Menurut (Inmon, 2005), *data warehouse* adalah sekumpulan dari data yang *subject-oriented, integrated, time-variant, dan non-volatile* untuk mendukung proses pembuatan keputusan manajemen. Sedangkan menurut (McLeod dan Schell, 2004), *data warehouse* merupakan perkembangan dari konsep *database* yang menyediakan suatu sumber data yang lebih baik bagi para pengguna dan memungkinkan pengguna untuk memanipulasi dan menggunakan data tersebut secara intuitif.

Menurut (Inmon, 2005), *data warehouse* memiliki beberapa karakteristik yaitu:

1. *Subject-Oriented* (Berorientasi Subjek)*, data* *warehouse* diorganisasikan berdasarkan subjek utama dari perusahaan (seperti pelanggan, produk, dan penjualan) dari pada berdasarkan area aplikasi utama (seperti pembuatan faktur pelanggan, pengendalian persedian dan penjualan produk). Hal ini menggambarkan data yang ada didalam *data* *warehouse* merupakan data untuk pengambilan keputusan, bukan data yang berorientasi aplikasi.
2. *Integrated* (Terintegrasi)*,* data didalam *data* *warehouse* berasal dari sumber data yang berbeda dari sistem aplikasi yang berbeda diseluruh perusahaan. Sumber data biasanya digunakan secara tidak konsisten, contohnya format yang berbeda. Sumber data yang terintegrasi harus dibuat konsisten untuk menampilkan pandangan terintegrasi dari data kepada *user.*
3. *Time Variant,* data didalam *data warehouse* hanya akurat dan *valid* dalam suatu waktu tertentu atau dalam interval waktu tertentu.
4. *Non-volatile,* data didalam *data warehouse* tidak di*update* secara *real time* tetapi diperbaharui dari sistem operasional secara berkala. Data baru selalu ditambahkan sebagai tambahan ke dalam *database* bukan sebagai penggantian.

### **2.2.6 *Naive Bayes Clasifier***

1. **Teorema *Bayes***

Teorema *Bayes* adalah teknik prediksi yang berbasis pada probabilitas sederhana yang berdasar pada penerapan teorema *Bayes* (atau aturan *Bayes*) dengan kemungkinan independensi atau ketidaktergantungan yang kuat yang disebut naif. Jadi dalam *Naive Bayes* model yang digunakan adalah fitur independen yaitu independen yang kuat dalam fitur satu dan fitur lainnya pada sebuah data tidak berhubungan dengan ada atau tidaknya fitur lain dalam data yang sama.

Rumus teorema *Bayes* pada umumnya dapat dilihat pada persamaan (1).

…………………………………………………… (1)

Parameter Keterangan

P(H|E) Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis H terjadi jika diberikan bukti (*evidence*) E terjadi.

P(E|H) Probabilitas sebuah bukti E terjadi akan dipengaruhi hipotesis H.

P(H) Probabilitas awal (priori) hipotesis H terjadi tanpa memandang bukti apapun.

P(E) Probabilitas awal (priori) bukti E terjadi tanpa memandang hipotesis/bukti yang lain.

Dasar ide aturan *Bayes* adalah diperkirakan berdasarkan pada beberapa bukti (E) yang diamati didapatkan hasil dari hipotesis atau peristiwa (H). Beberapa hal penting dari aturan *Bayes* adalah:

* + - 1. Sebuah probabilitas awal/prior H atau P(H) adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
      2. Sebuah probabilitas akhir H atau P(H|E) adalah probabilitas dari suatu hipotesis setelah bukti diamati.

1. **Pengklasifikasian *Naive Bayes***

Hubungan antara *Naive Bayes* dengan klasifikasi, korelasi hipotesis dan bukti klasifikasi adalah dalam teorema *Bayes* bahwa hipotesis merupakan label kelas yang menjadi target pemetaan dalam klasifikasi, sedangkan bukti merupakan fitur-fitur yang menjadikan masukkan dalam model klasifikasi. Jika X adalah vektor masukkan yang berisi fitur dan Y adalah label kelas, *Naive Bayes* dituliskan dengan P(Y|X). Notasi tersebut berarti probabilitas label kelas Y didapatkan setelah fitur-fitur X diamati. Notasi ini disebut juga probabilitas akhir (*posterior probability*) untuk Y, sedangkan P(Y) disebut probabilitas awal (*prior probability*) Y.

Selama proses pelatihan data latih yang didapat berdasarkan informasi harus dilakukan dengan pembelajaran probabilitas akhir P(Y|X) pada model untuk setiap kombinasi X dan Y. Dengan membangun model tersebut, suatu data uji X’ dapat diklasifikasikan dengan mencari nilai Y’ dengan memaksimalkan nilai P(X’|Y’) yang didapat.

Rumus *Naive Bayes*  untuk klasifikasi dinyatakan pada persamaan (2).

(2)

P(Y|X) adalah probabilitas data kelas Y jika diberikan data vektor X. P(Y) adalah probabilitas awal kelas Y. adalah perkalian semua probabilitas independen dari masing-masing fitur pada vektor X jika diketahui vektor tersebut berada pada kelas Y. Nilai P(X) selalu tetap sehingga dalam perhitungan prediksi nantinya tinggal dihitung bagian dengan memilih yang terbesar sebagai kelas yang dipilih sebagai hasil prediksi. Sementara probabilitas independen tersebut merupakan pengaruh semua fitur dari data terhadap setiap kelas Y, yang dinotasikan dengan , Setiap set fitur X = {X1, X2, X3,…, Xq} terdiri atas *q* atribut (*q* dimensi).

Untuk fitur dengan tipe numerik (kontinu) maka dilakukan beberapa langkah khusus sebelum dimasukan ke dalam *Naive Bayes* yaitu:

1. Melakukan diskretisasi pada setiap fitur kontinu dan mengganti nilai fitur kontinu tersebut dengan nilai interval diskret. Pendekatan ini dilakukan dengan mentransformasikan fitur kontinu ke dalam fitur ordinal.
2. Mengasumsikan bentuk tertentu dari distribusi probabilitas untuk fitur kontinu dan memperkirakan parameter distribusi dengan data pelatihan.
3. **Karakteristik *Naive Bayes***

Klasifikasi dengan *Naive Bayes* bekerja dengan memandang semua fitur dari data sebagai bukti dalam probabilitas berdasarkan teori probabilitas. Hal ini memberikan karakteristik *Naive Bayes* sebagai berikut:

* + - 1. Metode *Naive Bayes* bekerja teguh (*robust*) terhadap data-data yang terisolasi yang biasanya merupakan data dengan karakteristik berbeda (*outlier*). *Naive Bayes* juga bisa menangani nilai atribut yang salah dengan mengabaikan data latih selama proses pembangunan model dan prediksi.
      2. Tangguh menghadapi atribut yang tidak relevan.
      3. Atribut yang mempunyai korelasi bisa mendegradasi kinerja klasifikasi *Naive Bayes* karena asumsi independensi atribut tersebut sudah tidak ada.

Tabel 2.1 menunjukkan data *training* dari contoh kasus pengklasifikasian penyakit *stroke* berdasarkan fitur/kriteria yang diberikan, sedangkan tabel 2.2 berisi tentang data *testing* dari seorang yang diduga mengidap penyakit *stroke* pada contoh kasus tersebut.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **JENIS KELAMIN** | **KELOMPOK USIA\*)** | **KESADARAN** | **NYERI KEPALA** | **KEJANG** | **JENIS *STROKE* \*\*)** |
| **1** | LAKI-LAKI | 30 thn (Kel. 2) | MENURUN | TIDAK | YA | INF |
| **2** | LAKI -LAKI | 24 thn (Kel. 1) | MENURUN | TIDAK | YA | INF |
| **3** | PEREMPUAN | 28 thn (Kel. 2) | MENURUN | TIDAK | YA | INF |
| **4** | PEREMPUAN | 35 thn (Kel. 2) | TIDAK | YA | TIDAK | INF |
| **5** | LAKI-LAKI | 40 thn (Kel. 2) | TIDAK | YA | TIDAK | HEMO |
| **6** | LAKI-LAKI | 46 thn (Kel. 3) | MENURUN | TIDAK | TIDAK | HEMO |
| **7** | PEREMPUAN | 38 thn (Kel. 2) | MENURUN | YA | YA | HEMO |
| **8** | PEREMPUAN | 50 thn (Kel. 3) | TIDAK | YA | YA | HEMO |
| **9** | LAKI-LAKI | 22 thn (Kel. 1) | TIDAK | TIDAK | YA | INF |
| **10** | PEREMPUAN | 57 thn (Kel. 3) | TIDAK | TIDAK | YA | INF |
| **11** | PEREMPUAN | 47 thn (Kel. 3) | MENURUN | TIDAK | TIDAK | INF |
| **12** | PEREMPUAN | 31 thn (Kel. 2) | MENURUN | TIDAK | TIDAK | INF |
| **13** | LAKI-LAKI | 47 thn (Kel. 3) | MENURUN | YA | TIDAK | HEMO |
| **14** | LAKI-LAKI | 52 thn (Kel. 3) | TIDAK | YA | TIDAK | HEMO |
| **15** | LAKI-LAKI | 44 thn (Kel. 2) | TIDAK | YA | TIDAK | HEMO |

Tabel 2.1 Data *Training*

Keterangan:

\*) Fitur USIA telah didiskretisasi menjadi fitur KELOMPOK USIA sebagai berikut:

Kelompok 1 = usia 25 tahun

Kelompok 2 = 25 tahun usia 45 tahun

Kelompok 3 = usia 45 tahun

\*\*) INF = *Infrak*

HEMO = *Hemoragic*

Tabel 2.2 Data *Testing*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **JENIS KELAMIN** | **KELOMPOK USIA** | **KESADARAN** | **NYERI KEPALA** | **KEJANG** | **JENIS *STROKE*** |
| LAKI-LAKI | 35 thn (Kel. 2) | MENURUN | YA | TIDAK | ? |

Untuk dapat menentukan jenis penyakit *stroke* yang terdapat pada data *testing* di tabel 2.2, maka dilakukan tahapan perhitungan sebagai berikut:

**Tahap 1**: Menghitung probabilitas awal masing-masing kelas/label.

P(Y= INF) = 8/15

Nilai probabilitas ini didapatkan dari jumlah data “INF” pada kolom “JENIS STROKE” dibagi dengan total jumlah data pada kolom tersebut.

P(Y= HEMO) = 7/15

Nilai probabilitas ini didaptkan dari jumlah data “HEMO” pada kolom “JENIS STROKE” dibagi dengan total jumlah data pada kolom tersebut.

**Tahap 2**: Menghitung probabilitas masing-masing fitur (sesuai dengan data yang terdapat pada data *testing*) pada setiap kelas.

P(JENIS KELAMIN = LAKI-LAKI | Y= INF) = 3/8

Nilai probabilitas ini didapatkan dari jumlah data “LAKI-LAKI” pada kolom “JENIS KELAMIN” yang bersesuaian dengan baris “INF” pada kolom “JENIS STROKE” dibagi dengan total jumlah data “INF” pada kolom “JENIS STROKE”.

P(JENIS KELAMIN = LAKI - LAKI | Y= HEMO) = 5/7

Nilai probabilitas ini didapatkan dari jumlah data “LAKI-LAKI” pada kolom “JENIS KELAMIN” yang bersesuaian dengan baris “HEMO” pada kolom “JENIS STROKE” dibagi dengan total jumlah data “HEMO” pada kolom “JENIS STROKE”.

P(KELOMPOK USIA = Kel. 2 | Y= INF) = 4/8

Nilai probabilitas ini didapatkan dari jumlah data “Kel. 2” pada kolom “KELOMPOK USIA” yang bersesuaian dengan baris “INF” pada kolom “JENIS STROKE” dibagi dengan total jumlah data “INF” pada kolom “JENIS STROKE”.

P(KELOMPOK USIA = Kel. 2 | Y= HEMO) = 3/7

Nilai probabilitas ini didapatkan dari jumlah data “Kel. 2” pada kolom “KELOMPOK USIA” yang bersesuaian dengan baris “HEMO” pada kolom “JENIS STROKE” dibagi dengan total jumlah data “HEMO” pada kolom “JENIS STROKE”.

P(KESADARAN = MENURUN | Y= INF) = 5/8

Nilai probabilitas ini didapatkan dari jumlah data “MENURUN” pada kolom “KESADARAN” yang bersesuaian dengan baris “INF” pada kolom “JENIS STROKE” dibagi dengan total jumlah data “INF” pada kolom “JENIS STROKE”.

P(KESADARAN = MENURUN | Y= HEMO) = 3/7

Nilai probabilitas ini didapatkan dari jumlah data “MENURUN” pada kolom “KESADARAN” yang bersesuaian dengan baris “HEMO” pada kolom “JENIS STROKE” dibagi dengan total jumlah data “HEMO” pada kolom “JENIS STROKE”.

P(NYERI KEPALA = YA | Y= INF) = 1/8

Nilai probabilitas ini didapatkan dari jumlah data “YA” pada kolom “NYERI KEPALA” yang bersesuaian dengan baris “INF” pada kolom “JENIS STROKE” dibagi dengan total jumlah data “INF” pada kolom “JENIS STROKE”.

P(NYERI KEPALA = YA | Y= HEMO) = 6/7

Nilai probabilitas ini didapatkan dari jumlah data “YA” pada kolom “NYERI KEPALA” yang bersesuaian dengan baris “HEMO” pada kolom “JENIS STROKE” dibagi dengan total jumlah data “HEMO” pada kolom “JENIS STROKE”.

P(KEJANG = TIDAK | Y= INF) = 3/8

Nilai probabilitas ini didapatkan dari jumlah data “TIDAK” pada kolom “KEJANG” yang bersesuaian dengan baris “INF” pada kolom “JENIS STROKE” dibagi dengan total jumlah data “INF” pada kolom “JENIS STROKE”.

P(KEJANG = TIDAK| Y= HEMO) = 5/7

Nilai probabilitas ini didapatkan dari jumlah data “TIDAK” pada kolom “KEJANG” yang bersesuaian dengan baris “HEMO” pada kolom “JENIS STROKE” dibagi dengan total jumlah data “HEMO” pada kolom “JENIS STROKE”.

**Tahap 3**: Mengalikan semua hasil perhitungan pada tahap 1 dan 2 untuk masing-masing kelas (INF dan HEMO)

Aproksimasi nilai probabilitas untuk jenis *stroke* dengan kelas INF dapat dihitung dari perkalian berikut:

P(Y = INF) P(JENIS KELAMIN = LAKI–LAKI | Y = INF) P(KELOMPOK USIA = Kel. 2 | Y = INF) P(KESADARAN = MENURUN | Y = INF) P(NYERI KEPALA = YA | Y = INF) P(KEJANG = TIDAK | Y = INF)

= 8/15 3/8 4/8 5/8 1/8 3/8

= 0,003

Aproksimasi nilai probabilitas untuk jenis *stroke* dengan kelas HEMO dapat dihitung dari perkalian berikut:

P(Y = HEMO) P(JENIS KELAMIN = LAKI–LAKI | Y = HEMO) P(KELOMPOK USIA = Kel. 2 | Y = HEMO) P(KESADARAN = MENURUN | Y = HEMO) P(NYERI KEPALA = YA | Y = HEMO) P(KEJANG = TIDAK | Y = HEMO)

= 7/15 5/7 3/7 3/7 6/7 5/7

= 0,037

**Tahap 4**: Membandingkan hasil nilai probabilitas posterior kelasINF dan HEMO

Karena aproksimasi nilai probabilitas posterior P(Y = INF) lebih kecil dari aproksimasi nilai probabilitas posterior P(Y = HEMO) maka keputusanya adalah data *testing* masuk ke dalam kelas “HEMO”.

Hasil perhitungan klasifikasi data testing yang terdapat pada tabel 2.2, dapat dilihat pada tabel 2.3.

Tabel 2.3. Hasil

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **JENIS KELAMIN** | **KELOMPOK USIA** | **KESADARAN** | **NYERI KEPALA** | **KEJANG** | **JENIS *STROKE*** |
| LAKI-LAKI | 35 thn (Kel. 2) | MENURUN | YA | TIDAK | HEMO |

### **2.2.7.** **Penentuan Tingkat Keberhasilan Sistem**

Untuk mengetahui tingkat keberhasilan suatu sistem yang telah dibuat, maka dapat dilakukan pengujian terhadap sistem tersebut untuk mendapatkan besarnya persentasi data benar yang dihasilkan oleh sistem. Pada penelitian ini, pengujian dilakukan dengan menggunakan data latih (data *training*) yang sebelumnya juga sudah digunakan untuk membentuk sistem. Penentuan persentase keberhasilan sistem dilakukan dengan membandingkan jumlah data benar (sesuai dengan data latih yang digunakan untuk membentuk sistem) yang dihasilkan oleh sistem dengan jumlah total data yang dihasilkan oleh sistem, dimana selanjutnya hasil perbandingan ini dikalikan dengan 100%. Adapun rumus penentuan persentase keberhasilan sistem dapat dilihat pada persamaan (3).

……………………………... (3)

Keterangan:

Persentase Kebrhasilan = Persentase keberhasilan yang didapatkan dari hasil pengujian terhadap sistem yang telah dibuat.

Data Benar = Jumlah data benar yang sesuai dengan data latih pembentukan sistem, yang dihasilkan oleh sistem selama masa pengujian.

Total Data = Jumlah total data yang dihasilkan oleh sistem selama masa pengujian.

### **2.2.8 Penyakit *Stroke***

*Stroke* atau *Cerebrovaskular diseas* adalah gangguan fungsi pada saraf yang disebabkan oleh gangguan aliran darah ke otak yang terhenti akibat sumbatan atau pecahnya pembuluh darah yang muncul secara mendadak dan sangat cepat.

Menurut (Misbach, 1999), klasifikasi *stroke* antara lain:

1. Berdasarkan kelainan patologis anatomi dan penyebabnya:

a. *Stroke* Iskemik (Infark) yang disebabkan karena adanya hambatan atau sumbatan pada pembuluh darah otak tertentu sehingga daerah otak yang diperdarahi oleh pembuluh darah tersebut tidak mendapat pasokan energi dan oksigen, sehingga pada akhirnya jaringan sel-sel otak di daerah tersebut mati dan tidak berfungsi lagi.

* *Transient Ischemic Attack* (TIA) adalah serangan yang terjadi saat pasokan darah ke otak mengalami gangguan sesaat. Serangan ini umumnya berlangsung lebih singkat dari *stroke*, yaitu selama beberapa menit hingga beberapa jam dan akan pulih dalam 1 hari.
* Trombosis serebri merupakan salah satu penyakit pembuluh darah otak berupa penyumbatan pembuluh darah otak oleh karena thrombus yang menyebabkan iskemik atau infark jaringan yang menimbulkan gejala disfungsi otak lokal dengan defisit neurologi yang menetap atau sembuh dengan gejala sisa.
* Emboli serebri suatu kondisi di mana aliran darah terhambat akibat benda asing (embolus), seperti bekuan darah atau udara. Benda asing yang berada di dalam akliran darah akan ikut bersirkulasi sampai terhambat pada salah satu pembuluh darah. Hambatan ini menyebabkan aliran darah menjadi tidak lancar dan jaringan tubuh kekurangan oksigen sehingga menyebabkan kegagalan fungsi organ.

b. *Stroke* hemoragik adalah *stroke* yang terjadi bila pasokan darah ke otak Anda terganggu akibat pembuluh darah pecah dan berdarah di dalam otak. Otak mengalami pendarahan dan darah menekan otak sehingga mengakibat gangguan di seluruh tubuh.

* Perdarahan intraserebral yaitu *stroke* disebabkan oleh pendarahan di dalam otak.
* Perdarahan subarakhnoid *stroke* disebabkan oleh pendarahan di permukaan otak dalam ruang subarachnoid (ini dibentuk oleh dua lapisan membran di antara otak dan tulang tengkorak).

2. Berdasarkan stadium atau pertimbangan waktu:

a. *Transient Ischemic Attack* (TIA) Serangan ini umumnya berlangsung lebih singkat dari *stroke*, yaitu selama beberapa menit hingga beberapa jam dan akan pulih dalam 1 hari.

b. *Reversible Ischemic Neurologic Deficit* (RIND) gejala neurologis yang ada pada RIND akan menghilang antara 24 jam sampai 21 hari. Jika  pada TIA dokter jarang menyaksikan sendiri sehingga diagnosis ditegakkan berdasar riwayat dari pasien, maka RIND besar kemungkinan dokter masih dapat mengamati

c. *Progressing Stroke* atau *Stroke in evolution* Pada keadaan ini kelainan yang ada masih terus berkembang ke arah yang lebih berat. Diagnosis *stroke* ditegakkan pada saat awal dengan mengamati langsung perkembangannya, atau saat akhir dimana diambil riwayat penyakitnya.

d. *Completed stroke* diartikan bahwa kelainan neurologis yang ada  sifatnya sudah menetap, tidak berkembang lagi. Kelainan neurologis yang muncul bermacam-macam tergantung daerah otak yang mengalami lesi.

3. Berdasarkan sistem pembuluh darah:

a. Sistem karotis aliran darah ke otak yang melalui arteri karotis interna beserta cabang-cabangnya.

b. Sistem vertebrobasiler aliran darah ke otak yang melalui arteri vertebralis berserta cabang-cabangnya.

### **2.2.8 Rekam Medis**

Dalam Peraturan Menteri Kesehatan Republik Indonesia Nomor 269/Menkes/Per/III/2008 Tentang Rekam Medis pada Bab I Pasal 1a yaitu Rekam medis adalah berkas yang berisikan catatan dan dokumen tentang identitas pasien, pemeriksaan, pengobatan, tindakan dan pelayanan lain yang telah diberikan kepada pasien.

# **BAB III**

**METODE PENELITIAN**

## **3.1. Bahan dan Alat Penelitian**

### **3.1.1 Bahan** **Penelitian**

Bahan yang digunakan untuk penelitian ini adalah data rekam medik pasien penderita penyakit *stroke* di Rumah Sakit Umum Daerah Undata Palu.

### **3.1.2 Alat Penelitian**

Alat yang digunakan untuk penelitian ini yaitu sebuah laptop dengan spesifikasi Intel®CoreTm i3 2330M, 14.0” HD LED LCD, Intel® HD Graphics 3000, 2 GB DDR3 Memory, 500 GB HDD, dengan sistem operasi Windows 7,MySQL, dan software *Visual Studio* 2013.

## **3.2. Desain Penelitian**

### **3.2.1. Jenis Penelitian**

Jenis penelitian yang digunakan dalam penelitian ini yaitu penelitian terapan karena tujuan penelitian ini dimaksudkan untuk menerapkan metode *data mining* dalam mengklasifikasikan penyakit *stroke* berdasarkan kelainan patologis untuk menghasilkan sebuah aplikasi. Penelitian terapan dilakukan dengan tujuan menerapkan, menguji, dan mengevaluasi kemampuan suatu teori yang diterapkan dalam memecahkan masalah-masalah praktis (Sugiyono, 2002).

### **Tipe Penelitian**

Tipe penelitian yang digunakan adalah penelitian kualitatif berbasis ekperimen karena tipe penelitian eksperimen merupakan tipe penelitian dengan menguji sebab dan akibat dalam sebuah penelitian dengan melakukan perbandingan antara objek yang satu dengan objek yang lain.

* 1. **Objek, Waktu dan Lokasi Penelitian**

Penelitian ini dilaksanakan pada bulan September 2016 sampai dengan bulan Juli 2017 bertempat di Rumah Sakit Umum Daerah Undata Palu, Sulawesi Tengah.

## **3.4. Jenis dan Sumber Data**

Jenis data yang digunakan pada penelitian ini adalah data primer dan data sekunder, yang mana data primer adalah data yang diperloeh secara langsung di lapangan dan data sekunder adalah data yang diperoleh dari pihak yang diperlukan datanya. Sumber data pada penelitian ini adalah data hasil rekam medik pasien pada Rumah Sakit Umum Daerah Undata Palu.

## **3.5. Teknik Pengumpulan Data**

Teknik pengumpulan data yang tepat yaitu dengan mempertimbangkan penggunaannya berdasarkan jenis data dan sumbernya. Data yang objektif dan relevan dengan pokok permasalahan penelitian merupakan indikator keberhasilan suatu penelitian. Pengumpulan data penelitian ini dilakukan dengan cara sebagai berikut:

1. Observasi

Observasi adalah suatu cara pengumpulan data dengan pengamatan langsung dan pencatatan secara sistematis terhadap obyek yang akan diteliti. Observasi dilakukan di Rumah Sakit Umum Daerah Undata Palu dengan melakukan kunjungan pada rumah sakit tersebut.

1. Wawancara

Wawancara atau *interview* adalah suatu cara untuk mengumpulkan data dengan mengajukan pertanyaan langsung kepada seorang narasumber atau seorang ahli atau yang berwenang dalam suatu masalah. Pertanyaan-pertanyaan yang diajukan biasanya disiapkan terlebih dahulu yang diarahkan kepada informasi-informasi untuk topik yang akan teliti. Dalam hal ini peneliti melakukan wawancara dengan dokter ahli saraf bernama dr. Isnaniah, Sp.S mengenai penelitian yang akan dilakukan.

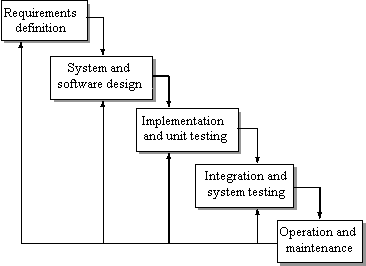
1. Studi Pustaka

Studi pustaka adalah segala usaha yang dilakukan oleh peneliti untuk menghimpun informasi yang relevan dengan topik atau masalah yang akan atau sedang diteliti. Informasi itu dapat diperoleh dari buku-buku ilmiah, laporan penelitian, karangan-karangan ilmiah, tesis dan disertasi, peraturan-peraturan, ketetapan-ketetapan, buku tahunan, ensiklopedia, dan sumber-sumber tertulis baik tercetak maupun elektronik.

## **3.6. Metode Analisis Data**

Metode analisis data adalah proses pencarian dan penyusunan data secara sistematis terhadap sesuatu untuk mendapatkan pola data yang telah disusun. Dalam penelitian ini digunakan beberapa diagram yang dapat menjelaskan penggunaan dan cara kerja sistem. Diagram yang digunakan antara lain adalah DFD (Data *Flow* Diagram), *Flowchart*, *Use Case* Diagram dan *Class* Diagram.

## **Metode Pengembangan Sistem**

Pembangunan perangkat lunak dalam penilitian ini menggunakan model *waterfall* (Sommerville, 2011), dimana alur prosesnya adalah sebagai berikut: 

Gambar 3.1. *Metode Waterfall*

(Sumber: Sommerville, 2011)

Model *waterfall* seperti yang digambarkan pada gambar 3.1 terdapat beberapa proses yang dilalui. Dalam model *waterfall* Sommerville terdapat kemungkinan untuk kembali ke tahap sebelumnya apabila terjadi kesalahan atau perbaikan.

a. *Requirement Definition,* Tahap awal dimana adanya analisis untuk menentukan kebutuhan, batasan, dan tujuan (*goal*) dari perangkat lunak sesuai yang diinginkan. Hal tersebut kemudian didefinisikan secara rinci dan terbentuk sebagai spesifikasi sistem. Pada tahap ini penulis melakukan studi literatur dengan mencari referensi tentang metode *naive bayes classifier*, *data mining* serta penyakit *stroke*.

b. *System and Software Design, System design* merupakan proses perancangan perangkat keras maupun perangkat lunak yang dilibatkan untuk menunjang sistem yang akan dibangun. Sementara itu *software design* merupakan proses perancangan yang melibatkan identifikasi dan menggambarkan dasar sistem serta hubungan satu sama lain. Pada tahap ini penulis membuat *design* aplikasi agar tampilan dari aplikasi ini nantinya dapat digunakan dengan mudah oleh *user* serta memiliki *design* yang menarik untuk digunakan.

c. *Implementation and Unit Testing,* Pada tahap ini, *software design* yang telah dilakukan sebelumnya kemudian diimplementasikan dalam bentuk unit program. Setelah unit program dibuat, kemudian dilakukan *testing* pada unit program tersebut untuk memastikan implementasi berjalan dengan baik. Pada tahap ini penulis menerapkan metode *naive bayes* dan kode-kode program yang mendukung aplikasi serta melakukan pengujian terhadap *form-form* yang telah dibuat untuk mengetahui apakah setiap *form* berjalan sesuai dengan yang diharapkan.

d. *Integration and System Testing,* Setelah semua unit program berhasil diimplementasikan dan lolos *testing* maka dilanjutkan dengan mengintegrasikan setiap unit untuk membentuk sistem yang diinginkan. Sistem yang sudah dibentuk kemudian di tes kembali untuk memastikan unit program dapat berjalan satu sama lain dalam sistem dan sistem yang dibuat sudah memenuhi kebutuhan. Pada tahap ini dilakukan penyatuan semua *form* yang telah dibuat sebelumnya menjadi sistem yang utuh, serta melakukan pengujian terhadap sistem untuk mengetahui apakah sistem berjalan dengan baik atau tidak.

e. *Operation and Maintenance,* Tahap ini merupakan tahap dimana sistem sudah dipasang kemudian melakukan perbaikan ketika terdapat kesalahan atau *error* yang tidak ditemukan sebelumnya saat pembangunan sistem berlangsung. Perbaikan juga dilakukan jika terdapat kebutuhan baru yang perlu ada pada sistem. Tahapan ini dilakukan untuk memperbaiki dan memaksimalkan setiap fungsi yang terdapat dalam sistem jika sistem masih memiliki kekurangan.

## **3.8 Tahapan dan Diagram Alir Penelitian**

Terdapat beberapa tahapan yang perlu dilakukan agar penelitian ini dapat berjalan dengan baik. Tahapan-tahapan tersebut adalah sebagai berikut:

1. Studi literatur

Studi litelatur adalah mencari referensi terlebih dahulu agar lebih memahami metode serta masalah yang diangkat dalam skripsi.

1. Pengumpulan Data

Data yang digunakan dari penelitian ini adalah data rekam medik pasien yang diperoleh dari Rumah Sakit Umum Daerah Undata Palu.

1. *Pre-Processing*

*Pre-Processing* adalah tahapan pengolahan data mentah dengan prosedur yang lebih mudah dan efektif untuk menghasilkan data yang siap digunakan. Tahapan yang di maksud antara lain dapat dilihat pada Gambar 3.2 *Flowchart* Tahapan Data *Mining* dengan langkah-langkah sebagai berikut:

* + - * 1. Seleksi Data

Data yang ada diperoleh sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

* 1. Pembersihan Data

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten. Pada umumnya data yang diperoleh, baik dari *database* suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik.

* 1. Transformasi Data

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*.

* 1. Proses *Mining*

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

* 1. Evaluasi Pola

Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses *data mining*.

* 1. Presentasi Pengetahuan

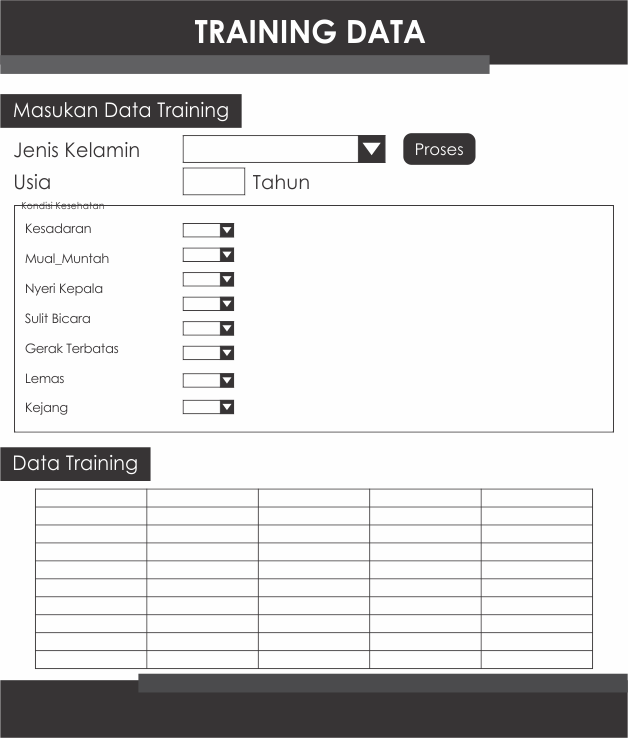
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat.



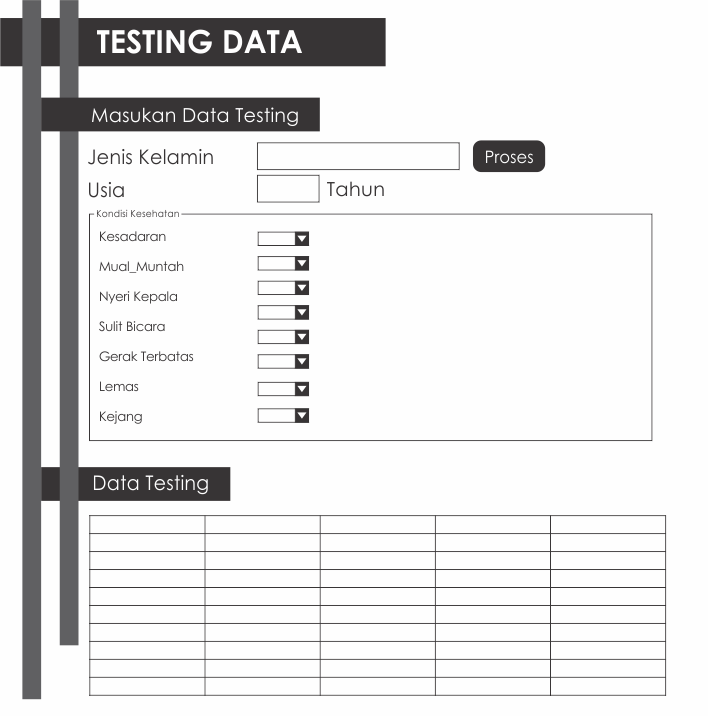
Gambar 3.2. *Flowchart* Tahapan Data *Mining*

1. Perancangan Aplikasi

Perancangan aplikasi dilakukan dengan membuat *form-form* rencana tampilan yang akan diterapkan pada aplikasi yang dibuat dalam penelitian seperti yang terlihat pada gambar 3.3 *Form Training* dan gambar 3.4 *Form Testing*.



Gambar 3.3. *Form Training*



Gambar 3.4. *Form Testing*

Pada gambar 3.3. yang merupakan *Form* untuk *training* data maka pilih Jenis Kelamin pasien, lalu ketik Usia pasien pada kotak, kemudian mengisi kondisi kesehatan yang ada dibawah dengan jawaban ya atau tidak dengan memilih pada *combobox*. Kemudian dilakukan dilakukan proses dengan menekan *button* “Proses”. Langkah yang sama juga dilakukan pada *Form Testing* pada gambar 3.4.

1. Pembuatan Aplikasi

Pembuatan aplikasi yaitu membuat desain dan pengkodean untuk membuat sistem klasifikasi penyakit *stroke* dengan metode *naive bayes classifier*.

1. Memasukan Data

Data yang telah diolah dimasukan ke dalam aplikasi untuk melakukan *training* data.

1. Pengujian dan Analisis

Setelah dilakukan *training* data maka dilakukan pengujian dan analisis apaka sistem dapat berjalan dengan baik untuk melakukan klasifikasi.

Tahapan-tahapan di atas dapat digambarkan pada diagram alir yang ditunjukan pada gambar 3.5 di bawah ini.



Gambar 3.3. *Flowchart* Tahapan Penelitian

# **BAB IV**

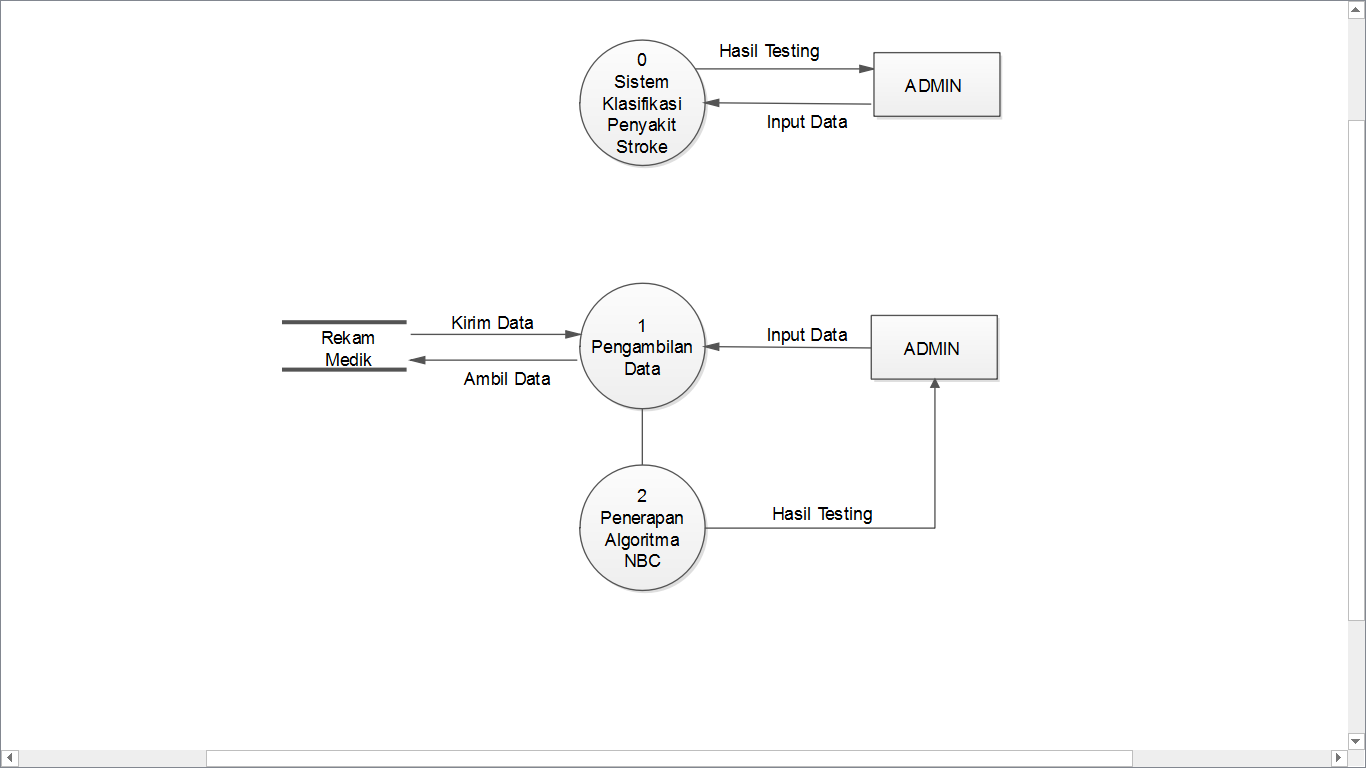
**HASIL DAN PEMBAHASAN**

## **4.1. Hasil Penelitian**

### **4.1.1. Analisa Sistem**

1. Diagram Konteks

Menurut (Jogiyanto, 2010), Diagram Konteks adalah diagram yang menggambarkan keseluruhan sistem yang ada dengan menunjukan sumber serta tujuan data yang akan diproses.



Gambar 4.1 Diagram Konteks Sistem Klasifikasi Penyakit *Stroke*

Diagram konteks Sistem Klasifikasi Penyakit *Stroke* yang telah penulis buat pada skripsi ini dapat dilihat pada Gambar 4.1. dimana keterangannya adalah sebagai berikut:

1. Admin

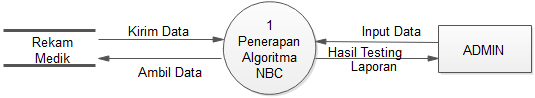
Admin yang di maksud pada diagram di atas adalah admin yang bekerja di Rumah Sakit yang melakukan *input* data pasien ke dalam sistem.

1. Sistem Klasifikasi Penyakit Stroke

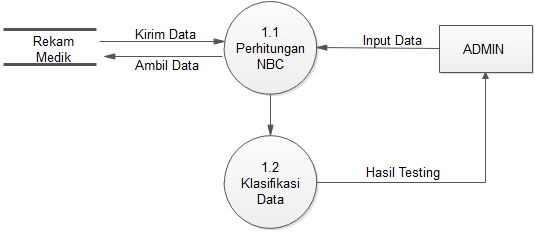
Sistem klasifikasi penyakit stroke yang dibuat di maksudkan untuk mengolah data pada *database* dan memberikan hasil klasifikasi dari data yang baru.

2. *Data Flow Diagram* (DFD)

Menurut (Jogiyanto, 2010), *Data Flow Diagram* (DFD) digunakan untuk menggambarkan sistem baru atau sistem yang telah ada sebelumnya yang dibuat secara logika tanpa melihat lingkungan fisik dimana data tersebut mengalir atau lingkungan fisik dimana data tersebut akan disimpan.



Gambar 4.2. DFD level 1 Proses pengambilan data



Gambar 4.3 DFD level 2 Proses Klasifikasi Data

*Data Flow Diagram*  (DFD) level 1 dapat dilihat pada gambar 4.2. sementara DFD level 2 dapat dilihat pada gambar 4.3. adapun keterangan gambar tersebut adlah sebagai berikut:

1. Admin

Admin yang memasukan data ke dalam sistem agar sistem dalam melakukan proses klasifikasi.

1. Rekam Medik

Rekam medik merupakan *database* yang menyimpan data pasien ataupun data gejala-gejala penyakit pasien yang di dapat dari rekam medik. Data-data tersebut diambil kedalam sistem untuk dapat diolah datanya.

1. Penerapan Algoritma *Naive Bayes Classifier* (NBC)

Setelah data diambil dari *database*,kemudian dilakukan tahapan berikutnya yaitu penerapan algoritma untuk menghasilkan data yang telah diklasifikasikan.

1. Klasifikasi Data

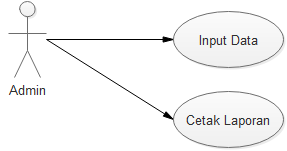
Data yang telah diambil dari *database* telah diklasifikasikan, dimana dari hasil klasifikasi ini akan terlihat data yang telah melalui proses pembelajaran. Setelah proses klasifikasi data selesai dilakukan, maka selanjutnya sistem siap untuk melakukan pengelompokkan jika ada data baru yang masuk.

1. Perhitungan *Naive Bayes Classifier* (NBC)

Dalam tahap ini memasukan rumus *Naive Bayes Classifier* yaitu pada persamaan (2) untuk mendapatkan hasil pengelompokan dari data yang baru dimasukkan dimana data baru ini disebut sebagai *data testing*.

3. *Usecase Diagram*

Menurut (Sukamto dan Shalahuddin, 2011), *Use case* atau diagram *use case* merupakan pemodelan untuk kelakuan *(behavior)* sistem informasi yang akandibuat. *Use case* mendeskripsikan sebuah interaksi antara satu atau lebih actordengan sistem informasi yang akan dibuat.



Gambar 4.4. *Usecase Diagram* sistem klasifikasi penyakit stroke

Dari gambar 4.4. dapat dilihat keterangan dari gambar *usecase diagram* sebagai berikut:

* 1. InputData

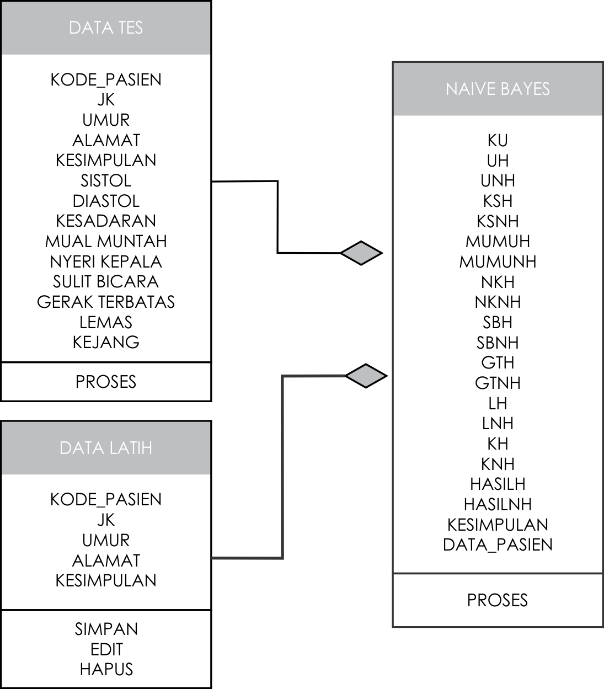
Pada proses input data, Admin disini memiliki tugas untuk mengurus sistem dengan melakukan input data-data pasien.

* 1. Cetak laporan

Pada proses cetak laporan, Admin bertugas untuk melakukan cetak laporan yang ada di dalam sistem. Menghasilkan output laporan data *training* dan data *testing*.

4. *Class Diagram*

(Sukamto dan Shalahuddin, 2013), Diagram kelas atau *class diagram* menggambarkan struktur sistem dari segi pendefinisian kelas-kelas yang akandibuat untuk membangun sistem.



Gambar 4.5. *Class diagram*  sistem klasifikasi penyakit *stroke*

1. Data Latih

Kelas ini merupakan kelas untuk menyimpan data latih atau *training data* yang menjadi tolak ukur dalam melakukan klasifikasi. Data latih merupakan data pasien yang telah diketahui riwayat penyakitnya dan dapat dijadikan sabagai parameter untuk menghitung data baru yang akan di klasifikasikan ke dalam kelas tertentu.

1. Data Test

Kelas ini digunakan untuk menginput data baru atau data yang belum diketahui kesimpulan atau kelasnya, sehingga dipakai parameter dari data latih untuk dapat menemukan hasil dari data baru dengan kelas yang sesuai.

1. Naive Bayes

Kelas ini merupakan proses *Naive Bayes Classifier* dimana dilakukan perhitungan probabilitas untuk kemungkinan kelas diterapkan. Proses untuk menghitung kemungkinan dari kedua kelas yang ada dan mengklasifikasikan kedalam kelas yang sesuai.

### **4.1.2. Implementasi Sistem**

#### **4.1.2.1. Implementasi Perangkat Keras dan Perangkat Lunak**

Agar sistem klasifikasi penyakit stroke dapat dioptimalkan maka perangkat keras yang digunakan untuk pembuatan sistem ini adalah sebuah laptop dengan spesifikasi Intel®CoreTm i3 2330M, 14.0” HD LED LCD, Intel® HD Graphics 3000, 2 GB DDR3 Memory, 500 GB HDD, dengan sistem operasi Windows 7.

Perangkat lunak yang digunakan dalam pembuatan sistem ini yaitu Bahasa pemrograman *Visual Basic* 2013 sebagai applikasi pembuatan dan perancangan sistem dengan *MySQL* sebagai *database* yang digunakan.

#### **4.1.2.2. Implementasi Basis Data**

Basis data atau *database* yang digunakan dalam pembuatan aplikasi yaitu *MySQL*. Dengan nama *database* “rekammedik” yang memiliki beberapa tabel yang digunakan dalam mengkalsifikasi penyakit *stroke* antara lain sebagai berikut:

1. Tabel Data Latih

Tabel data\_latih digunakan untuk menyimpan data pasien beserta beberapa informasi umum (kecuali nama pasien) yang terdapat pada data rekam medik di Rumah Sakit Umum Daerah Undata Palu, serta sebual *field* yang berisi kesimpulan yaitu hasil analisa penyakit pasien yang terkena HEMO (*Hemoragik*) atau INF (*Infark*). Tabel data latih ditunjukan pada Tabel 4.1.

Tabel 4.1 Data Latih

|  |  |  |  |
| --- | --- | --- | --- |
| *Field* | *Type* | *Length*/*Values* | *Index* |
| kode\_pasien | Varchar | 10 | Primary |
| Jk | Varchar | 2 | - |
| Umur | Varchar | 5 | - |
| Alamat | Varchar | 50 | - |
| Kesimpulan | Varchar | 9 | - |
| Sistol | INT | 5 | - |
| Distol | INT | 5 | - |
| Kesadaran | Varchar | 11 | - |
| mual\_muntah | Varchar | 5 | - |
| nyeri\_kepala | Varchar | 5 | - |
| sulit\_bicara | Varchar | 5 | - |
| gerak\_terbatas | Varchar | 5 | - |
| Lemas | Varchar | 5 | - |
| Kejang | Varchar | 5 | - |

1. Tabel Data Test

Tabel data\_test digunakan untuk menyimpan data baru yang masuk, data baru ini akan diklasifikasin pada kelas HEMO (Hemoragik) atau kelas INF (Infark). Tabel data\_test ditunjukan pada Tabel 4.2.

Tabel 4.2. Data *Test*

|  |  |  |  |
| --- | --- | --- | --- |
| *Field* | *Type* | *Length*/*Values* | *Index* |
| kode\_pasien | Varchar | 10 | Primary |
| Jk | Varchar | 2 | - |
| Umur | Varchar | 5 | - |
| Alamat | Varchar | 50 | - |
| Kesimpulan | Varchar | 9 | - |
| Sistol | INT | 5 | - |
| Distol | INT | 5 | - |
| Kesadaran | Varchar | 11 | - |
| mual\_muntah | Varchar | 5 | - |
| nyeri\_kepala | Varchar | 5 | - |
| sulit\_bicara | Varchar | 5 | - |
| gerak\_terbatas | Varchar | 5 | - |
| Lemas | Varchar | 5 | - |
| Kejang | Varchar | 5 | - |

3. Tabel Login

Tabel login digunakan untuk menampung data *admin*  yang memiliki akses untuk mengoperasikan sistem. Tabel login terdiri dari 3 *field* yang dapat di lihat pada tabel 4.3.

|  |  |  |  |
| --- | --- | --- | --- |
| *Field* | *Type* | *Length/Values* | *Index* |
| ID | Varchar | 10 | Primary |
| Nama | Varchar | 10 | - |
| Sandi | Varchar | 10 | - |

Tabel 4.3. Login

#### Berdasarkan tabel 4.1. sampai dengan tabel 4.3 dapat digambarkan hubungan antar entitas yang dapat dilihat pada gambar 4.6.

Data Latih

Data Test

Lemas

Kejang

Sulit Bicara

Diastol

Sistol

Gerak Terbatas

Kesadaran Menurun

Mual Muntah

Nyeri Kepala

kode\_pasien

kode\_pasien

kode\_pasien

kode\_pasien

kode\_pasien

kode\_pasien

kode\_pasien

kode\_pasien

kode\_pasien

kode\_pasien

ket

ket

ket

ket

ket

ket

ket

ket

ket

Memiliki

Pengambilan Keputusan

kode\_pasien

jk

jk

umur

umur

alamat

alamat

kesimpulan

kesimpulan

Sistol

Distol

Kesadaran Menurun

Nyeri Kepala

Mual Muntah

Gerak Terbatas

Sulit Bucara

Kejang

Lemas

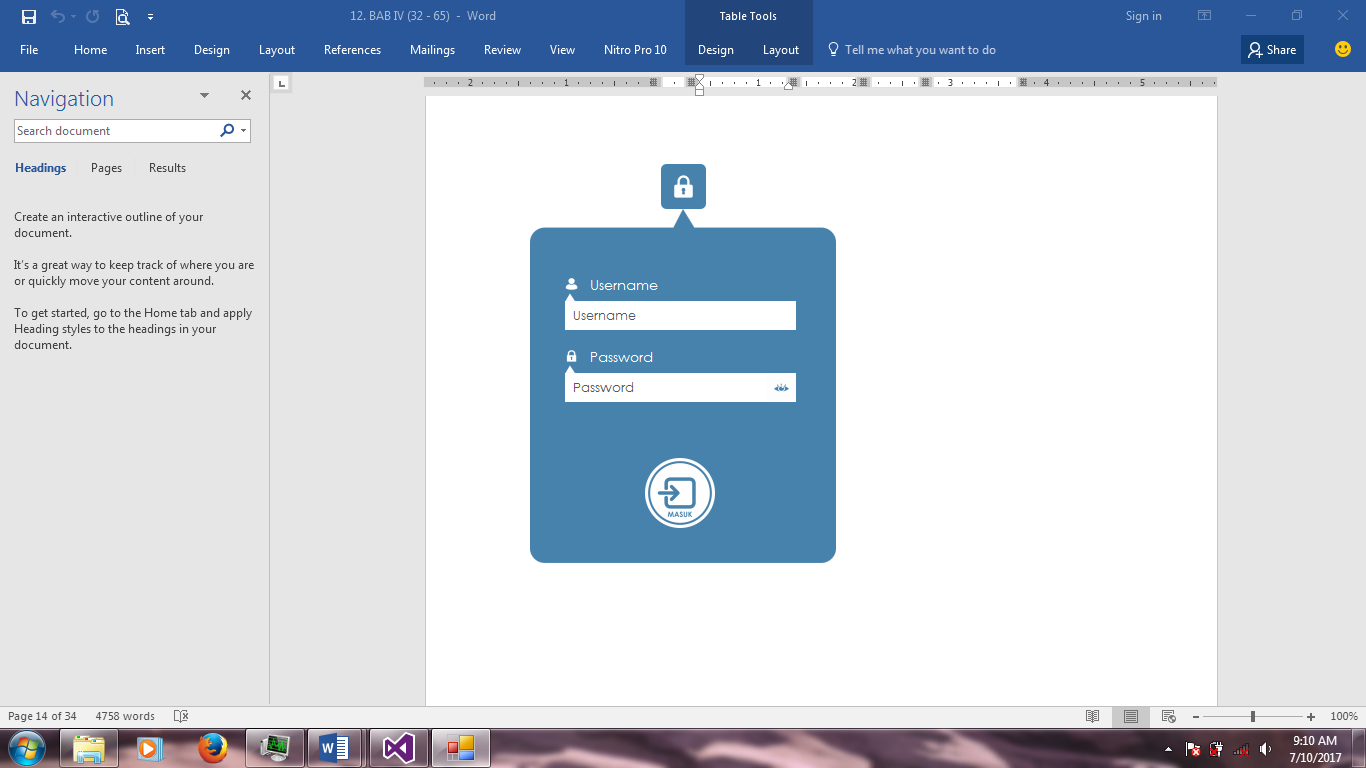
Gambar 4.6. *Entity Relationship Diagram* Sistem Klasifikasi Penyakit *Stroke*

Dari gambar 4.6. dapat diketahui hubungan antar tabel sebagai berikut: Data Latih memiliki fitur-fitur yang terdiri dari 9 *entinty* yaitu sistol, diastol, kesadaran menurun, mual muntah, nyeri kepala, gerak terbatas, sulit bicara, kejang, lemas. Masing-masing entity memiliki 2 atribut yaitu kode\_pasien dan ket (keterangan). Data latih juga memiliki relasi dengan Data Test, dimana Data Test mengambil keputusan berdasarkan Data Latih.

#### **4.1.2.3. Implementasi Input**

1. Form *Login*

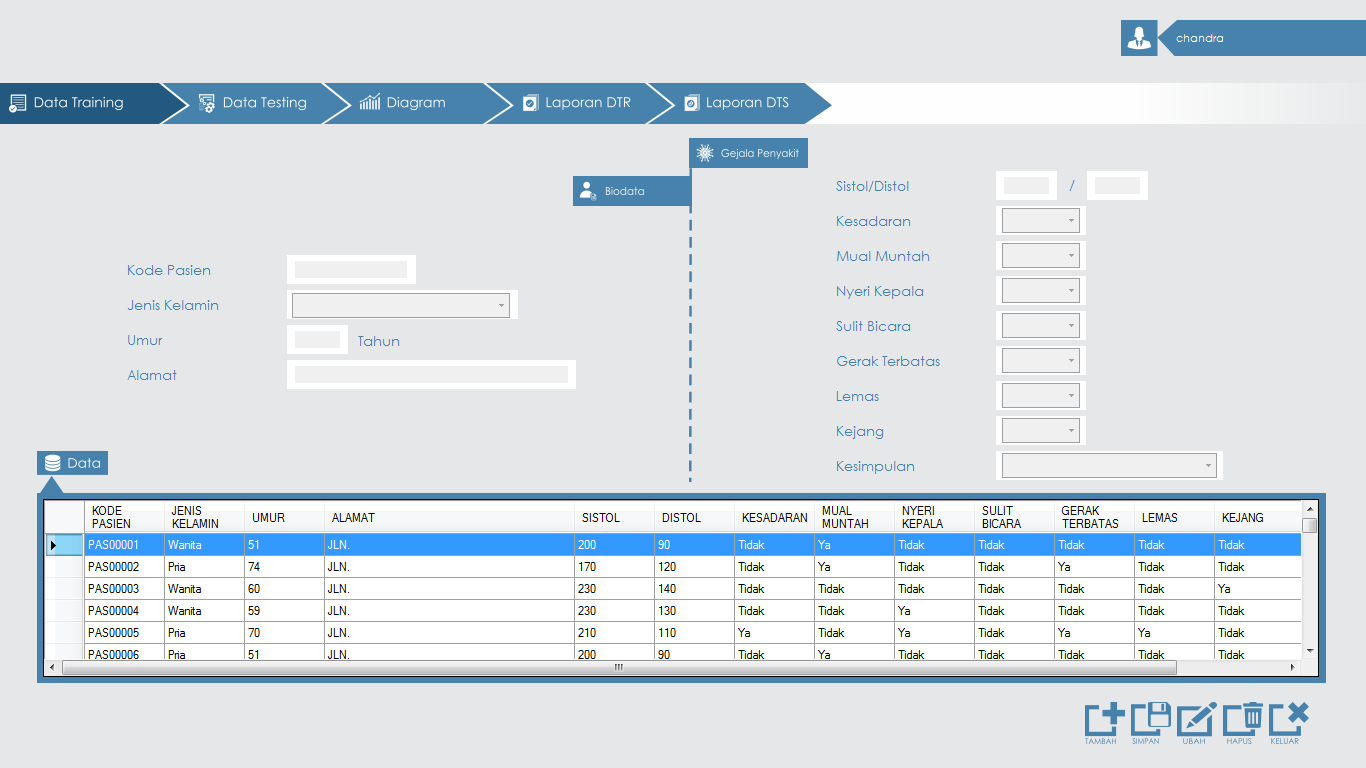
Agar dapat menggunakan Sistem Klasifikasi Penyakit *Stroke* ini, langkah awal yang dapat dilakukan adalah dengan memasukan *username* dan *password* yang terdapat pada form *login* yang akan muncul pertama kali saat membuka sistem. Tampilan form *login* dapat dilihat pada Gambar 4.7.



Gambar 4.7. Form *login* Sistem Klasifikasi Penyakit *Stroke*

1. Form *Training*

Form ini berisi data *training* atau data latih untuk menjadi parameter perhitungan kelas data baru atau data test. Form ini juga berfungsi menginputkan data *training* atau data latih baru. Tampilan form data *training*  dapat dilihat pada gambar 4.8.

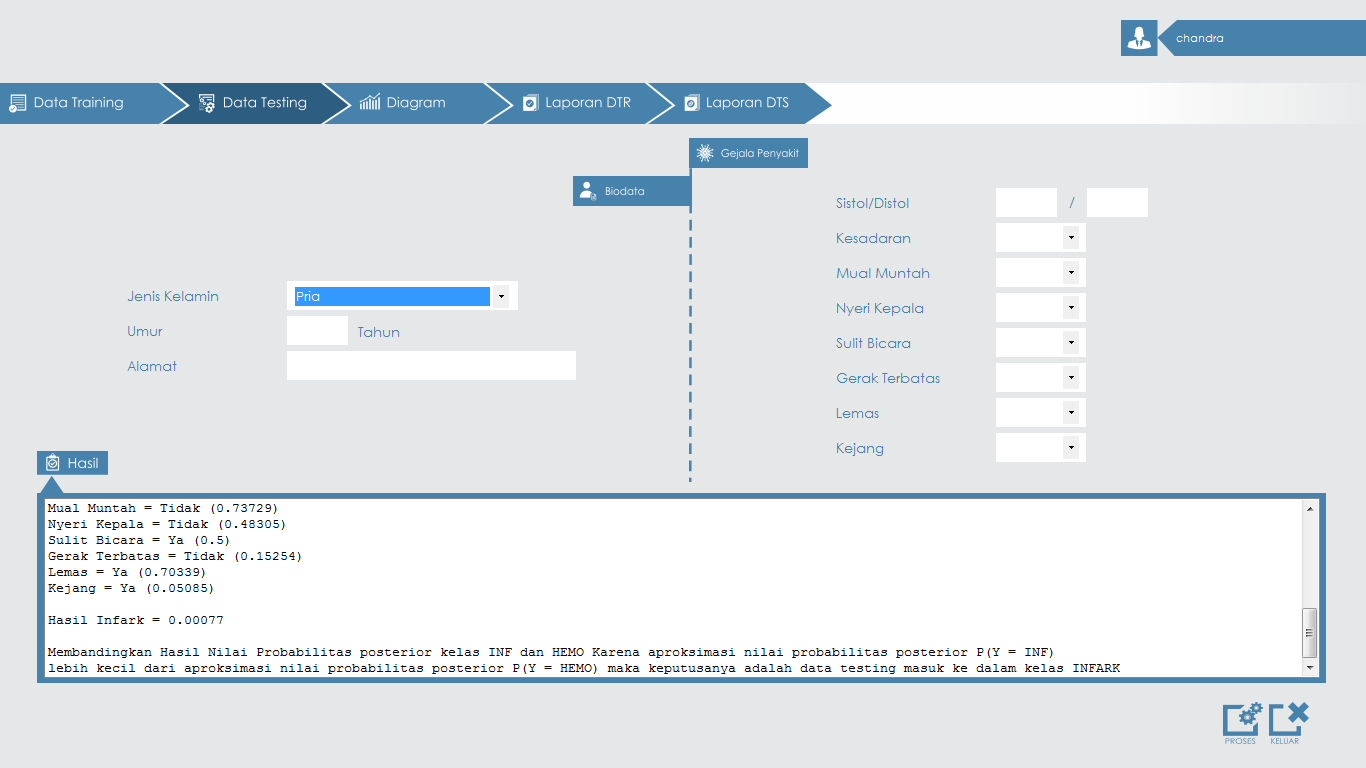


Gambar 4.8. Form data *training* Sistem Klasifikasi Penyakit *Stroke*

#### **4.1.2.4. Implementasi Output**

1. Form Data *Testing*

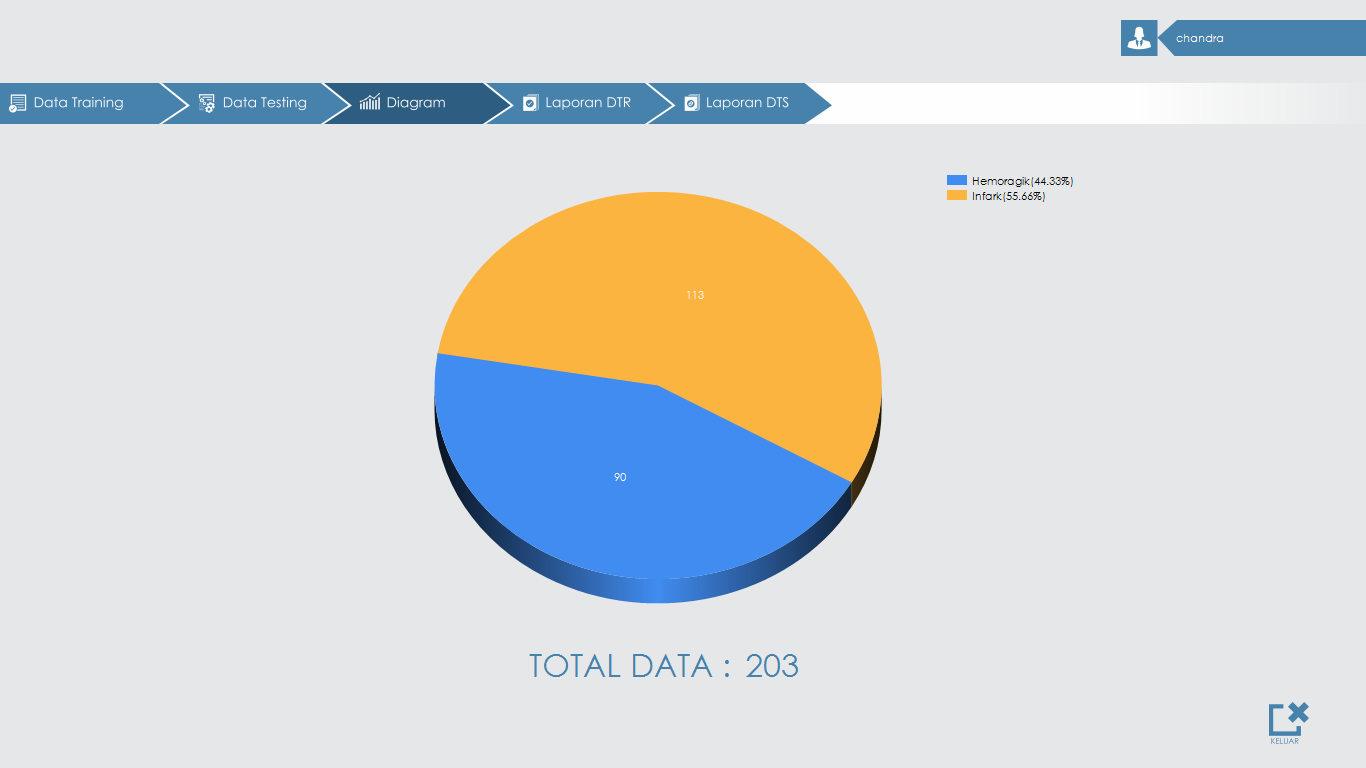
Form data testing digunakan untuk memasukan dan menguji data baru yang masuk juga menguji algoritma *Naive Bayes Classifier* apakah dapat mengklasfikasin dengan baik atau tidak. Tampilan form data *testing* dapat dilihat pada gambar 4.9.



Gambar 4.9. Form Data *Testing* Sistem Klasifikasi Penyakit *Stroke*

1. Form Diagram

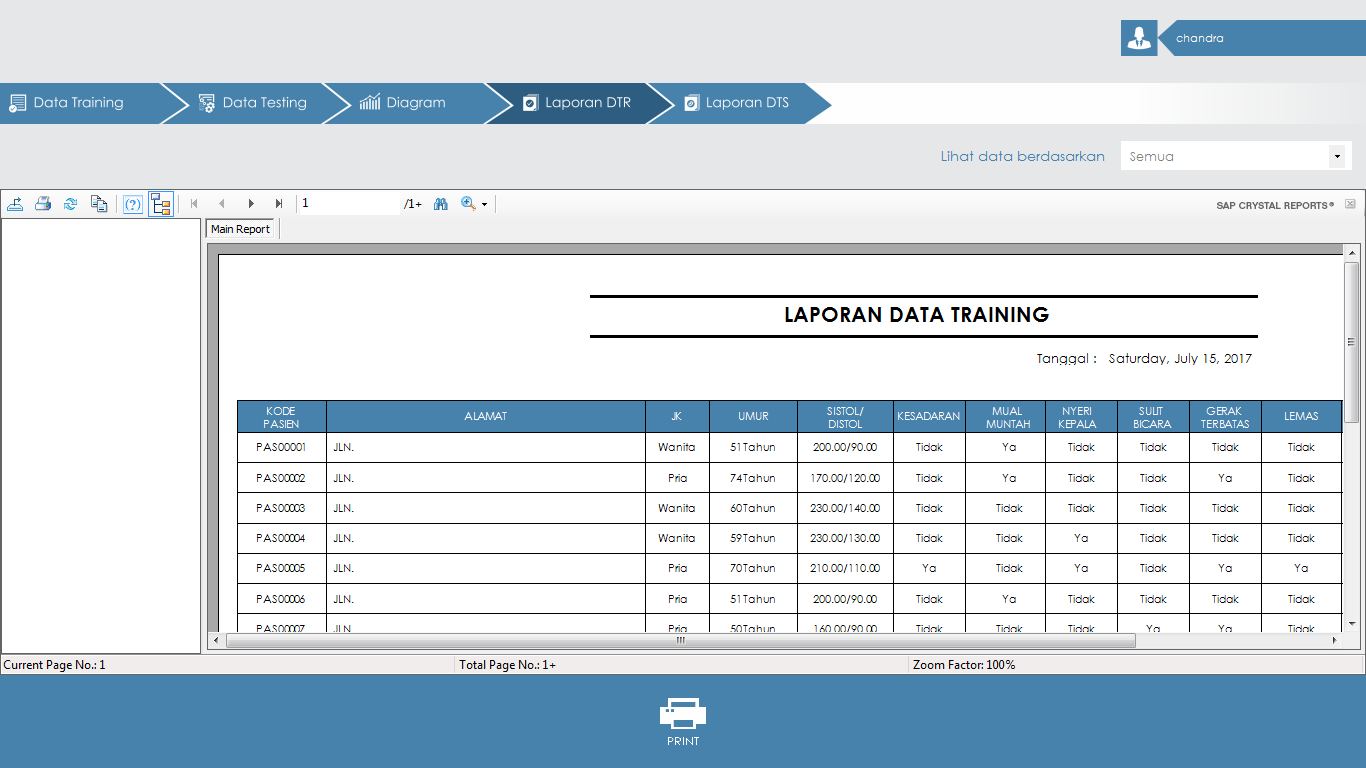
Form diagram dibuat untuk melihat persentasi hasil dari HEMO (*hemoragik*) dan INF (*infark*). Tampilan form diagram dapat dilihat pada gambar 4.10.



Gambar 4.10. Form Diagram Sistem Klasifikasi Penyakit *Stroke*

1. Form Laporan Data *Training*

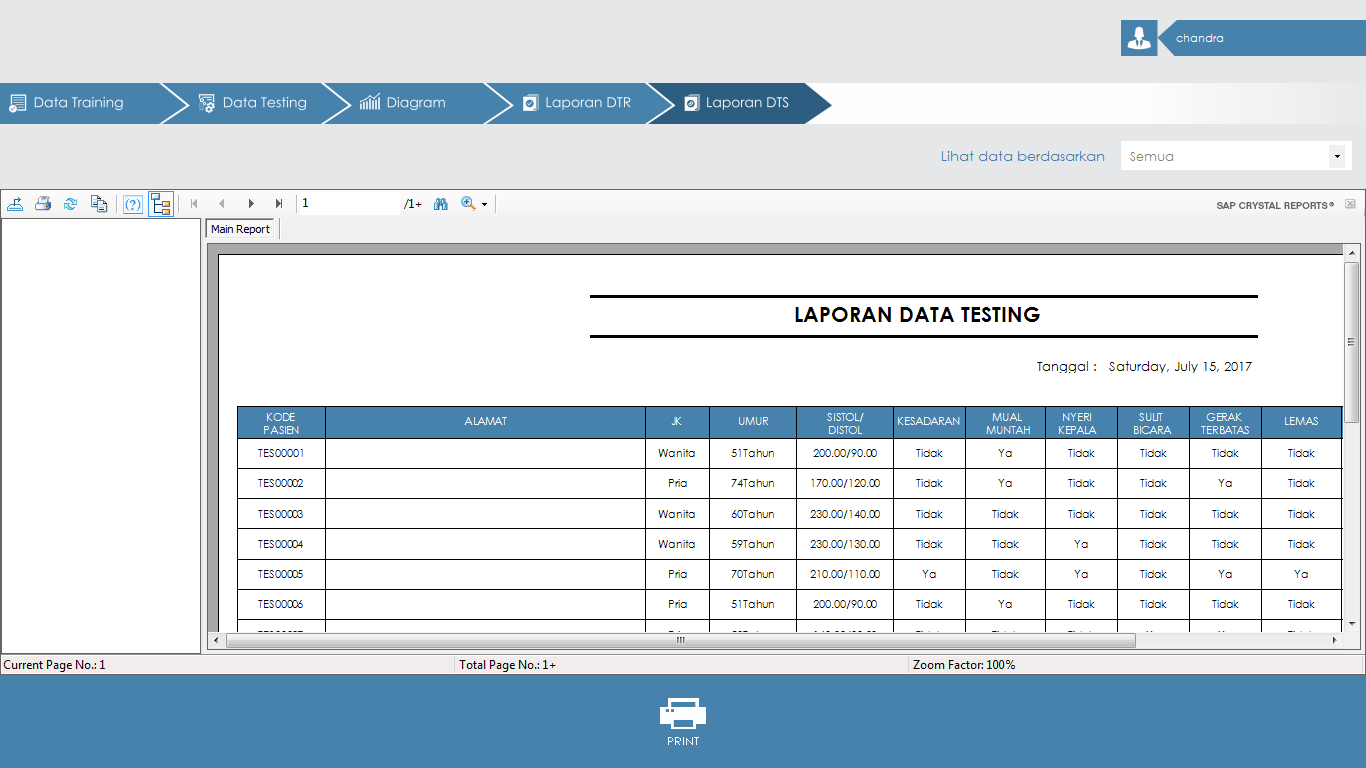
Form ini digunakan untuk menampilkan laporan data *training* yang ada dalam *database*. Tampilan form dapat dilihat pada gambar 4.11.



Gambar 4.11. Form Laporan Data *Training*

1. Form Laporan Data *Testing*

Form ini digunakan untuk menampilkan laporan data *testing* yang ada dalam *database*. Tampilan form dapat dilihat pada gambar 4.12.



Gambar 4.12. Form Laporan Data *Testing*

### **4.1.3. Pengujian Sistem**

Setelah proses pembuatan aplikasi telah selesai, tahapan selanjutnya yang akan dilakukan adalah pengujian sistem. Pengujian sistem melingkupi dua jenis pengujian yaitu pengujian fungsi sistem dan pengujian pada algoritma *Naive Bayes Classifier*. Pengujian pada sistem dilakukan dengan metode *blackbox* yang dapat menunjukan fungsi-fungsi dalam sistem telah berjalan sesuai dengan yang diharapkan atau tidak.

Tabel 4.4. Tabel Pengujian Fungsi Sistem

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Fungsi Yang Diuji | Ekspektasi Hasil | Hasil Pengujian |
| 1. | Verifikasi  Form *Login* | Jika admin melakukan *login* dan memasukan *username* dan *password*  dengan benar maka admin masuk kedalam sistem, jika salah satu tidak sesuai maka muncul pesan untuk mengganti *username* atau *password* yang salah. | Berhasil |
| 2. | Tambah Data  (Form *Data Training*) | Dapat menambahkan data pada *data Training*. | Berhasil |
| 3. | Simpan Data  (Form *Data Training*) | Dapat menyimpan data pada *data Training*. | Berhasil |
| 4. | Ubah Data  (Form *Data Training*) | Dapat mengubah data pada *data Training*. | Berhasil |
| 5. | Hapus Data  (Form *Data Training*) | Dapat menghapus data pada *data Training*. | Berhasil |
| 6. | Tambah Data  (Form *Data Testing*) | Dapat menambahkan data pada *data Testing*. | Berhasil |
| 7. | Proses Bayes  (Form *Data Testing*) | Dapat menyimpan dan melakukan proses *Bayes* pada *data* *Testing*. | Berhasil |
| 8. | Laporan  (Form *Data Training*) | Dapat melihat dan mencetak laporan *data Training*. | Berhasil |
| 9. | Laporan  (Form *Data Testing*) | Dapat melihat dan mencetak laporan *data Testing*. | Berhasil |

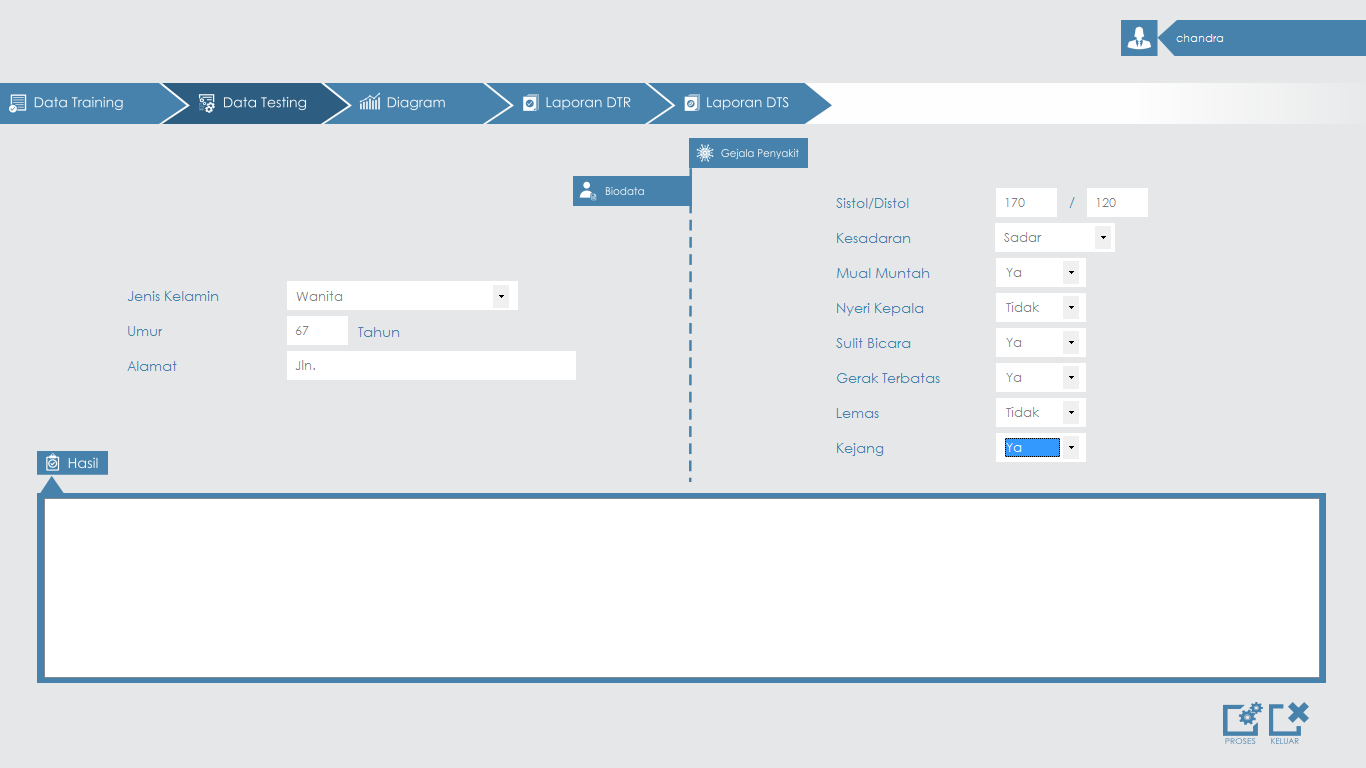
Selanjutnya dilakukan proses pengujian algoritma *Naive Bayes Classifier* antara data Latih atau data *training* dengan data uji atau data *testing* yang dapat dilihat pada tabel 4.5. sampai dengan tabel 4.12 di bawah ini:

Dari tabel pengujian fungsi sistem pada tabel 4.4. di atas dapat diketahui jika seluruh fungsi-fungsi sistem yang terdapat dalam sistem Klasifikasi Penyakit *Stroke* dapat berjalan sesuai dengan fungsi-fungsinya dan dapat digunakan dengan baik.

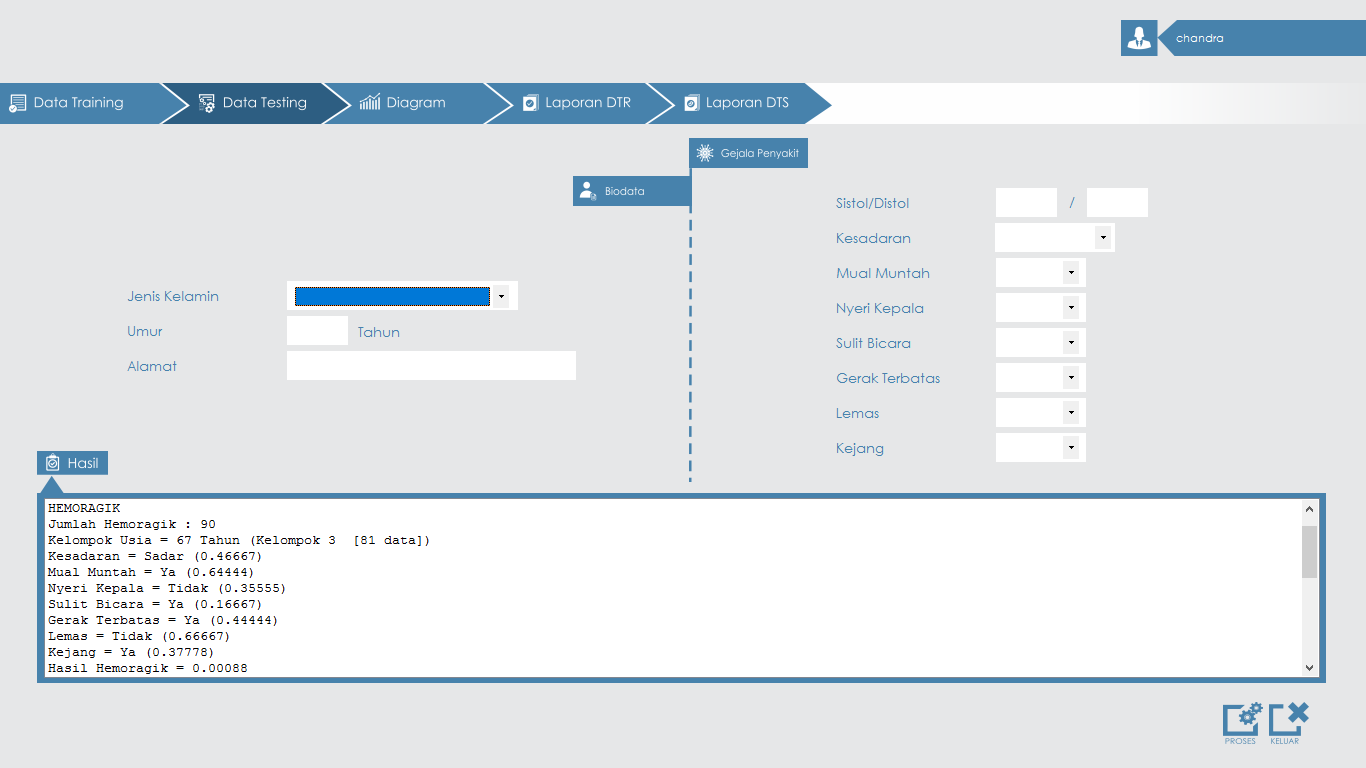
Tabel pengujian *Naive Bayes*  pada tabel 4.5. sampai tabel 4.12. di atas menunjukan pengujian sistem dengan 203 data yang akan mengukur tingkat keberhasilan dari sistem klasifikasi penyakit *stroke*. Dalam pengujian ini didapatkan hasil dari tingkat keberhasilan algoritma *Naive Bayes Classifier* adalah sebesar 89.65%, nilai persentase ini didaptkan berdasarkan persamaan (3) . Dalam pengujian *Naive Bayes* terdapat 21 data yang tidak sesuai dalam pengujian sehingga menjadi .

Tabel 4.13. Contoh Kasus Data *Testing* Pada Sistem

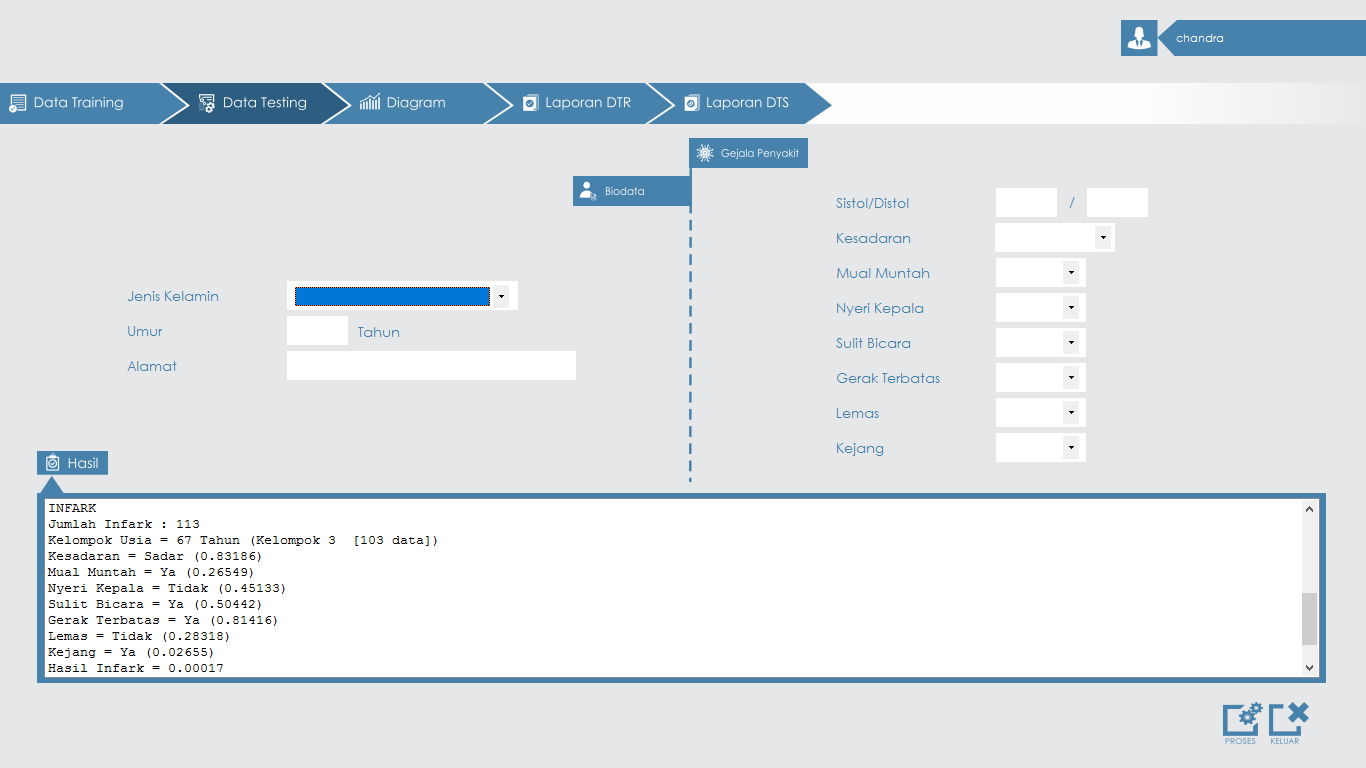
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Jenis Kelamin | Umur | Sistol/ Diastol | Kesadaran | Mual Muntah | Nyeri Kepala | Sulit Bicara | Gerak Terbatas | Lemas | Kejang |
| Wanita | 67 | 170/ 120 | Sadar | Ya | Tidak | Ya | Ya | Tidak | Ya |



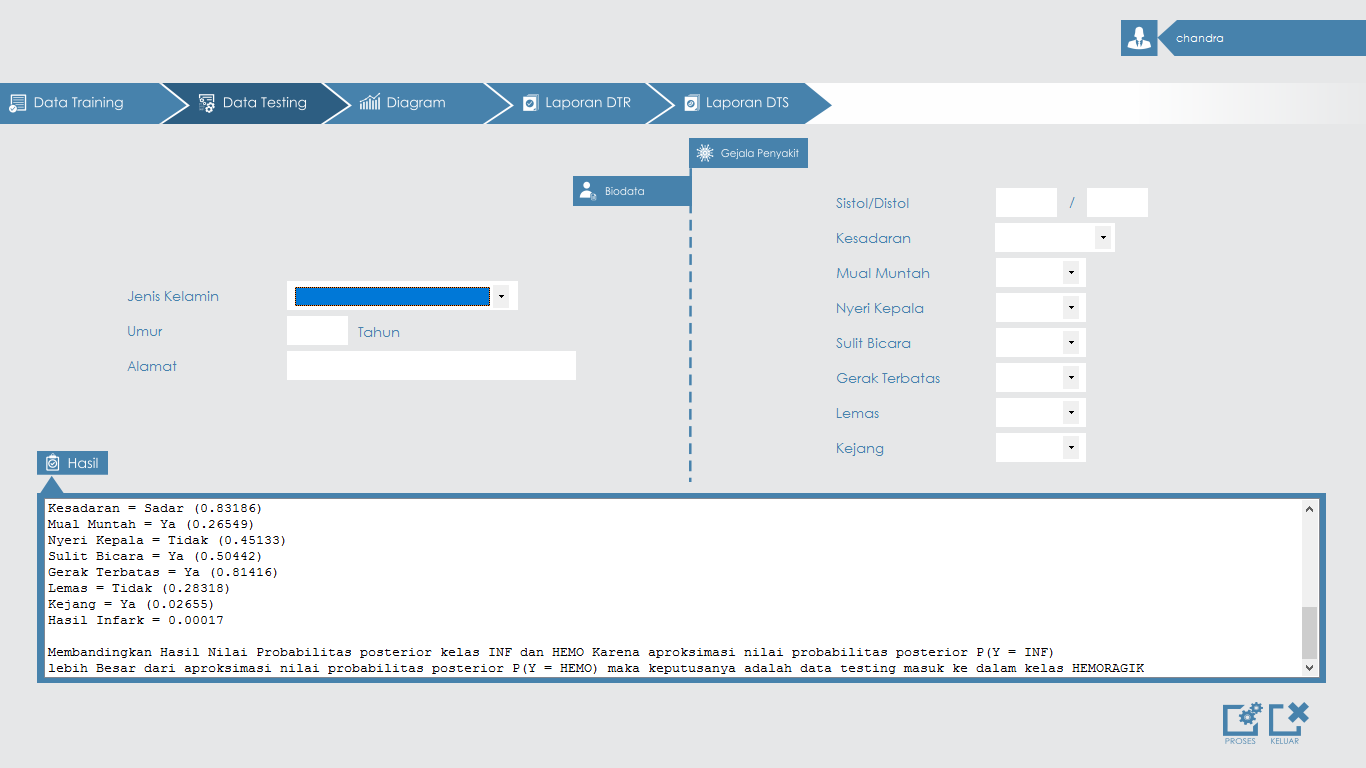
Gambar 4.13. Contoh Kasus Data *Testing* Pada Sistem



Gambar 4.14. Hasil Hemoragik untuk Data *Testing* Pada Sistem



Gambar 4.15. Hasil Infark untuk Data *Testing* Pada Sistem



Gambar 4.16. Kesimpulan untuk Data *Testing* Pada Sistem

Tahap 1: Menghitung probabilitas awal masing-masing kelas/label

P(Y= HEMO) = 90/203

P(Y= INF) = 113/203

Tahap 2: Menghitung probabilitas masing-masing fitur (sesuai dengan data yang terdapat pada data *testing*).

* P(Kesadaran = Sadar | Y=HEMO) = 42/90
* P(Mual Muntah = Ya | Y=HEMO) = 58/90
* P(Nyeri Kepala = Tidak | Y=HEMO) = 32/90
* P(Sulit Bicara = Ya | Y=HEMO) = 15/90
* P(Gerak Terbatas = Ya | Y=HEMO) = 40/90
* P(Lemas = Tidak | Y=HEMO) = 60/90
* P(Kejang = Ya | Y=HEMO) = 34/90
* P(Kesadaran = Sadar | Y=INF) = 94/113
* P(Mual Muntah = Ya | Y= INF) = 30/113
* P(Nyeri Kepala = Tidak | Y= INF) = 51/113
* P(Sulit Bicara = Ya | Y= INF) = 57/113
* P(Gerak Terbatas = Ya | Y= INF) = 92/113
* P(Lemas = Tidak | Y= INF) = 32/113
* P(Kejang = Ya | Y= INF) = 3/113

Tahap 3: Mengalikan semua hasil perhitungan pada tahap 1 dan 2 untuk masing-masing kelas (HEMO dan INF)

P(Y= HEMO) × P(Kesadaran = Sadar | Y=HEMO) × P(Mual Muntah = Ya | Y=HEMO) × P(Nyeri Kepala = Tidak | Y=HEMO) × P(Sulit Bicara = Ya | Y=HEMO) × P(Gerak Terbatas = Ya | Y=HEMO) × P(Lemas = Tidak | Y=HEMO) × P(Kejang = Ya | Y=HEMO)

= 90/203 × 42/90 × 58/90 × 32/90 × 15/90 × 40/90 × 60/90 ×34/90

= 0.00088

P(Y= INF) × P(Kesadaran = Sadar | Y= INF) × P(Mual Muntah = Ya | Y= INF) × P(Nyeri Kepala = Tidak | Y= INF) × P(Sulit Bicara = Ya | Y= INF) × P(Gerak Terbatas = Ya | Y= INF) × P(Lemas = Tidak | Y= INF) × P(Kejang = Ya | Y= INF)

= 113/203 × 94/113 × 30/113 × 51/113 × 57/113 × 92/113 × 32/113 × 3/113

= 0.00017

Tahap 4: Membandingkan hasil probabilitas antara nilai posterior kelas HEMO dan INF. Karena nilai posterior P(Y=HEMO) lebih besar dari nilai posterior P(Y=INF) maka hasil data *testing* diklasifikasin ke dalam kelas “HEMO

## **4.2. Pembahasan**

### **4.2.1. Fungsi Sistem**

1. Form *Login*

Form *login* dibuat untuk melakukan proses *login* dimana proses ini dimaksudkan untuk membatasi pengguna sistem yang memiliki akses ke dalam sistem ini. *Form* ini menggunakan *username* dan *password* untuk dapat melakukan verifikasi, jika salah satu kombinasi yaitu *username* atau *password* tidak sesuai maka pengguna tidak dapat masuk kedalam sistem. Sebaliknya jika *username* dan *password* sesuai atau cocok maka pengguna dapat masuk ke dalam sistem.

1. Form *Data Training*

Form *data training* digunakan untuk menyimpan data *training* atau data latihyang menjadi parameter untuk menghitung probabilitas data baru yang akan masuk ke kelas INF atau HEMO pada sistem Klasifikasi Penyakit *Stroke*. Form ini berisi data rekam medik pasien yang telah terkena penyakit *stroke* dengan hasil diagnosa dokter yang mengarah pada kesimpulan kelas dari penyakit yang diderita pasien, sehingga datanya dapat disebut *valid* untuk menjadi parameter dalam menghitung data baru. Dalam form ini pengguna dapat memasukan data baru yang telah *valid* sebelumnya yaitu yang telah diketahui hasilnya berdasarkan kesimpulan dokter dengan dilengkapi *button* simpan data, hapus data dan ubah data paisen.

1. Form *Data Testing*

Form *data testing* digunakan untuk menyimpan data pasien yang belum *valid*  atau memiliki kesimpulan kelas yang tepat. Data *test* atau data baru ini dapat menguji algoritma *Naive Bayes Classifier* untuk dapat mengklasifikasikan data baru dengan tepat ke kelas yang tepat. Data *test* yang dimasukan adalah data yang belum diketahui kesimpulan kelasnya, jadi di dalam form ini terdapat fungsi untuk menambahkan data dan melakukan proses *Bayes*  yang akan langsung tersimpan pada tabel data\_test. Selanjutnya pada proses perhitungan *Naive Bayes* dapat dilihat hasil kelas yang probabilitasnya paling besar.

1. Form Diagram

Form ini menunjukan persentasi dari jumlah data pasien yang terkena penyakit *stroke* dalam bentuk diagram *pie*. Warna kuning digunakan untuk menunjukan persentasi nilai HEMO sedangkan warna biru digunakan untuk menunjukan persentasi nilai INF.

1. Form Laporan

Form laporan dibuat untuk melihat laporan pasien yang diambil dari tabel data\_latih dan tabel data\_test. Form laporan juga meliki fungsi untuk mencetak laporan pasien. Jika data pasien dibutuhkan dalam bentuk cetak, maka dibuat form untuk itu.

### **4.2.2. Proses Algoritma Naive Bayes Classifier**

Dalam proses perhitungan sebuah kasus menggunakan algoritma *Naive Bayes Classifier* terdapat berbagai macam tingkat keberhasilan yang dapat diperoleh pada hasil akhir pengujian. Dalam kasus dan algoritma yang penulis gunakan hasil pengujiannya mendapatkan persentase 89.65%. Hasil persentasi ini didapat dari pengujian data latih atau data *training* menjadi data uji atau data *testing*, yaitu sebanyak 203 data. Hasil pengujian ini tidak dapat mencapai 100% karena dalam perhitungan probabilitas yang dilakukan terdapat kemungkinan data *error* yaitu klasifikasi data yang tidak tepat (*misclassification*) karena pengaruh tingkat ketergantungan (*dependency*) antar fitur yang cukup tinggi . Jika ada kemungkinan hasil probabilitas sama antara kelas HEMO dan INF maka langkah selanjutnya adalah melihat sebaran data terbanyak.

Fitur-fitur yang digunakan dalam pengujian antara lain kesadaran menurun, mual muntah, nyeri kepala, sulit bicara, gerak terbatas, lemas, dan kejang. Usia tidak dimasukkan ke dalam fitur karena dari hasil penelitian, seseorang dapat terkena penyakit *stroke* di berbagai lapisan usia. Dengan kata lain, usia bukan merupakan fitur yang dominan pada penelitian ini.

Sistol dan Diastol juga tidak dimasukkan ke dalam fitur karena tekanan darah seseorang bukan merupakan indikator yang kuat untuk menentukan apakah orang tersebut sudah terindikasi menderita penyakit *stroke* atau tidak. Jadi hanya 7 fitur saja yang dijadikan parameter perhitungan probabilitas penyakit *stroke* dengan algoritma *Naive Bayes Classifier*.

### **4.2.3. Tahapan Alir Sistem**

Dalam pembuatan sistem klasifikasi penyakit *Stroke* terdapat beberapa tahapan yang dilakukan yaitu sebagai berikut:

1. Melakukan Proses Simpan, Ubah, dan Hapus

Meng-*input­-kan* data *training* atau data latih yang didalamnya dapat melakukan penyimpanan data, ubah data, dan hapus data yang merupakan bagian dari sebuah sistem dengan memiliki proses simpan, ubah, dan hapus data.

2. Melakukan Input Data Uji

Melakukan input data uji atau data *Testing* untuk menguji data yang akan digunakan dalam melakukan pengujian proses *Bayes* yaitu dengan mengisi fitur-fitur yang terdapat pada bagian *input*  pada form data *Testing*.

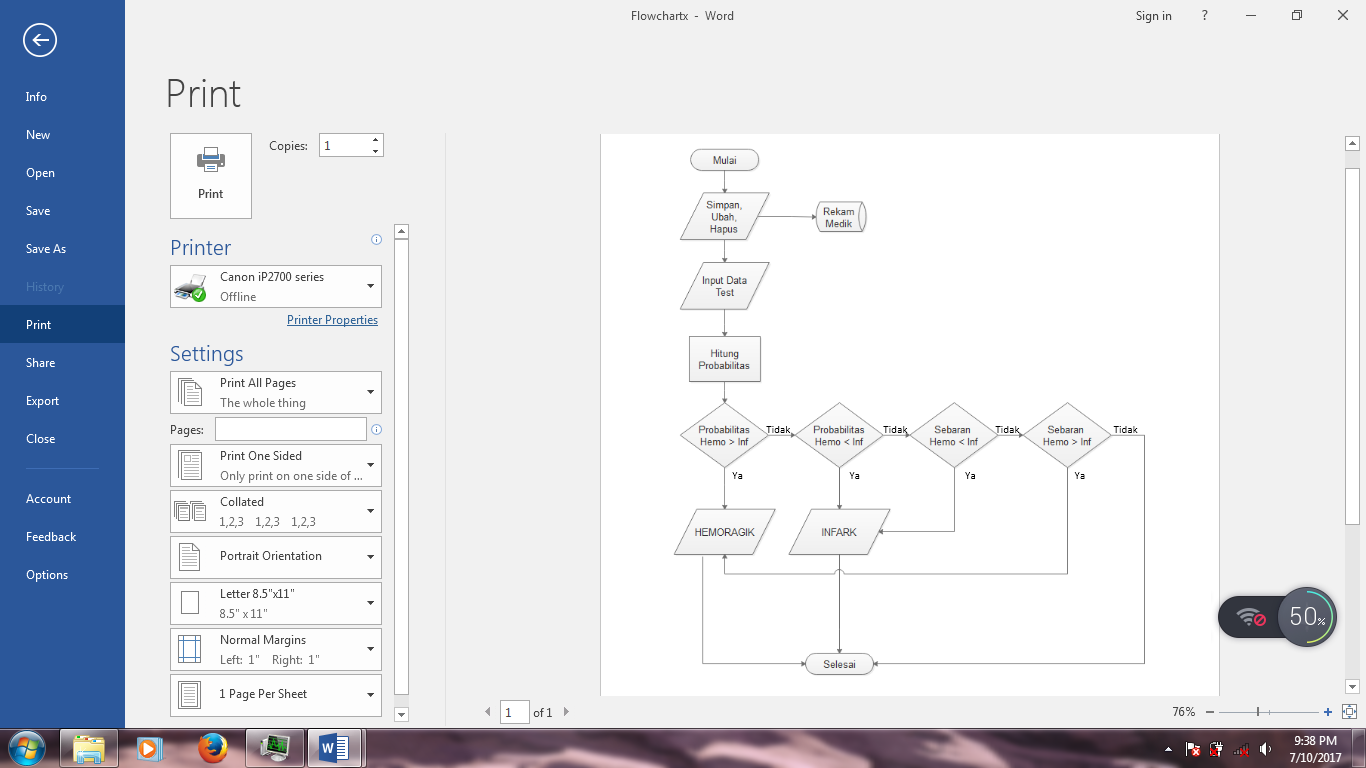
1. Menghitung Probabilitas

Menghitung probabilitas dilakukan setelah melakukan input data uji atau data *Testing* lalu melakukan proses *Bayes* dan melihat hasil probabilitas data uji yang akan masuk pad akelas HEMO atau INF.

1. Membandingkan Hasil Probabilitas

Dalam membandingkan hasil probabilitas terdapat empat kondisi untuk melihat kelas yang sesuai dengan data uji atau data *Testing*. Hasil dari kondisi yang tepat akan mengarah pada kesimpulan data uji akan masuk ke kelas HEMO atau INF.

Tahapan-tahapan alir sistem ini dapat dilihat pada gambar 4.17.



Gambar 4.17. *Flowchart* Tahapan Alir Sistem

# **BAB V**

**PENUTUP**

## **5.1. Kesimpulan**

Dari hasil penelitian yang telah dilakukan dalam pembuatan Sistem Klasifikasi penyakit *Stroke* menggunakan metode *Naive Bayes Classifier* yang dilakukan pada Rumah Sakit Umum Daerah Undata Palu maka dapat dibuat kesimpulan sebagai berikut:

* 1. Dalam pembuatan sistem klasifikasi penyakit *stroke* menggunakan metode *Naive Bayes Classifier* dilakukan 4 tahapan, yaitu: melakukan proses simpan, ubah dan hapus, melakukan input data uji, menghitung probabilitas, dan yang terakhir membandingkan hasil dari probabilitas.
  2. Algoritma *Naive Bayes Classifier* dapat memberikan hasil 89.65% untuk tingkat keberhasilan dalam melakukan klasifikasi. Hasil ini menunjukkan bahwa Algoritma *Naive Bayes Classifier* dapat digunakkan (valid) untuk melakukan proses klasifikasi penyakit *stroke*.
  3. Dengan 203 data latih yang digunakan sebagai parameter, terdapat 7 fitur-fitur yang digunakan anatar lain: Kesadaran Menurun, Mual Muntah, Nyeri Kepala, Sulit Bicara, Gerak Terbatas, Lemas, Kejang. Dengan nilai satuan Ya atau Tidak.
  4. Dari ketujuh fitur tersebut dapat diklasifikasikan dalam dua kelas yaitu Hemo (Hemoragik) dan Inf (Infark) berdasarkan kelainan patologisnya.

## **5.2. Saran**

Dari hasil penelitian yang telah dilakukan sebelumnya maka diharapkan pada peneliti yang akan melakukan penelitian selanjutnya untuk mengembangkan klasifikasi penyakit stroke selanjutnya, antaralain:

1. Dapat menggunakan algoritma pembanding selain *Naive Bayes Classifier* untuk dapat melihat hasil perbandingan dan keakuratan antara algoritma *Naive Bayes Classifier* dan algoritma lainnya. Selain itu dengan menghadirkan algoritma pembanding, dapat ditentukan algoritma mana yang memiliki tingkat keberhasilan tertinggi.
2. Sumber data yang akan digunakan sebagai data latih sebaiknya tidak hanya berasal dari satu sumber (dalam kasus penelitian ini hanya berasal dari satu rumah sakit) tetapi juga berasal dari sumber lain sehingga pada akhirnya akan didapatkan data latih dalam jumlah yang besar.

# **DAFTAR PUSTAKA**

Abdilah, S. 2015. *Penerapan Algoritma Decision Tree C4.5 Untuk Diagnosa Penyakit Stroke Dengan Klasifikasi Data Mining Pada Rumah Sakit Santa Maria Pemalang*. Universitas Dian Nuswantoro, Semarang.

Arifianto, A.S. Sarosa,M. Setyawati, O. 2014. *Klasifikasi Stroke Berdasarkan Kelainan Patologis dengan Learning Vector Quantization*.Universitas Brawijaya, Malang.

Budiharjo, S. P. 2014. *Gambaran Kadar Gula Darah Dan Derajat Keparahan Stroke Pada Penderita Stroke Iskemik Trombotik*. Universitas Katolik Widya Mandala, Surabaya.

Connolly, Thomas, C. B. 2010. *DataBase System : A Practical approach to Design, Implemantation, and Management Fifth Edition* : Pearson Education Inc.

Han, J., dan Kamber, M. 1999. *Introduction to Data Mining and Knowledge Discovery*. Two Crows Corporation. Third Edition, USA.

Huda, N. M. 2010. *Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa (Studi Kasus Di Fakultas Mipa Universitas Diponegoro)*. Universitas Diponegoro, Semarang.

Inmon, William H. 2005. “*Building The Data Warehouse (4th ed.)*”. Indianapolis: Wiley Publishing, Inc.

Jogiyanto, 2010. *Analisis dan Desain Sistem Informasi,* Edisi IV, Andi Offset, Yogyakarta.

Kusrini. Luthfi, E. T., 2009*. Algoritma Data Mining*. Yogyakarta : Penerbit Andi.

Larose, Daneil T. 2005. *Discovering Knowledge in Data : An Introduction to Data Mining.* Canada : John Wiley & Sons, Inc.

McLeod, R. Schell, G. 2004. *Sistem Informasi Manajemen* Edisi ke-8. Jakarta: PT. Indeks.

Misbach J.1999. Stroke, Aspek Diagnostik,Patofisiologi,Manajemen,edisi pertama, BPFK Universitas Indonesia, Jakarta.

Mukhlis, M. K, Martiana, E. 2011. *Diagnosa Kemungkinan Pasien Terkena Stroke Dengan Menggunakan Metode Naive Bayes Dan Metode Jaringan Syaraf Tiruan Berbasis Web*. Politeknik Institut Teknologi Sepuluh November, Surabaya.

Ponniah, P. 2001. *Datawarehouse Fundamentals: A Comprehensive Guide for IT Professional*. John Willey & Sons, Inc.

Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi menggunakana MATLAB.* ANDI, Yogyakarta.

Riskesdas. 2013. *Badan Penelitian dan Pengembangan Kesehatan*. Kementerian Kesehatan RI.

Sommerville, I. 2011. *Software Engineering* (Rekayasa Perangkat Lunak). Jakarta: Erlangga.

Sugiyono. 2002. Metode Penelitian Bisnis. PT. Gramedia Pustaka Utama, Jakarta.

Sukamto, R.A., dan Shalahuddin, M. 2011, *Modul Pembelajaran Rekayasa Perangkat Lunak*, Bandung.

Turban, E. 2005. *Decision Support Systems and Intelligent Systems*. Dwi Prabantini (Penterjemah). Yogyakarta: Andi.