**BAB II**

**LANDASAN TEORI**

**2.1. Tinjauan Studi**

Berikut ini adalah beberapa penelitian yang menjadi referensi penulis dalam menyusun Usulan Penelitian ini :

1. Penelitian dilakukan oleh Arief Jananto. Penelitian yang berjudul “*Algoritma Naive Bayes untuk mencari perkiraan Waktu Studi Mahasiswa*”. Metode yang digunakan pada penelitian ini adalah Algoritma Klasifikasi *Naive Bayes*. Data yang diperoleh adalah data yang berkaitan dengan keterangan diri dari mahasiswa UNISBANK lulusan tahun 2004-2007. Kesimpulan dari penelitian adalah sebagai berikut :

1. Ketepatan masa studi mahasiswa diprediksi berdasarkan latar belakang sekolah sebelumnya dan data akademik, serta pribadi saat berada diperguruan tinggi.
2. Dalam memprediksi ketepatan studi mahasiswa menggunakan data training dan testing dengan memanfaatkan fungsi prediksi dari teknik data mining yaitu menggunakan algoritma klasifikasi *Naive Bayes*.
3. Fungsi klasifikasi yang digunakan dalam prediksi ketepatan masa studi mahasiswa ini mempunyai tingkat kesalahan sebesar 20% sampai 34%, kesalahan tersebut dapat dipengaruhi oleh jumlah data *training* dan *testing* serta tingkat konsistensi data yang digunakan.

2. Penelitian yang dilakukan oleh Yuda Septian Nugroho. Penelitian yang berjudul “*Algoritma Naive Bayes untuk Prediksi Kelulusan Mahasiswa*”. Tujuan dari penelitian sebagai berikut :

1. Mengklasifikasi kelulusan mahasiswa Udinus Fakultas Ilmu Komputer angkatan tahun 2009.
2. Mendapatkan hasil akurasi kelulusan sebesar 82.08%. menggunakan metode klasifikasi *Naive Bayes* dan dataset berupa data mahasiswa Universitas Dian Nuswantoro Fakultas Ilmu Komputer angkatan tahun 2009,
3. Data yang kurang kompleksitas menyebabkan model dapat memprediksi cukup akurat.

3. Peneliti Mujid Ridwan, Tahun 2013. Judul yang diangkat “*Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma* *Naïve Bayes Classifier*”**.** Di latar belakang di jelaskan bahwa Masiswa merupakan salah satu aspek penting dalam evaluasi keberhasilan dalam penyelenggaraan program studi pada suatu perguruan tinggi. Pememtauan mahassiswa yang masuk, peningkatan kemampuan mahasiswa, prestasi yang di capai mahasiswa, dan rasio kelulusan terhadap jumlah total mahasiswa. Kompetensi lulusan mendapatkan perhatian yang serius untuk memperoleh kepercayaan stakeholder dalam menilai dan menetapkan lulusannya. Oleh karena itu akan di buat sebuah sistem untuk mengklasifikasikan kelulusan mahasiswa dengan cara mengevaluasi kinerja pada tahun pertama dan atau pada tahun kedua.

Pada penelitian ini digunakan teknik data mining untuk menemukan pola kelulusan mahasiswa yang sudah lulus, kemudian di jadikan dasar untuk memprediksi kelulusan mahasiswa pada tahun ke-2.

Data Mining adalah proses menemukan hubungan dalam data yang tidak di ketahui oleh pengguna dan menyajikannya dengan cara dapat di pahami sehingga hubungan tersebut dapat di jadikan dasar untuk pengambilan keputusan. Teknik Data Mining yang di gunakan adalah algoritma *Naïve Bayes Classifier*.

*Naïve Bayes Classifier* merupakan sebuah pengklasifikasi probabilitas sederhana yang mengaplikasikan *Teorema Bayes.* Ide dasar dari *Teorema Bayes* adalah menangani masalah yang bersifat hipotesis yakni mendesain suatu klasifikasi untuk memisahkan objek.

**2.2. Tinjauan Pustaka**

**2.2.1 Data Mining**

*Data mining* adalah sebuah istilah yang digunakan untuk menggambarkan penemuan ilmu pengetahuan dalam bidang database, sebuah bidang analisis informasi yang mencari pola tersembunyi dalam sekelompok data yang dapat digunakan untuk memprediksi perilaku masa depan (turbanet al,2007).

Menurut Gartner Group *data mining* adalah suatu proses menemukanhubungan yang berarti, pola, dan kecendurungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika.(Larose,2005).

*Data Mining* adalah proses menemukan pengetahuan yang menarik dari sejumlah besar data yang tersimpan dalam database, data warehouse, atau repositori penyimpanan data dan informasi lainnya (Han dan Kamber, 2006).

*Data Mining* adalah proses menemukan hubung dalam data yang tidak diketahui oleh pengguna dan menyajikannya dengan cara yang dapat dipaham hingga hubungan tersebut dapat menjadi dasar pengambilan keputusan (McLeod, 2007).

*Data Mining* memegang peranan penting dalam berbagai bidang seperti industri, keuangan, cuaca, medical, teknologi dan lain sebagainya. *Data Mining*  berkenaan dengan jumlah pengolahan data dalam skala besar. Data-data yang ada, tidak dapat langsung diolah dengan menggunakan sistem data mining. Data-data tersebut harus dipersiapkan terlebih dahulu agar hasil yang diperoleh lebih maksimal, dan waktu komputasinya lebih minimal.

Sebagai suatu rangkaian proses, *Data Mining* dapat dibagi menjadi beberapa tahap proses. Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantaraan *knowledge base*.

1. Pembersihan data *(Data Cleaning)*

Pembersihan data merupakan proses menghilang-kan *noise* dan data yang tidak konsisten atau data tidak relevan.

1. Integrasi data *(Data Integration)*

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru.

1. Seleksi data *(Data Selection)*

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

1. Transformasi data (Data Transformation)

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

1. Proses *Data Mining*

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Beberapa metode yang dapat digunakan berdasarkan pengelompokan *Data Mining.*

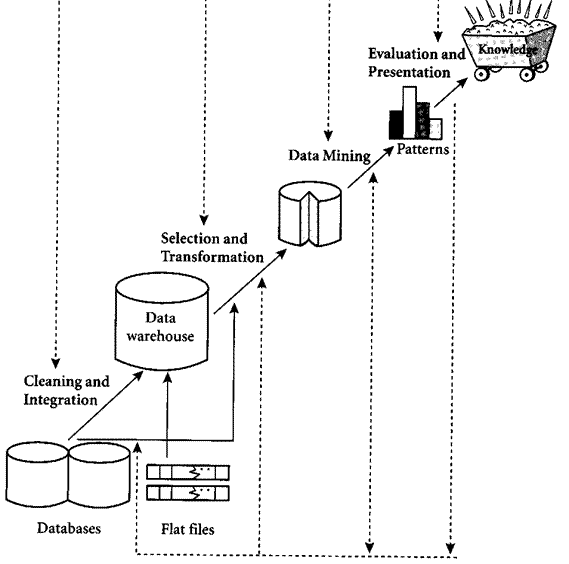
1. Evaluasi pola *(Pattern Evaluation)*

Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan.

1. Presentasi pengetahuan *(Knowledge Presentation)*

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

**2.2.2. Tahapan -Tahapan Data Mining**



**Gambar 2.2.** Tahapan Data Mining

Keterangan:

1. Pembersihan Data

Pembersihan data dilakukan untuk menghilangkan noise dan data yang tidak konsisten atau tidak relevan. Sering kali data yang diperoleh dari *database* suatu perusahaan maupun diperoleh melalui hasil eksperimen, memiliki isian-isian data yang tidak lengkap seperti data yang hilang, tidak valid dan atau hanya salah ketik. Selain itu terdapat atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan lebih baik dibuang dan tidak digunakan dalam proses. Pembersihan data akan mempengaruhi performasi dari teknik data mining. Karena data yang diproses akan berkurang jumlah dan kompleksitasnya.

1. Integrasi Data

Integrasi data merupakan penggabungan data dariberbagai database ke dalam satu database yang baru. Data yang digunakan dalam data mining tidak hanya didapatkan dari satu database namun juga didapatkan dari beberapa database atau *file teks*. Integrasi data dapat dilakukan pada atribut-atribut yang mengidentifikasikan entitas-entitas yang unik seperti atribut nama, jenis produk, nomer pelanggan dan lainya. Dalam melakukan integrasi data harus dilakukan dengan cermat agar hasil tidak menyimpang dan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk namun menggabungkan produk dari kategori yang berbeda, maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada. Dalam melakukan integrasi data diperlukan transformasi dan pemberisihan data dikarenakan sering kali data dari dua database berbeda cara penulisannya dan bahkan data yang ada disatu database tidak ada di database lainya.

1. Seleksi Data

Tidak semua data yang ada dalam database dipakai, oleh sebab itu hanya data yang sesuai akan diambil untuk dianalisa. Sebagai contoh sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus market basket analisis, tidak perlu mengambil nama pelanggan, tetapi cukup dengan id pelanggan.

1. Transformasi Data

Beberapa teknik data mining memerlukan format datayang khusus sebelum dapat diaplikasikan. Sebelum diproses dalam data mining data akan diubah dan di digabungkan ke dalam format yang sesuai. Beberapa metode data mining memerlukan format data khusus agar dapat diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagibagi menjadi beberapa interval. Dalam proses ini sering disebut transformasi data. Transformasi dan pemilihan data ini menentukan kualitas dari hasil data mining nantinya, karena ada beberapa karakteristik teknik data mining tertentu yang tergantung pada tahap ini.

1. Proses Mining

Tahap ini merupakan proses utama saat metode diterapkan untuk menemukan informasi atau pengetahuan yang berharga dan tersembunyi dari data.

1. Evaluasi Pola

Untuk mengidentifikasi pola-pola menarik kedalam knowledge based yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas ataupun model prediksi akan dievaluasi untuk menilai apakah hipotensa yang ada memang tercapai. Namun bila hasil yang didapatkan tidak sesuai hipotesa maka akan dilakukan beberapa alternatif, seperti menjadikan umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain dan menerima hasil ini sebagai hasil yang diluar dugaan yang mungkin bermanfaat.

1. Presentasi Pengetahuan

Tahap yang terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami tentang data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang mampu dipahami semua orang dalam satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga dapat membantu mengkomunikasikan hasil dari data mining.

**2.2.3 Cara Kerja Algoritma *Naïve Bayes Classifier***

Perhitungan Sederhana Naïve Bayes untuk Predeksi Kelulusan.

**Tabel 2.3.** Tabel Training

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **NO** | **JENIS KELAMIN** | **STATUS MAHASISWA** | **STATUS PRENIKAHAN** | **IPK Semester 1-6** | **STATUS PRESTASI** |
| **1** | LAKI – LAKI | MAHASISWA | BELUM | 3.17 | TERCAPAI |
| **2** | LAKI – LAKI | BEKERJA | BELUM | 3.30 | TERCAPAI |
| **3** | PEREMPUAN | MAHASISWA | BELUM | 3.01 | TERCAPAI |
| **4** | PEREMPUAN | MAHASISWA | MENIKAH | 3.25 | TERCAPAI |
| **5** | LAKI – LAKI | BEKERJA | MENIKAH | 3.20 | TERCAPAI |
| **6** | LAKI – LAKI | BEKERJA | MENIKAH | 2.50 | BELUM |
| **7** | PEREMPUAN | BEKERJA | MENIKAH | 3.00 | BELUM |
| **8** | PEREMPUAN | BEKERJA | BELUM | 2.70 | BELUM |
| **9** | LAKI – LAKI | BEKERJA | BELUM | 2.40 | BELUM |
| **10** | PEREMPUAN | MAHASISWA | MENIKAH | 2.50 | BELUM |
| **11** | PEREMPUAN | MAHASISWA | BELUM | 2.50 | BELUM |
| **12** | PEREMPUAN | MAHASISWA | BELUM | 3.50 | TERCAPAI |
| **13** | LAKI – LAKI | BEKERJA | MENIKAH | 3.30 | TERCAPAI |
| **14** | LAKI – LAKI | MAHASISWA | MENIKAH | 3.25 | TERCAPAI |
| **15** | LAKI – LAKI | MAHASISWA | BELUM | 2.30 | BELUM |

1. Jika seorang mahasiswa dengan data sebagai berikut;

**Tabel 2.3.1.** Tabel Testing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| KELAMIN | STATUS | PRENIKAHAN | IPK | KETERANGAN |
| LAKI – LAKI | MAHASISWA | BELUM | 2.70 | ??? |

1. Jawab :

Tahap 1 menghitung jumlah class/label :

P(Y= TERCAPAI)  = 8/15   ‘ jumlah data “TERCAPAI” pada komom ‘STATUS PRESTASI’ dibagijumlah data

 P(Y= BELUM) = 7/15 ‘ jumlah data “BELUM” pada komom ‘STATUS PRESTASI’ dibagi jumlah data

Tahap 2 menghitung jumlah kasus yang sama dengan class yang sama

1. P(JENIS KELAMIN = LAKI - LAKI | Y= TERCAPAI) = 5/8

‘jumlah data jenis kelamin “laki-laki” dengan keterangan “TERCAPAI” dibagi jumlah data TERCAPAI

P(JENIS KELAMIN = LAKI - LAKI | Y= BELUM) = 3/7

‘jumlah data jenis kelamin “laki-laki” dengan keterangan “BELUM” dibagi jumlah data BELUM

1. P(STATUS MAHASISWA = MAHASISWA | Y= TERCAPAI) = 5/8

‘jumlah data dengan status mahasiswa  dengan keterangan “TERCAPAI” dibagi jumlah data TERCAPAI

P(STATUS MAHASISWA = MAHASISWA | Y= BELUM) = 3/7

‘jumlah data dengan status mahasiswa  dengan keterangan “BELUM” dibagi jumlah data BELUM

1. P(STATUS PRENIKAHAN = BELUM| Y= TERCAPAI) = 4/8

‘jumlah data dengan status pernikahan “Belum menikah” dan  keterangan “TERCAPAI” dibagijumlah data TERCAPAI

P(STATUS PRENIKAHAN = BELUM| Y= BELUM) = 4/7

‘jumlah data dengan status pernikahan “Belum menikah” dan keterangan “BELUM”dibagi jumlah data BELUM

1. P(IPK = 2.70| Y= TERCAPAI) = 0/8

‘jumlah data IPK “2.70” dengan keterangan “TERCAPAI” dibagi jumlah data TERCAPAI

P(IPK = 2.70| Y= BELUM) = 1/7

‘jumlah data IPK “2.70” dengan keterangan “BELUM” dibagi jumlah data BELUM

Tahap 3 kalikan semua hasil variable TERCAPAI & BELUM

1. P (KELAMIN=LAKI – LAKI), (STATUS MHS=MAHASISWA), (PRENIKAHAN = BELUM), (IPK =2.70 ) | TERCAPAI)

= {P(P(KELAMIN =LAKI-LAKI|Y= TERCAPAI). P(STATUS MHS = MAHASISWA | Y= TERCAPAI) . P(PRENIKAHAN = BELUM|Y= TERCAPAI). P(IPK = 2.70| Y= TERCAPAI)

=    5/8    .      5/8 .   4/8   .   0/8   . 8/15

= 0

1. P (KELAMIN=LAKI – LAKI), (STATUS MHS=MAHASISWA), (PRENIKAHAN = BELUM), (IPK =2.70 ) |BELUM)

= {P(P(KELAMIN =LAKI-LAKI|Y= BELUM). P(STATUS MHS = MAHASISWA | Y= BELUM) . P(PRENIKAHAN = BELUM|Y= TERCAPAI). P(IPK = 2.70| Y= BELUM)

= 3/7  .  3/7    .  4/7 .  1/7 . 7/15

= 0,0069

Tahap 4 Bandingkan hasil class TERCAPAI & BELUM

Karena hasil (P|BELUM) lebih besar dari (P|BELUM) maka keputusanya adalah “BELUM”

**Tabel 2.4.** Tabel Hasil

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| KELAMIN | STATUS | PRENIKAHAN | IPK | KETERANGAN | |
| LAKI – LAKI | MAHASISWA | BELUM | 2.70 | | BELUM |

(Sumber : Marselina, Tahun 2010 dan Sakti Sarjono, Tahun 2010)

**2.2.4. Prestasi Akademik**

Prestasi akademik adalah hasil pelajaran yang diperoleh dari kegiatan belajar di sekolah atau perguruan tinggi yang bersifat kognitif dan biasanya ditentukan melalui pengukuran dan penilaian. Prestasi akademik pada penelitian ini dinilai berdasarkan IPK (Indeks Prestasi Kumulatif). Indeks Prestasi (IP) adalah penilaian keberhasilan studi semester yang dilakukan pada tiap akhir semester. Penilaian ini meliputi semua mata kuliah yang direncanakan mahasiswa dalam Kartu Rencana Studi (KRS).

*IP* =

∑

*n*

*I=*1

*I=*1

*Ni ki*

*n*



*ki*

Dimana :

1. IP : Indeks Prestasi
2. Ni : Nilai Mutu Kuliah i
3. K : Bobot SKS Mata Kuliah i
4. N : Jumlah Mata Kuliah i
   * 1. **Indeks Prestasi Kumulatif (IPK)**

IPK adalah angka yang menunjukkan prestasi atau keberhasilan studi mahasiswa dari semester pertama sampai dengan semester terakhir yang telah ditempuh secara kumulatif. IPK digunakan untuk :

a. Menentukan beban studi yang dapat diambil mahasiswa pada semester berikutnya.

b. Evaluasi Akademik per semester.

c. Evaluasi hasil studi pada akhir program.

* 1. **Metode *Naive Bayes***

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma mengunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Definisi lain mengatakan Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya.

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi paremeter yang diperlukan dalam proses pengklasifikasian. Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan.

**2.3.1. Persamaan Metode *Naive Bayes***

**𝑃(𝑋|𝐻). 𝑃(𝐻)**

**𝑃(𝐻|𝑋) =**

**𝑃(𝑋)**

Di mana :

*X* : Data dengan *class* yang belum diketahui

*H* : Hipotesis data merupakan suatu *class* spesifik

*P(H|X)* : Probabilitas hipotesis *H* berdasar kondisi *X* (posteriori probabilitas)

*P(H)* : Probabilitas hipotesis *H* (prior probabilitas)

*P(X|H)* : Probabilitas *X* berdasarkan kondisi pada hipotesis *H*

*P(X)* : Probabilitas *X*

Untuk menjelaskan metode *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode *Naive Bayes* di atas disesuaikan sebagai berikut:

**(𝐶)(𝐹1…𝐹𝑛|𝐶)**

**(𝐹1…𝐹𝑛)**

**(𝐶|𝐹1…𝐹𝑛) =**

Di mana Variabel *C* merepresentasikan kelas, sementara variabel *F1 ... Fn* merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas *C* (*Posterior*) adalah peluang munculnya kelas *C* (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas *C* (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik*-*karakteristik sampel.

**𝑝𝑟𝑖𝑜𝑟 𝑥 𝑙𝑖𝑘𝑒𝑙𝑖ℎ𝑜𝑜𝑑**

**𝑒𝑣𝑖𝑑𝑒𝑛𝑐𝑒**

**𝑃𝑜𝑠𝑡𝑒𝑟𝑖𝑜𝑟** =

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai-nilai *posterior* kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus *Bayes* tersebut dilakukan dengan menjabarkan (𝐶|𝐹1,…,) menggunakan aturan perkalian sebagai berikut :

(𝐶|𝐹1,…,=𝑃(𝐶)𝑃(𝐹1,…,𝐹𝑛|𝐶)

=(𝐶)𝑃(𝐹1|𝐶)𝑃(𝐹2,…,𝐹𝑛|𝐶,𝐹1)

=𝑃(𝐶)𝑃(𝐹1|𝐶)𝑃(𝐹2|𝐶,𝐹1 )𝑃(𝐹3,…,𝐹𝑛|𝐶,𝐹1,𝐹2

=(𝐶)𝑃(𝐹1|𝐶)𝑃(𝐹2|𝐶,𝐹1 )𝑃(𝐹3|𝐶,𝐹1,𝐹2)𝑃(𝐹4,…,𝐹𝑛|𝐶,𝐹1,𝐹2,𝐹3)

=𝑃(𝐶)𝑃(𝐹1|𝐶)𝑃(𝐹2|𝐶,𝐹1 )𝑃(𝐹3|𝐶,𝐹1,𝐹2)…𝑃(𝐹𝑛|𝐶,𝐹1,𝐹2,𝐹3,…,𝐹𝑛−1)

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor - faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Di sinilah digunakan asumsi independensi yang sangat tinggi (*naif*), bahwa masing-masing petunjuk *(F1,F2...Fn)* saling bebas *(independen)* satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

**𝑃(𝐹𝑖|𝐹𝑗) =**

**𝑃(𝐹𝑗)**

**𝑃(𝐹𝑗)**

**𝑃(𝐹𝑗)**

**𝑃(𝐹𝑖)𝑃(𝐹𝑗)**

**𝑃(𝐹𝑖∩ 𝐹𝑗)**

**=**

**2.3.2. Kelebihan dan Kekurangan *Naïve Bayes***

* + - * 1. **Kelebihan *Naïve Bayes Classifier***

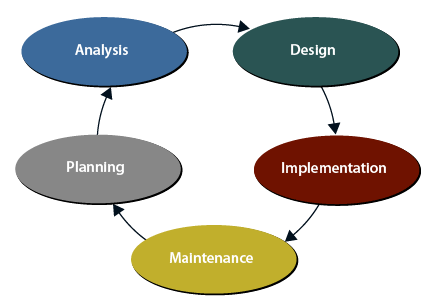
Kelebihan dari penggunaan *Naïve Bayes Classifier* dalam Sasaran Kinerja Pegawai dapat di tinjau dari prosesnya yang mengambil aksi berdasarkan data-data yang telah ada sebelumnya. Oleh karena itu, Sasaran Kinerja Pegawai dengan metode ini dapat dipersonalisasi, maksudnya adalah proses Sasaran Kinerja Pegawai dapat di sesuaikan sesuai dengan sifat dan kebutuhan masing-masing orang.

* + - * 1. **Kekurangan *Naïve Bayes Classifier***

Kekurangan dari metode *Naïve Bayes Classifier* ini adalah banyaknya celah untuk mengurangi keefektifan metode in dan akibatnya meloloskan dokumen kedalam kelas tertentu padahal jelas-jelas dokumen tersebut tidak layak berada di kelas tersebut. (sumber ; Sem Tahun 2011).

**2.3.3. Siklus Hidup Pengembangan Sistem**

Menurut Jogiyanto (2005 :41), Proses pengembangan sistem melewati beberapa tahapan dari mulai dari sistem itu direncanakan sampai dengan sistem itu direncakan sampai dengan sistem tersebut diterapkan, dioperasikan dan dipelihara. Bila operasi yang sudah dikembangkan masih timbul kembali permasalahan-permasalahan yang tidak dapat diatasi dalam tahap pemeliharaan. Maka perlu dikembangkan kembali suatu sistem untuk mengatasinya dan proses ini kembali ke tahap yang pertama, yaitu tahap perencanaan sistem. Siklus ini disebut dengan siklus hidup suatu sistem *(systems life cycle)*. Daur atau siklus hidup dari pengembangan sistem merupakan suatu bentuk yang digunakan untuk menggambarkan tahapan utama dan langkah-langkah didalam tahapan tersebut dalam proses pengembangannya. Berikut langka-langkah yang digunakan :



**Gambar 2.4.** Siklus Hidup Pengembangan Sistem

**2.3.4 Analisis Sistem**

Model analisis perangkat lunak, seperti kita ketahui, sesungguhnya merupakan spesifikasi rinci dari kebutuhan-kebutuhan pengguna dan bekerja sebagai Iangkah pertama saat kita kelak akan mengembangkan model perancangan perangkat lunak. (Nugroho, 2010)

Setelah proses perencanaan atau *planing* tahapan selanjutnya adalah analisis terhadap data – data mengenai kebutuhan pengguna yang dipadukan dengan

hasil studi kelayakan yang meliputi data-data ketersediaan sumber daya yang ada seperti manusia. infrastruktur teknologi yang tersedia dan lain-lain, yang kemudian menetapkan solusi yang tepat untuk diterapkan. Pada tahapan ini seorang analis sistem melakukan pengenalan terhadap segenap permasalahan yang timbul pada penguna melalui dekomposisi dan realisasi permasalahan lebih lanjut, mengenali komponen-komponen sistem/perangkat lunak, objek - objek. Hubungan antar obiek

dan sebagainya, dimana secara keseluruhan dinotasikan dalam bentuk *use case.*

**2.3.5. Design Sistem**

  Menurut Robert J. Verzello dan John Reuter yang dikutip Jogianto Hartono (2005: 196) Desain sistem adalah tahap setelah analisis dari siklus pengembangan sistem, pendefenisisan dari kebutuhan-kebutuhan fungsional dan persiapan untuk rancang bangun implementasi, menggambarkan bagaimana suatu sistem dibentuk.

Menurut John Burch dan Gerry Grodnetski yang dikutip Jogianto Hartono (2005:196) Desain sistem adalah dapat di identifikasikan sebagai penggambaran perencanaan dalam pembuatan sketsa atau pengaturan dari beberapa elemen yang terpisah ke dalam satu kesatuan yang utuh dan berfungsi.

Desain sistem dibagi menjadi dua yaitu desain sistem secara umum (*general system design)* dan desain sistem secara terinci (*detailed system design)*.

**2.3.5.1. Desain sistem secara umum**

Tujuan dari desain sistem secara umum adalah untuk memberikan gambaran secara umum kepada user tentang sistem yang baru. Desain sistem secara umum merupakan persiapan dari desain terinci. Desain secara umum mengidentifikasikan komponen-komponen sistem informasi yang akan didesain secara rinci. Desain terinci dimaksudkan untuk pemrogram komputer dan ahli teknik lainnya yang akan mengimplementasi sistem. Tahap desain sistem secara umum dilakukan setelah tahap analisis sistem selesai dilakukan dan hasil analisis disetujui oleh manajemen.

Analisis sistem dapat mendesain model dari sistem informasi yang diusulkan dalam bentuk physical systems dan logical model. Bagan alir sistem (*systemsflowchart*) merupakan alat yang tepat digunakan untuk menggambarkan physicalsystems. Simbol-simbol bagan allir sistem ini menunjukkan secara tepat arti fisiknya, seperti simbol terminal, hard disk, laporan-laporan.

Logical model dari sistem informasi lebih menjelaskan kepada user bagaimana nantinya fungsi-fungsi di sistem informasi secara logika akan bekerja. Logical model dapat digambar dengan menggunakan diagram arus data (*data flow diagram*). Arus dari data di diagram arus data dapat dijelaskan dengan menggunakan kamus data (*data dictionary*).

Sketsa dari physical systems dapat menunjukkan kepada user bagaimana nantinya sistem secara fisik akan diterapkan. Pengolahan data dari sistem informasi berbasis komputer membutuhkan metode-metode dan prosedur-prosedur. Metode-metode dan prosedur-prosedur ini merupakan bagian dari model sistem informasi (model prosedur) yang akan mendefinisikan urut-urutan kegiatan untuk menghasilkan output dari input yang ada. Metode (method) adalah suatu cara untuk melakukan suatu kegiatan. Suatu prosedur merupakan rencana tahap demi tahap untuk menerapkan suatu metode. Bagan alir sistem (systems flowchart) merupakan alat berbentuk grafik yang dapat digunakan untuk menunjukkan urut-urutan kegiatan dari sistem informasi berbasis komputer ini. Seringkali gambar bagan alir sistem untuk sistem informasi juga dapat digabung dengan bagan alir formulir dalam perusahaan untuk menunjukkan hubungan dan prosedur antara sistem informasi dengn sistem-sistem lainnya di perusahaan.

Pada tahap desain secara umum, komponen-komponen sistem informasi dirancang dengan tujuan untuk dikomunikasikan kepada user bukan untuk pemrogram. Komponen sistem informasi yang didesain adalah :

1. Desain Model

Analisis sistem dapat mendesain model dari sistem informasi yang diusulkan dalam bentuk physical sistem dan logical model. Bagan alir sistem merupakan alat yang tepat digunakan untk menggambarkan physical system, logical model dapat digambar dengan diagram arus data. (Jogiyanto : 2005 : 211).

Bagan alir sistem merupakan bagan yang menunjukan arus pekerjaan secara keseluruhan dari sistem. Bagan alir sistem digambar dengan simbol-simbol sebagai berikut :

**Tabel 2.5.** Simbol – Simbol Bagan Alir Sistem

|  |  |
| --- | --- |
| Simbol Dokumen | Menunjukan dokumen *input* dan *output* baik untuk proses manual, mekanik atau komputer |
| Pekerjaan Manual | Menunjukan Pekerjaan Manual |
| Proses | Menunjukan kegiatan proses dari operasi program komputer |
| Simpanan Offline  C  N | File non komputer yang diarsip urut angka *(numerical),* dan file non komputer yang diarsip urut tanggal *(chronological)* |
| Simpanan Offline  N | File non komputer yang diarsip urut huruf *(alphabetical)* |
| Kartu Plong | Menunjukan input/output yang menggunakan kartu plong (punched card) |
| Pita Magnetik | Menunjukan input/output menggunakan pita magnetik |
| Harddisk | Menunjukan input/output menggunakan harddisk |
| Diskette | Menunjukan input/output menggunakan diskette |
| Pita Kertas berlubang | Menunjukan input/output menggunakan pita kertas berlubang |
| Keyboard | Menunjukan input menggunakan on-line keyboard |
| Display | Menunjukan output yang ditampilkan di monitor |
| Garis Alir | Menunjukan arus dari proses |
| Penjelasan | Menunjukan penjelasan dari suatu proses |
| Penghubung | Menunjukan penghubung ke halaman yang masih sama |

(Sumber : Jogiyanto 2005).

Untuk mempermudah penggambaran suatu sistem yang ada atau sistem yang baru yang akan dikembangkan secara logika tanpa memperhatikan lingkungan fisik dimana data tersebut mengalir atau lingkungan fisik dimana data tersebut akan disimpan, maka digunakan Diagram Arus Data (DAD) atau Data Flow Diagram (DFD). Dalam menggambarkan sistem perlu dilakukan pembentukan simbol, berikut ini simbol-simbol yang sering digunakan DAD:

1. *External entity* (kesatuan luar) atau *boundary* (batas sistem)

Setiap sistem pasti mempunyai batas sistem *(boundary)* yang memisahkan suatu sistem dengan lingkungan luarnya. Sistem akan menerima input dan menghasilkan output kepada lingkungan luarnya. Kesatuan luar *(external entity)* merupakan kesatuan di lingkungan luar sistem yang dapat berupa orang, organisasi atau sistem lain yang berada di lingkungan luarnya yang akan memberikan input serta menerima output dari sistem. (sumber ; Jogiyanto, 2000).

**Gambar.** Notasi Kesatuan Luar

1. *Data Flow* (Arus Data)

Arus data ini menunjukkan arus atau aliran data yang dapat beupa masukkan untuk sistem atau hasildari proses sistem. (sumber ; Jogiyanto, 2000)

Gambar Arus Data

1. *Proses* (Memproses Data)

*Process* (proses). Suatu proses adalah kegiatan atau kerja yang dilakukan oleh orang, mesin atau komputer dari hasil suatu arus dara yang masuk kedalam proses untuk dihasilkan arus data yang masuk ke dalam proses untuk dihasilkan arus data yang akan keluar dari proses. (Jogiyanto, 2005)

Identifikasi

Nama

Proses

**Gambar** .Proses Data

1. *Data Store* (Simpanan Data)

Simpanan data pada DAD dapat disimbolkan dengan sepasang garis horizontal paralel yang tertutup di salah satu ujungnya. (sumber ; Jogiyanto,2000)

Media Data Store

**Gambar 2.7** Penyimpanan Data (*Data* *Storage*)

1. Desain Output

Output (keluaran) adalah produk dari sistem informasi yang dapat dilihat. Istilah output ini kadang-kadang membingungkan, karena output dapat terdiri dari macammacam jenis. Output dapat berupa hasil di media keras (seperti misalnya kertas, microfilm) atau hasil di media lunak (berupa tampilan di layar video). Disamping itu output dapat berupa hasil dari suatu proses yang akan digunakan oleh proses lain dan tersimpan di suatu media seperti tape, disk atau kartu. Yang akan dimaksud dengan 5 Desain Sistem Secara Umum output pada tahap desain ini adalah output yang berupa tampilan di media keras atau di layar video.

1. Desain Input

Bila berpikir tentang input, biasanya juga akan berpikir tentang alat input (*inputdevice*) yang akan digunakan, semacam keyboard, card reader dan lain sebagainya.Alat input dapat digolongkan ke dalam 2 golongan, yaitu alat input langsung (*online input device*). Alat input langsung merupakan alat input yang langsung dihubungkandengan CPU, misalnya adalah keyboard, mouse, touch screen dan lain sebagainya.Alat input tidak langsung adalah alat input yang tidak langsung dihubungkan denganCPU, misalnya KTC (key-to-card), KTT (key-to-tape) dan KTD (key-to-disk).

1. Desain Database

Basis data (database) merupakan kumpulan dari data yang saling berhubungan satu dengan yang lainnya, tersimpan di simpanan luar komputer dan digunakan perangkat lunak tertentu untuk memanipulasinya. Database merupakan salah satu komponen yang penting di sistem informasi, karena berfungsi sebagai basis penyedia informasi bagi para pemakainya. Penerapan database dalam sistem informasi disebut dengan database system. Sistem basis data (database system) ini adalah suatu sistem informasi yang mengintegrasikan kumpulan dari data yang saling berhubungan satu dengan lainnya dan membuatnya tersedia untuk beberapa aplikasi yang bermacammacam di dalam suatu organisasi.

Untuk tahap desain database secara umum, yang perlu dilakukan oleh analis adalah mengidentifikasi terlebih dahulu file-file yang diperlukan oleh sistem informasi. File-file database yang dibutuhkan oleh sistem dapat dilihat pada desain model yang digambarkan dalam bentuk diagram arus data.

**2.3.5.1.1 . Desain sistem secara terinci (detailed system design).**

1. Desain Input

Masukan merupakan awal dimulainya proses informasi. Bahan mentah dari informasi adalah data yang terjadi dari transaksi-transaksi yang dilakukan oleh organisasi. Data hasil transaksi merupakan masukan untuk sistem informasi. Hasil dari sistem informasi tidak lepas dari data yang dimasukan. Disain input terinci. Dimulai dari desain dokumen dasar sebagai penangkap input yang pertama kali. Jika dokumen dasar tidak di desain dengan baik, kemungkinan input yang tercatat dapat salah bahkan kurang.

1. Desain Ouput

Desain output terinci dimaksudkan untuk mengetahui bagaiman dan seperti apa bentuk output-output dari sistem yang baru. Desain output terinci terbagi atas dua yaitu desain output berbentuk laporan dimedia kertas dan desain output dalam bentuk dialog pada layar terminal.

1. Desain Database

Basis data (Database) merupakan kumpulan dari data yang saling berhubungan satu dan dengan yang lainnya, tersimpan disimpanan luar komputer dan digunkan perangkat lunak tertentu untuk memanipulasinya. Database merupakan salah satu komponen yang penting disitem informasi karna berfungsi sebagai basis penyedia informasi bagi para pemakainya. Penerapan database dalam sistem nformasi disebut database system. Sistem basis data (database system) adalah suatu sistem informasi yang mengintegrasikan kumpulan dari data yang salin berhubungan dengan satu dengan yang lainnya dan membuatnya tersedia untuk bebrapa aplikasi yang bermacam-macam didalam suatu organisasi. Dalam sisytem basis data, tiap-tiap orang atau bagian dapat memandang database dari beberapa sudut pandang yang berbeda.

**2.3.6. Implementasi**

*Whitten, et al* (2004:34) mengungkapkan “*System Implementation* adalah konstruksi, instalasi, pengujian dan pengiriman sistem ke dalam produksi. Tahap implementasi sistem merupakan tahap meletakkan sisitem supaya siap untuk di operasikan.

Implementasi sistem mengkonstruksi sistem informasi baru dan menempatkannya ke dalam operasi. Selama implementasi, sistemlah perangkat keras dan perangkat lunaksistem yang baru di instal dan di uji. Semua perangkat lunak aplikasi dan database yang di beli, akan di instal dan di konfigurasi. Semua perangkat lunak dan database *custom* akan di konstruksi dengan menggunakan cetak biru dan spesifikasi teknis yang dikembangkan selama desain sistem.

Tahap Implementasi sistem dapat terdiri dari langkah langkah sebagai berikut:

1. Menerapkan Rencana Implementasi

Rencana Implementasi merupakan kegiatan awal dari tahap implementasi sistem. Rencana Implementasi dimaksudkan terutama untuk mengatur biaya dan waktu yang dibutuhkan selama tahap implementasi.

1. Melakukan Tahap Implementasi

Kegiatan implementasi dilakukan dengan dasar kegiatan yang telah direncanakan dalam tahap implementasiini adlah sebagai berikut:

1. Pemeliharaan dan pelatihan personil

Telah diketahui bahwa manusia merupakan faktor yang perlu di pertimbagkan dalam sistem informasi. Jika sistem informasi ingin sukses, maka personil personil yang terlibat harus diberi pengertian dan pengeyahuan yang cukup tentang sistem informasi dan posisi serta tugas mereka.

1. Persipan tempat dan instalasi perangkat kerras dan perangkat lunak

Jika peralatan baru akan dimiliki, maka temapt atau ruangan untuk peralatan ini perlu dipersiapkan terlebih dahulu. Kemanan fisik dari tempat ini perlu juga dipertimbangkan. Sistem komputer yang besar membutuhkan temapy dengan lingkungan yang lebih harus diperhitungkan. Langkah selajutnya setelah persipana fisik tempat adalah menginstalasi pegnakat keras yang sudah dikirim dan menginstalasi perangkat lunak yang sudah ada.

1. Pemograman dan pengetesan sistem

Pemograman merupakan kegiatan menulis kode program yang akan di eksekusi oleh komputer. Kode program yang dituliskan oleh pemogram harus bedasarkan dokumentasi yang disediakan oleh analis sistem hasil dari desain sistem secara rinci. Sebelum program diterapkan, maka program harus terlebih dahulu bebas dari kesalahan kesalahan. Oleh sebab itu, program harus di uji untuk menemukan kesalahan kesalahan yang mungkin dapat terjadi. Prgram di tes untuk tiap tiap modul dan dilanjutkan dengan pengetasan untuk semua modul yang telah dirangkai.

1. Pengetesan sistem

Pnegetesan sistem biasanya dilakukan setelah pengetesan program. Pengetesan sisitem biasanya dilakukan untuk memeriksa kekompakan anatara kompoen sistem yang di implementasikan. Tujuan untama dari pengetesan sisitem ini adlah untuk memastikan elemen elemen atau komponen komponen dari sistem telah berfungsi sesuai dengan yang di harapkan.

**2.3.7. Perencanaan**

Perencanaan sistem menyangkut estimasi dari kebutuhan-kebutuhan fisik, tenaga kerja dan dana yang dibutuhkan untuk mendukung pengembangan sistem ini serta untuk mendukung operasinya setelah diterapkan.(chandra Pamungkas : 2012).

Proses Perencanaan Sistem dapat dikelompokkan dalam 3 proses utama yaitu sbb :

1. Merencanakan proyek-proyek sistem yang dilakukan oleh staf perencana system.
2. Menentukan proyek-proyek sistem yang akan dikembangkan dan dilakukan oleh komite pengarah.
3. Mendefinisikan proyek-proyek sistem dikembangkan dan dilakukan oleh analis sistem.

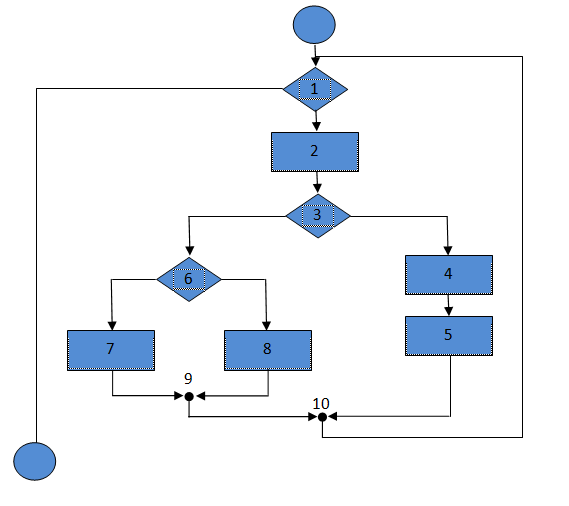
**2.4. Tehnik Pengujian Sistem**

**2.4.1 *White Box***

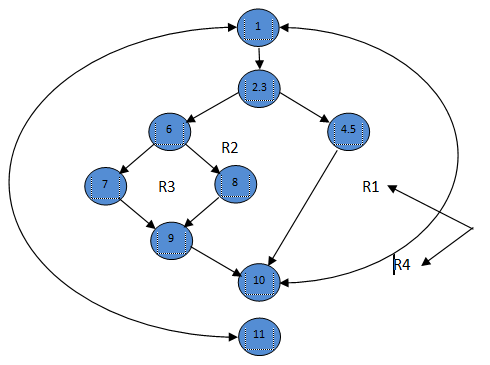
Pengujian perangkat lunak adalah elemen kritis dari jaminan kualitas perangkat lunak dan merepresentasikan kajian pokok dari spesifikasi, desain dan pengkodean. Pengujian sistem/perangkat lunak memiliki sejumlah aturan yang berfungsi sebagai sasaran pengujian, diantaranya adalah sebagai berikut:

1. Pengujian adalah proses eksekusi suatu program dengan maksud menemukan kesalahan.
2. *Test case* yang baik adalah *test case* yang memiliki probabilitas tinggi untuk menemukan kesalahan yang belum pernah ditemukan sebelumnya.
3. Pengujian yang sukses adalah pengujian yang mengungkap semua kesalahan yang belum pernah ditemukan sebelumnya.

Pengujian *White Box,* adalah metode pengujian yang menggunakan struktur kontrol desain prosedur untuk memperoleh *test case.* Dengan menggunakan metode *white box,* perekayasa sistem dapat melakukan *test case* yang memberikan jaminan bahwa semua jalur independen pada suatu modul telah digunakan paling tidak satu kali, menggunakan semua keputusan logis pada sisi *true* dan *false,* mengeksekusi semua *loop* pada batasan mereka dan pada batas operasional mereka, dan menggunakan stuktur data internal untuk menjamin validitasnya. Pengujian *Basis Path* adalah teknik pengujian *white box* yang diusulkan pertama kali oleh Tom McCabe. Metode *basis path* ini memungkinkan desainer *test case* mengukur kompleksitas logis dari desain prosedural dan menggunakannya sebagai pedoman untuk menetapkan basis set dan jalur eksekusi. (Roger S. Pressman, 2002 : *536)*

*.***Gambar 2.6.** Contoh Bagan Alir

Bagan alir digunakan untuk menggambarkan struktur kontrol program. Dan untuk menggambarkan grafik alir, harus memperhatikan representasi desain prosedural pada bagan alir. Pada gambar dibawah ini, grafik alir memetakan bagan alir tersebut ke dalam grafik alir yang sesuai (dengan mengasumsikan bahwa tidak ada kondisi senyawa yang diisikan di dalam diamond keputusan dari bagan alir tersebut). Masing-masing Iingkaran, yang disebut *simpul* grafik alir, merepresentasikan satu atau lebih statemen prosedural. Urutan kotak proses dan permata keputusan dapat memetakan simpul tunggal. Anak panah tersebut yang disebut *edges* atau *link,* merepresentasikan aliran kontrol dan analog dengan anak panah bagan alir. Edge harus berhenti nada suatu simpul, meskipun bila simpul tersebut tidak mempresentasikan statemen prosedural. (Pressman. 2002 : *536).*



**Gambar 2.7.** Contoh Grafik Alir

*Kompleksitas sikiomatis* adalah metriks perangkat lunak yang memberikan pengukuran kuantitatif terhadap kompleksitas logis suatu program. Bila metriks ini digunakan dalam konteks metode pengujian basis path, maka nilai yang terhitung untuk kompleksitas siklomatis menentukan jumlah *jalur independen.* Jalur independen adalah jalur yang melalui program yang mengintroduksi sedikitnya satu rangkaian statemen proses baru atau suatu kondisi baru. Bila dinyatakan dengan terminologi grafik alir, jalur independen harus bergerak sepanjang paling tidak satu edge yang tidak dilewatkan sebelum jalur tersebut ditentukan. Sebagai contoh, serangkaian jalur independen untuk gratik alir yang ditunjukkan pada gambar 2.7 adalah:

Jalur l : 1—11

Jalur 2 : 1—2—3 *—4—5—10—1—11*

Jalur 3 : 1—2—3 —6—8—9—10—I—11

Jalur 4 : 1—2—3 —6—7—9—10—1—11

Jalur 1, 2, 3, dan 4 yang ditentukan di atas terdiri dari sebuah basis set untuk grafik alir pada gambar 2.7. Bagaimana kita tahu banyaknya jalur yang dicari? Komputasi kompleksitas siklomatis memberikan jawaban. Fondasi kompleksitas siklomatis adalah teori grafik, dan memberi kita metriks perangkat lunak yang sangat berguna. Kompleksitas dihitung dalam salah satu dari tiga cara berikut:

1. Jumlah *region* grafik alir sesuai dengan kompleksitas siklomatis.
2. Kompleksitas siklomatis, *V(G),* untuk grafik alIr G ditentukan sebagai   
   *V(G)* = *E - N* + 2 di mana *E* adalah jumlah *edge* grafik alir dan *N* adalah jumlah simpul grafik alir.
3. Kompleksitas siklomatis, *V(G),* untuk grafik alir G juga ditentukan sebagai *V(G)* = P + 1, dimana P adalah jumlah simpul predikat yang dilsikan dalam grafik alir G.

Pada gambar 2.7 grafik alir, kompleksitas siklomatis dapat dihitung dengan menggunakan masing-masing dari algoritma yang ditulis di atas:

1. Grafik alir mempunyai 4 region
2. V(G) = 11 edge – 9 simpul + 2 = 4
3. *V(G) =* 3 simpul yang diperkirakan + 1 = 4

Dengan demikian, kompleksitas siklomatis dan grafik alir pada gambar 2.7 adalah 4. Yang Iebih penting, nilai untuk *V(G)* memberi kita batas atas untuk jumlah jalur independen yang membentuk basis set, dan implikasinya, batas atas jumlah pengujian yang harus didesain dan dieksekusi untuk menjamin semua statemen program.

**2.4.2 *Black Box***

*Black box aproach* adalah suatu sistem dimana input dan outputnya dapat didefinisikan tetapi prosesnya tidak diketahui atau tidak terdefinisi. Metode ini hanya dapat dimengerti oleh pihak dalam (yang menangani sedangkan pihak luar hanya mengetahui masukan dan hasilnya). Sistem ini terdapat pada subsistem tingkat terendah.

Metode ujicoba *black box* memfokuskan pada keperluan fungsional dan *software.* Karena itu ujicoba *black box* memungkinkan pengembang *software* untuk membuat himpunan kondisi input yang akan melatih seluruh syarat-syarat fungsional suatu program. Ujicoba *black box* bukan merupakan alternatif dari uji coba *white box,* tetapi merupakan pendekatan yang melengkapi untuk menemukan kesalahan Iainnya, selain menggunakan metode *white box.* Ujicoba *black box* berusaha untuk menemukan kesalahan dalam beberapa kategori, diantaranya:

1. Fungsi-fungsi yang salah atau hilang.
2. Kesalahan *interface*.
3. Kesalahan dalam struktur data atau akses database ekstemal.
4. Kesalahan performa dan kesalahan inisialisasi.

Tidak seperti metode *white box* yang dilaksanakan diawal proses, ujicoba *black box* diaplikasikan dibeberapa tahapan berikutnya. Karena ujicoba *black box* dengan sengaja mengabaikan struktur kontrol. sehingga perhatiannya difokuskan pada informasi *domain.* Ujicoba didesain untuk dapat menjawab pertanyaan pcrtanyaan berikut:

1. Bagaimana validitas fungsionalnya diuji?
2. Jenis input seperti apa yang akan menghasilkan kasus uji yang balk?
3. Apakah sistem secara khusus sensitive terhadap nilai input tertentu?
4. Bagaimana batasan-batasan kelas data diisolasi?
5. Berapa rasio data dan jumlah data yang dapat ditoleransi oleh sistem?
6. Apa akibat yang akan timbul dan kombinasi spesifik data pada operasi sistem?

Dengan mengaplikasikan ujicoba *black box*, diharapkan dapat menghasilkan sekumpulan kasus uji yang memenuhi kriteria berikut:

1. Kasus uji yang berkurang, jika jumlahnya lebih dari 1, maka jumlah dari uji kasus tambahan harus didesain untuk mencapai ujicoba yang cukup beralasan.
2. Kasus uji yang memberitahukan sesuatu tentang keberadaan atau tidaknya suatu jenis kesalahan, dari pada kesalahan yang terhubung hanya denan suatu ujicoba yang spesifik**.**

**2.5. Perangkat Lunak pendukung**

**Tabel 2.8.** Tabel Perangkat Lunak Pendukung

|  |  |  |
| --- | --- | --- |
| **NO** | **Tools Yang digunakan** | **Kegunaan/Fungsi** |
|
| **1** | **PHP** | PHP ( *PHP : Hypertext Preprocessor*) adalah bahasa server – side scripting yang menyatu dengan HTML untuk membuat halaman web yang dinamis.  Salah satu keunggulan yang dimiliki oleh PHP adalah kemampuannya untuk melakukan koneksi keberbagai macam software manajemen basis data/Databases Managemen Sistem (DBMS). (PHP dan MySQL).  (sumber ; M Rudyanto Arief, 2011) |
| **2** | **MySQL** | MySQL server adalah salah satu MySQL (*My Strukture Query Languange*) sebagai bahasa dasar untuk menagkses databasenya. Selain itu bersifat free (tidak perlu membayar untuk menggunakannya) berbagai platform (kecuali pada windows, yang bersifat software atau perlu membayar setelah evaluasi dan memutuskan untuk digunakan pada keperluan produksi). Mysql termasuk jenis RDBMS (*Relations Database Managemen Sistem*). Itulah sebabnya istilah seperti table, baris, dan kolom digunakan dalam MySQL. |

**2.2.7 Konsep Klasifikasi**

Klasifikasi merupakan pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu (1) pembangunan model sebagai prototipe untuk disimpan sebagai memori dan (2) penggunaan model tersebut untuk melakukan pengenalan/ klasifikasi/ prediksi pada suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang sudah disimpannya (Prasetyo, 2012).

Klasifikasi adalah metode data mining yang dapat digunakan untuk proses pencarian sekumpulan model (fungsi) yang dapat menjelaskan dan membedakan kelas-kelas data atau konsep, yang tujuannya supaya model tersebut dapat digunakan memprediksi objek kelas yang labelnya tidak diketahui atau dapat memprediksi kecenderungan data-data yang muncul di masa depan. Metode klasifikasi juga bertujuan untuk melakukan pemetaan data ke dalam kelas yang sudah didefinisikan sebelumnya berdasarkan pada nilai atribut data (Han dan Kamber, 2006). Proses klasifikasi tersebut seperti terlihat pada gambar berikut :

**Algoritma Pelatihan**

**Masukkan Data Latih (x,y)**

**Gambar 2.9. Konsep Klasifikasi**

**Penerapan Model**

**Keluaran Data Uji (x,y)**

**Masukkan Data Uji (x,?)**

**Pembangunan Model**