

## **BAB 2**

### **TINJAUAN PUSTAKA**

#### **2.1. Traveloka**

Traveloka adalah perusahaan teknologi terkemuka di Asia Tenggara yang memungkinkan pengguna untuk menemukan dan memesan beragam produk transportasi, akomodasi, gaya hidup, dan layanan keuangan. Portofolio komprehensif yang dimiliki oleh Traveloka termasuk pemesanan tiket untuk transportasi, seperti tiket pesawat, bus, kereta api, penyewaan mobil, *airport transfer*, serta inventaris akomodasi terbesar di Asia Tenggara, termasuk hotel, apartemen, *guest house*, *homestay*, *resort*, dan *villa*, menjadikan traveloka sebagai *platform* pemesanan dengan pilihan akomodasi dan paket terlengkap.

Traveloka juga merupakan pelaku industri utama di sektor gaya hidup yang menyajikan pesanan untuk beragam tempat rekreasi dan kegiatan, konektivitas, serta direktori kuliner. Melalui produk layanan keuangan, Traveloka juga menawarkan solusi keuangan, pembayaran, dan asuransi untuk masyarakat underbanked guna membantu pengguna dalam mengatasi kendala keuangan dalam memenuhi kebutuhan perjalanan dan gaya hidup mereka, seperti *Paylater*, *Paylater Credit Card* serta produk asuransi yang memudahkan pengguna untuk mendapatkan perlindungan menyeluruh. Traveloka memiliki layanan *customer service* yang tersedia 24/7 dalam bahasa lokal, serta kemudahan bertransaksi dengan lebih dari 40 metode pembayaran, online dan offline [6].

#### **2.2. Web Scraping**

*Web Scraping* merupakan teknik mengumpulkan data selain mengumpulkan data melalui API, teknik ini sering digunakan karena kita hanya membuat program

otomatis yang berhubungan dengan *server web* yang kemudian program tersebut dapat langsung mendapatkan data [7].

### **2.3. Text Mining**

Teks memiliki sifat tidak terstruktur, tidak berbentuk, dan susah untuk ditangani. Meskipun demikian teks adalah media untuk pertukaran informasi. Untuk mendapatkan pola dan informasi dari teks digunakan metode text mining. Pengertian text mining adalah proses menganalisis teks untuk mengambil intisari informasi yang bermanfaat untuk keperluan tertentu. Tujuan dari text mining adalah mencari pola dan informasi dari suatu teks [8].

### **2.4. Preprocessing**

Sebuah data yang didapat dari hasil *web scraping* belum bisa langsung diklasifikasikan karena data tersebut masih terdapat simbol dan kata-kata yang tidak diperlukan. Namun ada beberapa tahap yang dilakukan dalam *preprocessing* ini yang yaitu : *Case Folding*, *Tokenizing*, *Cleaning*, *Stemming*, dan *Stopword Removal*. [9].

#### **1. Case Folding**

*Case Folding* adalah proses penyeragaman bentuk huruf menjadi huruf kecil (*lowercase*). Dalam hal ini hanya menerima huruf latin dari a hingga z. Dimana dalam jika ditemukan dalam suatu dokumen akan terdapat beberapa huruf saja memiliki huruf kapital seperti awalan kalimat, nama orang, nama kota, dll[1].

#### **2. Tokenizing**

*Tokenizing* adalah sebuah proses memecah sekumpulan karakter dalam sebuah teks ke dalam sebuah kata. Sebagai contoh karakter *whitespace* seperti *enter*, spasi, tabulasi dianggap sebagai pemisah kata. Namun karakter petik

tunggal (‘), titik (.), semikolon (;), titik dua (:) atau lainnya dapat memiliki peran sebagai pemisah kata. [1]

### 3. *Cleaning*

Pada tahap *cleaning* akan dilakukan pembersihan kata atau simbol yang tidak diperlukan pada proses hasil klasifikasi sentimen. Karena dalam ulasan hotel pasti terdapat banyak atribut-atribut yang tidak terpakai maka pada tahap ini akan dilakukan penghapusan simbol atau emoticon, penghapusan angka, dan juga penghapusan tanda baca [2].

### 4. *Stemming*

*Stemming* merupakan suatu proses untuk mengubah kata-kata yang terdapat dalam suatu dokumen ke dalam kata akarnya dengan menggunakan aturan tertentu. Proses stemming bahasa Indonesia dilakukan dengan menghilangkan suffix, prefix, dan konfiks pada dokumen. Ada beberapa tahap pada proses stemming yang akan digunakan yaitu :

1. Cari kata yang akan distem dalam kamus. Jika diteukan maka asumsikan bahwa kata tersebut adalah rootword. Maka metode berhenti
2. Inflection Suffixes (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dibuang. Jika berupa particles (“-lah”, “-kah”, “-tah” atau “-pun”) maka langkah ini diulangi lagi untuk menghapus Possesive Pronouns (“-ku”, “-mu”, atau “- nya”), jika ada.\
3. Hapus Derivation Suffixes (“-i”, “-an” atau “-kan”). Jika kata ditemukan di kamus, maka metode berhenti. Jika tidak maka ke langkah 3a.
  - a. Jika “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “- k”, maka “-k” juga ikut

- dihapus. Jika kata tersebut ditemukan dalam kamus maka metode berhenti. Jika tidak ditemukan maka lakukan langkah 3b
- b. Akhiran yang dihapus (“-i”, “-an” atau “-kan”) dikembalikan, lanjut ke langkah 4.
4. Hapus Derivation Prefix. Jika pada langkah 3 ada sufiks yang dihapus maka pergi ke langkah 4a, jika tidak pergi ke langkah 4b.
    - a. Periksa tabel kombinasi awalan – akhiranyang tidak diijinkan. Jika ditemukan maka metode berhenti, jika tidak pergi ke langkah 4b
    - b. For  $i = 1$  to 3, tentukan tipe awalan kemudian hapus awalan. Jika root word belum juga ditemukan lakukan langkah 5, jika sudah maka metode berhenti.
  5. Melakukan recording
  6. Jika semua langkah telah selesai tetapi tidak juga berhasil maka kata awal diasumsikan sebagai *rootword*. Proses selesai.

## 5. *Convert Negation*

*Convert Negation* adalah proses yang bertujuan untuk menggabungkan kata negasi dengan kata setelahnya seperti contoh “ga suka” menjadi “gasuka”. Proses ini dilakukan karena dapat menentukan hasil deteksi teks.

## 6. **Stopword Removal**

*Stopword Removal* merupakan tahap menghilangkan kata yang tidak sesuai dengan topik dokumen dengan menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting).

## 2.5. Particle Swarm Optimization (PSO)

*Particle Swarm Optimization* (PSO) adalah metode optimasi yang dapat digunakan untuk menentukan parameter-parameter proses yang menghasilkan nilai respon yang optimum. PSO meniru sosial perilaku sekawanan burung atau ikan pada habitat alami. Setiap individu atau partikel berperilaku dengan cara menggunakan kecerdasannya (*intelligence*) sendiri dan juga dipengaruhi perilaku kelompok kolektifnya. Saat satu partikel atau seekor burung menemukan jalan tersebut meskipun lokasi mereka jauh di kelompok tersebut. Setiap individu atau partikel diperlakukan seperti sebuah titik pada suatu dimensi ruang tertentu. Terdapat dua faktor yang memberikan karakter terhadap status partikel pada ruang pencarian yaitu posisi partikel dan kecepatan partikel [10]. Berikut merupakan formulasi matematika yang menggambarkan posisi dan kecepatan partikel pada suatu dimensi ruang tertentu :

$$x_i(i) = x_1(1), x_2(1), \dots, x_{jN}(i) \quad (2.1)$$

$$v_j(i) = v_1(1), v_2(1), \dots, v_{jN}(i) \quad (2.2)$$

Dengan :

- x = posisi partikel
- v = kecepatan partikel
- i = iterasi ke-i
- j = indeks partikel
- N = jumlah partikel

Persamaan untuk mekanisme pembaruan status partikel adalah sebagai berikut (Kennedy dan Eberhart, 1995) :

$$v_j(i) = v_j(i - 1) + c_1 r_1 (P_{best,j} - x_j(i - 1)) + c_2 r_2 (G_{best} - x_j(i - 1)) \quad (2.3)$$
$$x_j(i) = v_j + x_j(i - 1)$$

Dengan :

$j = 1, 2, \dots, N$  merepresentasikan jumlah partikel

$Obj_i$  = respon dari eksperimen yaitu gaya tekan, torsi, delaminasi lubang masuk dan delaminasi lubang keluar.

$P_{best,j} = P_{best,1}, P_{best,2}, \dots, P_{best,2}$  merepresentasikan personal best dari partikel ke  $-j$

$G_{best}$  = merepresentasikan global best dari seluruh kawanan.

$c_1, c_2$  = learning factor.

$r_1, r_2$  = konstanta berupa bilangan random yang bernilai antara 0 sampai 1

## 2.6. Analisis Sentimen Berbasis Aspek

Analisis sentimen adalah analisis yang bertujuan untuk mengidentifikasi opini dan emosi yang diekspresikan oleh seseorang. ABSA melakukan analisis sentimen yang lebih dalam dari teks ulasan. Contohnya saat melihat ulasan dari sebuah lagu, opini yang dihasilkan tidak hanya keseluruhan sentimen tapi juga aspek spesifik seperti vokal, lirik, kualitas rekaman, dan lain-lain. Untuk mendapatkan beberapa aspek – aspek yang berbeda maka dibutuhkan aspect sentence labeling. Tujuan dari aspect sentence labeling yaitu melakukan pengidentifikasian dan menentukan kata yang memiliki kategori yang sama untuk nantinya dapat ditentukan aspek yang relevannya [11].

Ada beberapa langkah pada analisis sentimen berbasis aspek ini diantaranya yaitu :

### 1. *Clustering*

*Clustering* dilakukan untuk menentukan kata pada dataset yang memiliki kategori yang sama. Misalnya pada kata “panjang” dan “pendek” yang keduanya merupakan kategori ukuran. *Dataset* yang belum memiliki label kemudian akan dikelompokkan pada satu *cluster*.

## 2. Identifying Initial Document for Each Class

Pada tahap ini dilakukan identifikasi inisial dari setiap cluster atau yang disebut dengan aspek.

## 3. Labeling Document

*Labeling document* dilakukan pada kata yang sudah memiliki *cluster* dan diberi penilaian serta memutuskan aspek.

## 4. Klasifikasi Analisis Sentimen

Klasifikasi merupakan teknik yang mempelajari sekumpulan data sehingga dihasilkan aturan yang bisa mengklasifikasikan atau mengenali data-data baru yang belum pernah dipelajari. Klasifikasi dapat didefinisikan sebagai proses untuk menyatakan suatu objek data sebagai salah satu kategori (kelas) yang telah didefinisikan sebelumnya.

## 2.7. Teorema Bayes

Teorema bayes merupakan teorema yang dikenalkan oleh Thomas Bayes yang dimana untuk menghubungkan tingkat keyakinan (prior) dengan keyakinan baru (posterior) setelah mengetahui adanya observasi baru (evidence) berdasarkan likelihood tertentu dan dinyatakan dengan rumus [12] :

$$P(B) = \frac{P(A).P(A)}{P(B)} \quad (2.4)$$

Keterangan :

$P(A|B)$  : Probabilitas hipotesis berdasarkan kondisi (*posteriori probability*)

$P(A)$  : Probabilitas hipotesis (*prior probability*)

$P(B|A)$  : Probabilitas berdasarkan kondisi pada hipotesis

$P(B)$  : Probabilitas X

Bayes adalah teknik prediksi berbasis probabilistic sederhana yang berdasarkan pada penerapan teorema Bayes dengan asumsi independensi

(ketidaktergantungan) yang kuat (naif). Dalam Bayes, maksud independensi yang kuat pada fitur adalah bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama.

Ide dasar dari aturan Bayes adalah bahwa hasil dari hipotesis atau peristiwa (H) dapat diperkirakan berdasarkan pada beberapa bukti € yang diamati. Ada beberapa hal penting dari aturan Bayes tersebut, yaitu :

1. Sebuah probabilitas awal/prior A atau P(A) 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

## 2.8. *Naïve Bayes Classifier*

Algoritma *Naïve Bayes* merupakan salah satu bentuk pengklasifikasian dengan mempertimbangkan nilai probabilitas statistik. Sehingga, prediksi keanggotaan kelas suatu data baru dilakukan dengan memprediksi probabilitas keanggotaan dari data tersebut (Feldman, 2007). Metode *Naïve Bayes Classifier* adalah algoritma untuk klasifikasi yang sangat efektif untuk mendapatkan hasil yang tepat dan efisien karena proses penalaran yang dilakukan memanfaatkan input yang ada dengan cara yang relatif cepat. *Naïve Bayes Classifier* menggunakan pendekatan teorema *Bayes* untuk menghitung probabilitas kategori berdasarkan dokumen yang telah diketahui [13]. Perhitungan nilai probabilitas tersebut menggunakan persamaan :

$$P(H|X) = \frac{P(H).P(H)}{P(X)} \quad (2.5)$$

*Naïve bayes classifier* menggunakan asumsi yang sangat kuat (naif) yang akan independensi dari masing-masing kondisi atau kejadian, dimana masing-



masing petunjuk saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu persamaan sebagai berikut :

$$P(X) = P(H) \prod_{i=1}^n P(X_i|H) \quad (2.6)$$

Keterangan :

X : Data testing yang kelasnya belum diketahui

H : Hipotesis data X yang merupakan suatu kelas yang lebih spesifik.

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

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

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

P(X) : Probabilitas hipotesis X (*predictor prior probability*)

Algoritma *Naïve Bayes Classifier* sangat cocok untuk melakukan klasifikasi pada dataset bertipe nominal. Untuk dataset bertipe nominal, perhitungannya menggunakan persamaan (1). Apabila dataset bertipe numerik maka digunakan perhitungan distribusi Gaussian. Perhitungan distribusi Gaussian dapat dilihat dari persamaan (3), dimana dihitung terlebih dahulu nilai rata-rata  $\mu$  sesuai pada persamaan (4), dan standar deviasi  $\sigma$  sesuai pada persamaan (5) [referensi]. Tipe data nominal adalah jenis data yang diperoleh dengan cara kategorisasi atau klasifikasi, dan menunjukkan beberapa objek yang berbeda, contohnya : kode pos, jenis kelamin, nama kota, dll. Sedangkan tipe data numerik adalah jenis data yang diperoleh dengan cara pengukuran dimana jarak dua titik pada skala sudah diketahui, contohnya : umur, berat badan, tinggi badan, jumlah uang, dll.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.7)$$

$$\sigma = \sqrt{\frac{\sum_i^n n(x_i-\mu)^2}{n-1}} \quad (2.8)$$

Keterangan :

$F(x)$	: Nilai gaussian
$X$	: Nilai data
$\mu$	: Nilai rata-rata (mean)
$\sigma$	: Standar deviasi
$\pi$	: Nilai phi (3,146 atau 22/7)
$e$	: 2,7183
$x_i$	: Nilai data ke-i
$n$	: Jumlah data

Langkah-langkah *Algoritma Naïve Bayes Classifier* adalah sebagai berikut

:

1. Menyiapkan dataset
2. Hitung jumlah kelas pada data training
3. Hitung jumlah kasus yang sama dengan kelas yang sama. Jika data atribut bertipe nominal maka gunakan persamaan (1). Sedangkan jika data atribut bertipe numerik maka gunakan persamaan (3).
4. Kalikan semua hasil sesuai dengan data testing yang akan dicari kelasnya dengan menggunakan persamaan (2). Kemudian kalikan dengan hasil dari langkah kedua (jumlah kelas pada data training).
5. Bandingkan hasil per kelas, nilai tertinggi ditetapkan sebagai kelas baru.

## 2.9. Naïve Bayes Untuk Teks

Metode naïve bayes dapat digunakan dalam berbagai kebutuhan salah satunya yaitu untuk pengklasifikasian teks [14]. Berikut merupakan penjelasan dari Naïve Bayes untuk mengklasifikasikan teks :

$$P(\text{positif} | a)$$

$$P(\text{negatif} | a)$$

jika diterapkan pada rumus (2.1) maka akan menghasilkan persamaan berikut ini :

$$P(\text{positif} | a) = \frac{P(a | \text{positif}) P(\text{positif})}{P(a)}$$

$$P(\text{negatif} | a) = \frac{P(a | \text{negatif}) P(\text{negatif})}{P(a)}$$

karena denominator pada persamaan adalah sama, maka denominator bisa diabaikan dan yang dicari adalah nilai numerator.

$$P(\text{positif}) = \frac{\text{jumlah kalimat pada kelas positif}}{\text{jumlah total kalimat dalam set pelatihan}}$$

$$P(\text{negatif}) = \frac{\text{jumlah kalimat pada kelas negatif}}{\text{jumlah total kalimat dalam set pelatihan}}$$

jadi untuk ke tahapan numerator selanjutnya yaitu :

$$P(\text{positif}) = \frac{\text{berapa kali kalimat tersebut muncul di kelas itu}}{\text{jumlah total kalimat di set pelatihan}}$$

$$P(\text{negatif}) = \frac{\text{berapa kali kalimat tersebut muncul di kelas itu}}{\text{jumlah total kalimat dalam set pelatihan}}$$

Karena kalimat yang ingin diprediksi tidak muncul pada data training maka kalimat akan diubah menjadi kata, dan tiap kata akan diasumsikan saling independen atau dilihat masing-masing kata nya. Kalimat a terdiri dari n kata maka akan didapatkan sebagai berikut :

$$P(a | \text{positif}) = P(a_1 | \text{positif}) \times \dots \times P(a_n | \text{positif})$$

$$P(a | \text{negatif}) = P(a_1 | \text{negatif}) \times \dots \times P(a_n | \text{negatif})$$

tahap selanjutnya yaitu akan digunakan untuk mencari probabilitas kata yang akan muncul pada salah satu kelas, berikut persamaan untuk menghitung probabilitas :

$$P(w_k | c_i) = \frac{(n_{wk})c_i + 1}{n_{ci} + n_k} \quad (2.9)$$

Keterangan :

- a = kalimat yang akan diprediksi kelasnya
- $P(w_k | c_i)$  = peluang kemunculan  $w_k$  pada kategori  $c_i$
- $w_k$  = kata yang dicari peluang kemunculannya pada data latih
- $c_i$  = kategori yang terdapat pada dataset
- $(n_{wk}) c_i$  = jumlah kemunculan kata  $w_k$  pada kelas/kategori  $c_i$
- $n_{ci}$  = total kata yang ada pada kelas/kategori  $c_i$
- $n_k$  = total kata yang ada pada data latih

Pada saat tahap pengujian apabila terdapat kata yang belum muncul pada data latih maka nilai  $P(c_i) = 1$ .

## 2.10. Pembobotan Kata TF

TF (Term Frequency) adalah salah satu metode pembobotan kata untuk dokumen yang jumlahnya tunggal, tiap kalimat dianggap sebagai dokumen. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata tersebut di dalam dokumen itu. Bobot kata semakin besar jika kata tersebut sering muncul dalam suatu dokumen dan bobotnya akan semakin kecil jika kata tersebut muncul dalam banyak dokumen [15].

Konsep term frequency memandang suatu dokumen sebagai kantong kata dimana urutan dari kemunculan suatu kata diabaikan dan hanya jumlah kemunculan dari kata itu saja yang penting. Konsep term frequency memiliki kelemahan yaitu semua kata dianggap setara. Hal ini mengakibatkan relevansi suatu kata menjadi sangat tinggi jika kata itu sering muncul dalam suatu kumpulan dokumen. Terdapat 4 buah algoritma untuk mendapatkan nilai TF diantaranya yaitu :

- a. Raw TF yaitu nilai TF sebuah term yang dihitung berdasarkan kemunculan term tersebut dalam dokumen
  - b. Logarithmic TF. Merupakan cara memperoleh nilai Tf dengan menggunakan fungsi logaritma dalam matematika  $TF = 1 + \log(TF)$
  - c. Binary TF yaitu cara menghasilkan nilai Boolean berdasarkan kemunculan term pada dokumen tersebut. Dan akan bernilai 0 apabila term tersebut ada dalam dokumen sehingga banyaknya kemunculan term pada dokumen tidak berpengaruh.
  - d. Augmented Tf  $TF = 0.5 + 0.5 \times Tf_{max}(TF)$
- Dalam penelitian ini digunakan algoritma Raw TF. Raw TF diperoleh dari perhitungan frekuensi kemunculan suatu istilah pada dokumen.

**2.11. Confusion Matrix**

*Confusion Matrix* merupakan sebuah matriks yang berupa data-data aktual dan data prediksi yang telah diklasifikasikan oleh sistem. Dalam sebuah penelitian untuk mengetahui tingkat akurasi dari hasil penggunaan metode klasifikasi biasanya akan menghitung jumlah data uji yang kelas prediksinya dapat diprediksi dengan benar. Berbeda dengan fungsi dari *confusion matrix* dimana metode ini digunakan untuk mengetahui kinerja dari sistem yang akan dievaluasi.

**Tabel 2. 1** Tabel *confusion matrix*

		Hasil Prediksi	
		1	0
Kelas Sebenarnya	1	TP	FN
	0	FP	TN

Keterangan :

1. *True Positive* (TP) merupakan jumlah dokumen dari kelas 1 yang benar diklasifikasikan sebagai kelas 1.
2. *False Positive* (FP) merupakan jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1.
3. *False Negative* (FN) merupakan jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0.
4. *True Negative* (TN) merupakan jumlah kelas dari kelas 0 yang benar diklasifikasikan sebagai kelas 0.

Selanjutnya akan dihitung akurasi, akurasi merupakan perhitungan antara jumlah prediksi benar dan dibagikan dengan jumlah prediksi, untuk menghitung akurasi akan digunakan persamaan sebagai berikut :

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Setelah itu akan dilakukan perhitungan untuk *precision*, *precision* digunakan untuk menghitung rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif., untuk menghitung nilai *precision* digunakan persamaan sebagai berikut :

$$precision = \frac{TP}{TP + FP}$$

## 2.12. *Binary Classification*

*Binary classification* atau klasifikasi biner merupakan metode yang akan membantu permasalahan yang ada pada penelitian ini. Dimana pada penelitian ini

akan dilakukan sebuah analisis sentimen berbasis aspek yang akan mengenali aspek pada sentimen ulasan. Pada satu ulasan bisa terdapat lebih dari satu aspek, maka dari itu untuk mengenali aspek pada suatu ulasan dibutuhkan pendekatan multilabel. Salah satu metode yang dapat digunakan untuk melakukan klasifikasi multi label adalah dengan menggunakan *binary classification*. Pada *binary classification* dilakukan dengan mengambil satu label dari keseluruhan label, dan mengklasifikasikan masing-masing label secara terpisah. Terdapat dua pendekatan utama pada *binary classification* yaitu *one-versus-one* dan *one-versus-all*. Pendekatan OVO dan OVA ini merupakan metode yang sering digunakan pada saat pengklasifikasian multiclass. OVA yang dikenal juga dengan OVR atau one-versus-rest berdasarkan pelatihan pengklasifikasian individu untuk setiap kelas yang dibandingkan dengan kelas yang lain.

### 2.13. Data Flow Diagram (DFD)

DFD sering digunakan untuk menggambarkan suatu sistem yang telah ada atau sistem baru yang akan dikembangkan secara logika tanpa mempertimbangkan lingkungan fisik dimana data tersebut mengalir. DFD merupakan alat yang digunakan pada metodologi pengembangan sistem yang terstruktur dan dapat mengembangkan arus data di dalam sistem dengan terstruktur dan jelas.

Simbol-Simbol yang digunakan pada DFD :

1. *External Entity* atau *Boundary*

*External entity* atau kesatuan luar merupakan kesatuan di lingkungan luar sistem yang dapat berupa orang, organisasi atau sistem lainnya yang berada di lingkungan luarnya yang akan memberikan input atau menerima output dari sistem. *External entity* disimbolkan dengan notasi kotak.

2. Arus Data

Arus data (*data flow*) di DFD diberi simbol panah. Arus data ini mengalir diantara proses, simpanan data (*data store*) dan kesatuan

luar (*external entity*). Arus data ini menunjukkan arus data yang dapat berupa masukan untuk sistem atau hasil dari proses sistem.

3. Proses

Proses merupakan kegiatan yang dilakukan oleh orang, mesin, atau komputer dari hasil suatu arus data yang masuk ke dalam proses untuk menghasilkan arus data yang akan keluar dari proses. Simbol proses berupa lingkaran atau persegi panjang bersudut tumpul.

4. Simpanan Data

Simpanan data merupakan simpanan dari data yang dapat berupa hal-hal sebagai berikut, sebagai gambaran :

A. Suatu file atau database di sistem komputer.

B. Suatu arsip atau catatan manual

C. Suatu kotak tempat data di meja seseorang

D. Suatu tabel acuan manual

Simpanan data di DFD disimbolkan dengan sepasang garis horizontal paralel yang tertutup di salah satu ujungnya.

### **2.13.1. Diagram Konteks**

Menurut Jogiyanto (2005) Diagram konteks adalah diagram yang terdiri dari suatu proses dan menggambarkan ruang lingkup suatu sistem. Diagram konteks merupakan level tertinggi dari DFD yang menggambarkan seluruh input ke sistem atau output dari sistem. Ia akan memberi gambaran tentang keseluruhan sistem. Sistem dibatasi oleh boundary (dapat digambarkan dengan garis putus). Dalam diagram konteks hanya ada satu proses. Tidak boleh ada store dalam diagram konteks.

### **2.13.2. Data Flow Diagram Level 0**

DFD Level 0 adalah langkah selanjutnya setelah context diagram. Pada langkah ini, digambarkan proses – proses yang terjadi dalam sistem informasi.



### **2.13.3. Data Flow Diagram Level 1**

DFD Level 1 merupakan penjelasan dari DFD Level 0. Pada proses ini dijelaskan proses apa saja yang dilakukan pada setiap proses yang terdapat di DFD Level 0.