

BAB 2 TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Hosseini dan Gholaiman (2010) dan Pradono dkk (2011) melakukan pengelompokan konsumen dengan data yang bersifat *metric* dan menggunakan metode *K-Means*. Hosseini dan Gholaiman (2010) melakukan pengelompokan konsumen untuk mengembangkan *Customer Relationship Management* di sebuah perusahaan. Dalam penelitiannya, Hosseini dan Gholaiman (2010) mengkombinasikan metode *K-Means* dan RFM sehingga memberikan hasil pengelompokan yang lebih akurat. Sedangkan Pradono dkk (2011) menggunakan metode *K-Means* untuk mengelompokkan konsumen pengguna jasa kereta api pada jurusan tertentu dengan bantuan *software* SPSS. Dalam penelitian ini didapatkan bahwa segmentasi untuk layanan KRL ekspres dan ekonomi AC belum tepat sasaran.

Indrianti dan Satrio (2013) dan Rahayu (2012) juga melakukan penelitian segmentasi konsumen dengan menggunakan *K-Means*, hanya saja dalam kedua penelitian ini menggunakan data yang bersifat *non metric*. Indrianti dan Satrio (2013) memetakan persaingan antar produk kartu seluler, melakukan segmentasi konsumen dan mengetahui atribut yang menjadi prioritas bagi konsumen dan dari hasil pengolahan data didapatkan bahwa terbentuk 4 *cluster* dengan *cluster* 1 merupakan konsumen yang mengutamakan jangkauan sinyal, *cluster* 2 merupakan konsumen yang mengutamakan perangkat lunak promosi, *cluster* 3 merupakan konsumen yang mengutamakan kekuatan sinyal dan *cluster* 4 merupakan konsumen yang mengutamakan tarif. Untuk persaingan antar produk sendiri didominasi oleh kartu seluler Simpati dan IM3. Sedangkan Rahayu (2012) melakukan segmentasi UMKM di kota Malang agar dapat menyusun strategi pemasaran yang sesuai dengan segmen yang terbentuk. Dari hasil pengolahan data didapatkan 4 *cluster* yang terbentuk dengan *cluster* 1 merupakan pemimpin pasar, *cluster* 2 merupakan penceruk pasar, *cluster* 3 merupakan penantang pasar dan *cluster* 4 merupakan pengikut pasar. Dalam pengolahan datanya, kedua penelitian ini menggunakan *software* SPSS untuk melakukan analisis *cluster*.

Penelitian lain yang dilakukan adalah membandingkan beberapa metode analisis *cluster* yang ada. Alfina dkk (2012) dan Laraswati (2014) melakukan

perbandingan metode hirarki dan *K-Means* dalam melakukan analisis *cluster*. Alfina dkk (2012) membandingkan metode hirarki dan *K-Means* serta mengkombinasikan kedua metode tersebut dalam melakukan analisis *cluster* dengan jenis data yang *non metric*. Dari hasil perbandingan, didapatkan bahwa metode yang paling baik adalah kombinasi antara hirarki yakni *Average Linkage* dengan *K-Means*. Laraswati (2014) membandingkan kinerja antara *complete linkage*, *average linkage* dan *K-Means* dengan menggunakan data yang bersifat *metric*. Pada penelitian ini menggunakan *software* SPSS untuk melakukan pengolahan data. Dari hasil penelitian didapatkan bahwa metode *complete linkage* dan *average linkage* sama-sama memiliki nilai rasio terendah sehingga kedua metode ini merupakan metode yang paling baik dibandingkan dengan *K-Means*.

Tohari (2014) dan Ogbeide (2015) melakukan segmentasi konsumen dengan jenis data yang bersifat *non metric*. Tohari (2014) melakukan segmentasi konsumen pada Plaza Araya dengan menggunakan *average linkage*. Karena data yang digunakan *non metric* maka dalam penelitian ini, Tohari (2014) menggunakan *Minitab* untuk melakukan standarisasi data dan menggunakan SPSS untuk melakukan analisis *cluster*. Ogbeide (2015) menggunakan metode *ward* untuk melakukan analisis *cluster*. Dalam pengolahan data, Ogbeide (2015) menggunakan *software* yang jarang digunakan yakni *Strata 12 Software*.

Analisis *cluster* untuk segmentasi konsumen di sebuah retail juga dilakukan oleh Kaushal dan Prasad (2013). Kaushal dan Prasad (2013) melakukan segmentasi konsumen di sebuah pusat perbelanjaan agar dapat memberikan layanan yang lebih lagi. Sama dengan penelitian Tohari (2014), jenis data yang digunakan merupakan data *non metric*. Dalam penelitiannya, Kaushal dan Prasad (2013) menggunakan metode *K-Means* untuk melakukan analisis *cluster* dan dibantu dengan menggunakan *software* SPSS untuk melakukan pengolahan data. Untuk melakukan analisis *cluster*, Tohari (2014) mempertimbangkan komponen psikografis sebagai atributnya. Kaushal dan Prasad (2013) menggunakan demografi sebagai atribut dalam segmentasi konsumen.

Dari tinjauan pustaka diatas, dapat dilihat bahwa dalam segmentasi konsumen data yang sering digunakan adalah jenis data *non metric*. Dalam pengolahannya, penelitian yang telah ada menggunakan *software* SPSS untuk melakukan analisis *cluster*. Untuk dapat digunakan pada SPSS, tentunya data-data tersebut

haruslah diubah menjadi *metric*. Jika tetap dipaksakan menggunakan data *non metric* maka pengukuran jarak yang dilakukan tentunya tidak ada artinya karena data *non metric* tersebut hanyalah berupa simbol.

Oleh karena itu, penelitian yang sekarang ini dilakukan adalah membuat perangkat lunak yang sesuai untuk data *non metric*. Dalam pengukuran jarak digunakan *proximity matrix*s dan menggunakan metode analisis *cluster* BSAS. Perangkat lunak yang dibuat ini murni untuk jenis data yang bersifat *non metric*.

2.2. Dasar Teori

2.2.1. Analisis Multivariat

Teknik analisis multivariat merupakan teknik analisis dalam statistika yang sangat terkenal dikarenakan kemampuannya yang dapat dijadikan pertimbangan dalam pengambilan keputusan. Analisis multivariat merujuk pada semua teknik statistika yang menganalisis berbagai pengukuran secara simultan pada individu atau objek yang diteliti. Kebanyakan teknik multivariat merupakan penggabungan dari analisis *univariate* (analisis distribusi variabel tunggal) dan analisis *bivariate* (*cross-classification*, *correlation*, analisis varian, dan *simple regression* untuk menganalisis dua variabel). (Hair dkk, 2009)

Menurut Supranto (2004), analisis multivariat dibagi menjadi dua kelompok yaitu:

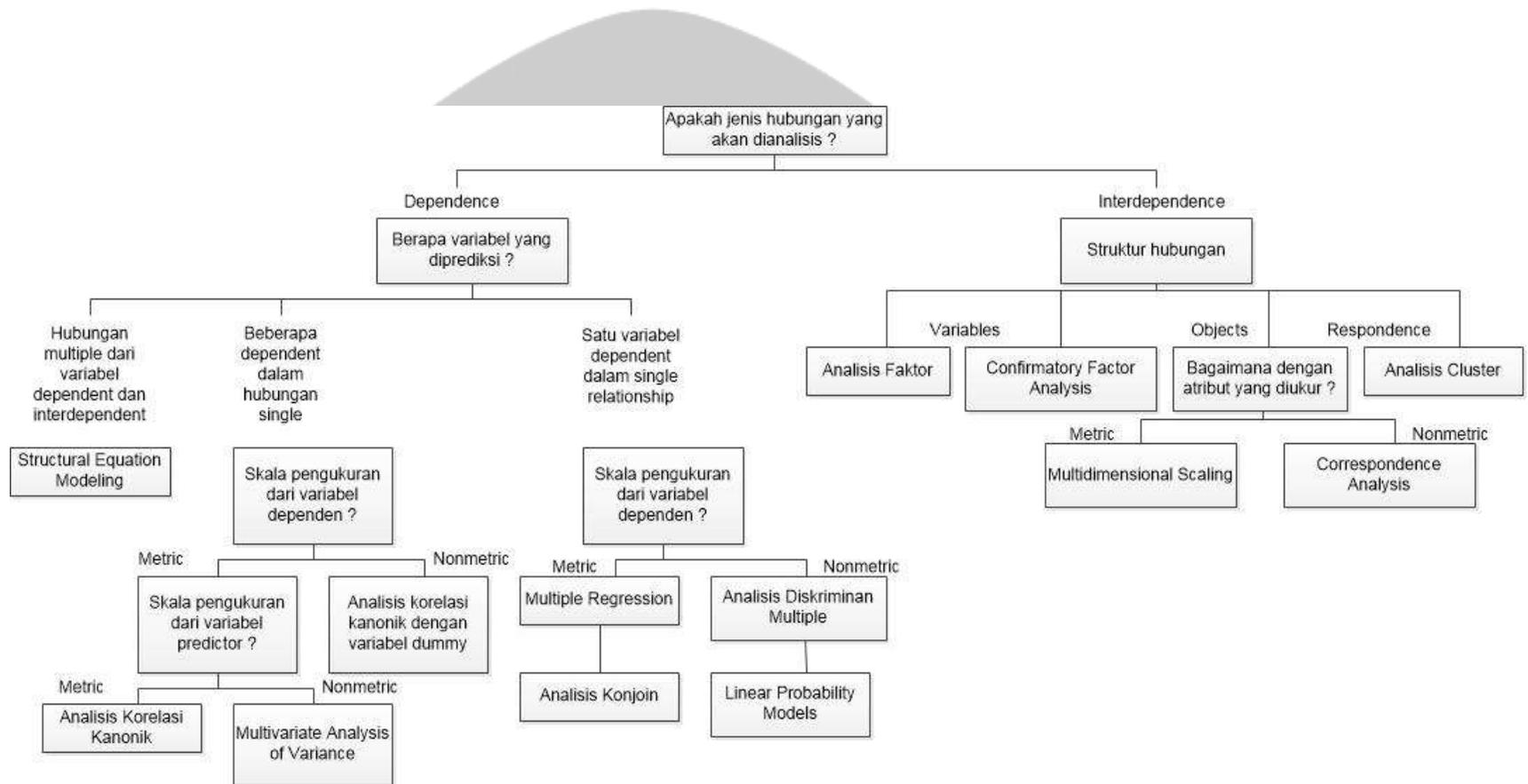
a. Analisis Dependensi / Ketergantungan

Analisis ini bertujuan untuk menjelaskan atau meramalkan nilai variabel tak bebas berdasarkan lebih dari satu variabel bebas yang mempengaruhinya. Hair dkk (2009) mengungkapkan bahwa analisis dependensi dapat memprediksi atau menjelaskan satu atau seperangkat variabel dari variabel lain yang diketahui sebagai independent variabel.

b. Analisis Interdependensi/ Saling Ketergantungan

Analisis ini bertujuan untuk mengelompokkan beberapa variabel menjadi kelompok yang lebih sedikit jumlahnya.

Hair dkk (2009) menggambarkan diagram untuk memilih analisis multivariat yang akan digunakan. Gambar 2.1. ini merupakan dasar dalam memilih analisis multivariat yang digunakan.



Gambar 2.1. Pemilihan Teknik Multivariat

2.2.2. Skala Pengukuran

Pengukuran merupakan hal yang penting dalam merepresentasikan keakuratan dari metode analisis multivariat yang dipilih. Data dapat diklasifikasikan menjadi dua kategori didasarkan pada atribut atau karakteristik yakni *non metric* (kualitatif) dan *metric* (kuantitatif). (Hair dkk, 2009)

Data *metric* diperoleh dengan mengukur dan bisa mempunyai desimal. Data *metric* dikategorikan menjadi data interval dan rasio. Sedangkan data *non metric* adalah data yang didapat dengan cara menghitung, tidak mempunyai desimal dan dapat dilakukan kategorisasi. Data *non metric* dibagi menjadi data nominal dan data ordinal. (Santoso, 2005).

2.2.3. Analisis Cluster

Menurut Hair dkk (2009) dalam buku *Multivariate Data Analysis*, analisis *cluster* merupakan kelompok teknik multivariat yang bertujuan untuk mengelompokkan objek ke dalam karakteristik yang dimiliki. Analisis ini juga dikenal sebagai *Q analysis*, *typology construction*, *classification analysis* dan *numerical taxonomy*. Berbagai istilah analisis *cluster* ini muncul karena pemanfaatannya dalam berbagai disiplin ilmu. Namun, secara umum metode-metode tersebut memiliki dimensi yang mirip yaitu klasifikasi yang didasarkan pada hubungan antar objek yang akan dikelompokkan.

Analisis *cluster* mengelompokkan objek sehingga setiap objek yang sama berada dalam satu *cluster* berdasarkan karakteristik yang dipilih. Hasil pengelompokan objek ini harus memiliki homogenitas yang tinggi di dalam anggota tiap *cluster* dan heterogenitas yang tinggi antar *cluster* yang terbentuk.

Mallapragada dkk (2010) mengungkapkan bahwa *clustering* bertujuan untuk mempartisi sejumlah n objek sebagai suatu poin dalam d ruang dimensi atau sebagai matriks persamaan $n \times n$. Pendekatan dalam analisis *cluster* terbagi menjadi dua yaitu pendekatan parametrik dan non parametrik. Pendekatan parametrik menekankan pada struktur data. Pendekatan ini telah banyak berhasil digunakan dalam berbagai aplikasi analisis data *discrete* dan *continue*. Akan tetapi, pendekatan parametrik hanya efektif bila digunakan untuk data yang diketahui distribusinya.

Sedangkan Chang Chien dkk (2012) mengatakan bahwa analisis *cluster* merupakan metode untuk mengelompokkan data berdasarkan persamaannya ke

dalam kelompok yang sama dan memiliki perbedaan yang besar antar *cluster*. Dari sudut pandang statistik, metode *clustering* secara umum dibagi menjadi dua kategori. Kategori pertama adalah pendekatan berdasarkan model probabilistik sedangkan kategori kedua adalah pendekatan *non parametric*. Pendekatan berdasarkan model probabilistik mengasumsikan bahwa data memiliki distribusi probabilistik. Pendekatan *non parametric* berdasarkan pengukuran persamaan atau perbedaan sehingga *clustering partitional* sangat baik digunakan.

2.2.4. Algoritma Clustering

Theodridis dkk (2003) mengungkapkan bahwa algoritma *clustering* merupakan sebuah skema yang dapat dipilih dengan mempertimbangkan hanya sebagian kecil dari himpunan data yang mungkin dipartisi. Hasil partisi ini tergantung dari algoritma yang spesifik dan kriteria yang digunakan. Algoritma *clustering* merupakan prosedur yang mencoba untuk mengidentifikasi karakteristik spesifik dari *cluster* berdasarkan data. Algoritma *clustering* terbagi menjadi beberapa jenis sebagai berikut.

a. Algoritma *Sequential*

Merupakan algoritma yang menghasilkan sebuah *cluster* single. Algoritma ini merupakan metode yang cukup mudah dan cepat. Kebanyakan dari metode ini melakukan iterasi hanya sekali atau beberapa kali (biasanya tidak lebih dari lima atau enam kali). Hasil akhir dari algoritma ini biasanya tergantung dari urutan vektor yang digunakan.

b. Algoritma Hirarki

Algoritma ini terbagi menjadi :

i. Algoritma *Agglomerative*

Algoritma ini menghasilkan urutan jumlah *cluster* yang menurun pada setiap langkah. Pengelompokan dari setiap langkah sebelumnya digabungkan menjadi satu.

ii. Algoritma *Divisive*

Algoritma ini merupakan kebalikan dari *agglomerative*. Algoritma ini menghasilkan *cluster* yang bertambah pada setiap langkahnya. Hasil pengelompokan dari langkah sebelumnya akan dipisah menjadi dua.

c. Algoritma *Clustering* berdasarkan *Cost Function Optimization*

Jenis ini merupakan algoritma yang “*sensible*” dengan menghitung fungsi biaya. Algoritma ini menggunakan konsep differential dalam kalkulus dan mencoba menghasilkan *cluster* yang optimum. Algoritma ini juga dikenal

dengan *iterative function optimization schemes*. Algoritma ini terdiri dari beberapa subkategori, yakni :

- a. *Hard* atau *crisp clustering algorithms*
- b. *Probabilistic clustering algorithms*
- c. *Fuzzy clustering algorithms*
- d. *Possibilistic clustering algorithms*
- e. *Boundary detection algorithms*

2.2.5. Basic Sequential Algorithmic Scheme

BSAS merupakan algoritma *clustering* yang paling dasar dan simpel. Dalam algoritma ini, yang dibutuhkan adalah pengukuran persamaan, batas dari persamaan dan jumlah *cluster* maksimal yang diizinkan. Algoritma BSAS dinyatakan sebagai berikut.

1. $m = 1; C_m = (x_1); // \text{Init first cluster} = \text{first sample}$
2. *for every sample x from 2 to N*
 - a. *find cluster C_k such that min d(x,C_k)*
 - b. *if (x,C_k) > th and (m < q)*
 $m = m + 1; C_m = (x) // \text{create a new cluster}$
 - c. *else*
 $C_k = C_k + (x) // \text{Add sample to the nearest cluster}$
Update representative if needed
3. *end algorithm.*

Ide utama dari algoritma ini adalah menetapkan vektor ke dalam *cluster* yang sudah ada atau membuat *cluster* baru berdasarkan jarak ke *cluster*. (Spagnolo dkk, 2007).

Theodridis dkk (2003) dalam algoritma ini tidak perlu diketahui jumlah *cluster* dalam kasus. *Cluster* baru akan terbentuk dengan sendirinya sesuai dengan evolusi dari algoritma. Hal terpenting yang mempengaruhi hasil dari algoritma ini adalah pemilihan dari *threshold* (batas). Jika nilai *threshold* terlalu kecil akan membentuk kelompok yang tidak diperlukan. Sedangkan jika nilai *threshold* terlalu besar akan membentuk jumlah *cluster* yang tidak sesuai.

2.2.6. Ukuran Jarak

Laraswati (2014) dalam skripsinya yang berjudul "Perbandingan Kinerja Metode *Complete Linkage*, Metode *Average Linkage* dan Metode *K-Means* dalam Menentukan Hasil Analisis *Cluster*" mengungkapkan sesuai dengan tujuan

analisis *cluster* yaitu untuk mengelompokkan obyek yang mirip dalam *cluster* yang sama maka beberapa ukuran diperlukan untuk mengetahui seberapa mirip atau berbeda objek-objek tersebut. Terdapat tiga metode yang dapat diterapkan dalam mengukur kesamaan antara objek, yaitu :

a. Ukuran Asosiasi

Ukuran asosiasi dipakai untuk mengukur data berskala *non metric* (nominal atau ordinal) dengan cara mengambil bentuk-bentuk koefisien korelasi pada tiap objeknya dengan memuatkan korelasi-korelasi yang bernilai negatif.

b. Ukuran Korelasi

Ukuran korelasi dapat diterapkan pada data dengan skala *metric*, namun ukuran korelasi jarang digunakan karena titik beratnya pada nilai suatu pola tertentu, padahal titik berat analisis *cluster* adalah besarnya objek. Kesamaan antar objek dapat dilihat dari koefisien korelasi antar pasangan objek yang diukur dengan beberapa variabel.

c. Ukuran Kedekatan

Metode ukuran jarak diterapkan pada data berskala *metric*. Ukuran ini sebenarnya merupakan ukuran ketidakmiripan, dimana jarak yang besar menunjukkan sedikit kesamaan sebaliknya jarak yang pendek/kecil menunjukkan bahwa suatu objek semakin mirip dengan objek lain. Perbedaan dengan ukuran korelasi adalah bahwa ukuran korelasi bisa saja tidak memiliki kesamaan nilai tetapi memiliki kesamaan pola, sedangkan ukuran jarak lebih memiliki kesamaan nilai meskipun memiliki pola yang berbeda.

2.2.7. Pengukuran Jarak Antara Vektor Biner

Barbara (2007) mengungkapkan terdapat beberapa pengukuran jarak antara vektor biner. Pengukuran ini dikenal sebagai *similarity coefficients* dan memiliki nilai antara 0 dan 1. Nilai 1 mengindikasikan bahwa kedua vektor tersebut benar-benar sama dan nilai 0 mengindikasikan bahwa kedua vektor sama sekali tidak sama. Perbandingan antara kedua vektor biner ini dimisalkan dalam p dan q dapat dilambangkan dalam variabel sebagai berikut:

M_{01} = jumlah dimana p adalah 0 dan q adalah 1

M_{10} = jumlah dimana p adalah 1 dan q adalah 0

M_{00} = jumlah dimana p dan q adalah 0

M_{11} = jumlah dimana p dan q adalah 1

Didasarkan pada variabel diatas maka rumus dari *Simple Matching Coefficient* adalah sebagai berikut :

$$SMC = \frac{M_{11} + M_{00}}{M_{01} + M_{10} + M_{11} + M_{00}} \quad 2.1$$

Pengukuran jarak lainnya yang biasa digunakan adalah *Jaccard Coefficient*. Adapun rumus dari *Jaccard Coefficient* adalah sebagai berikut :

$$J = \frac{M_{11}}{M_{01} + M_{10} + M_{11}} \quad 2.2$$

Secara konsep, SMC menyamakan dengan total jumlah data yang cocok, sedangkan *Jaccard Coefficient* hanya menganggap bahwa kedua vektor yang bernilai 1 lebih penting. Sangat penting untuk mengetahui pengukuran mana yang lebih cocok antara keduanya berdasarkan kondisi yang dihadapi.

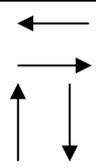
2.2.8. Definisi *Flowchart*

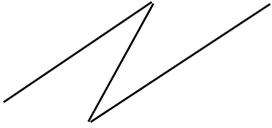
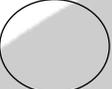
Al Bahra (2013) dalam penelitian Marono (2014) menjelaskan *flowchart* adalah bagan-bagan yang mempunyai arus yang menggambarkan langkah-langkah penyelesaian suatu masalah. Dalam *flowchart* terdapat simbol-simbol yang digunakan untuk menggambarkan langkah-langkah tersebut. Simbol-simbol *flowchart* dibagi menjadi tiga kelompok sebagai berikut :

- a. *Flow Direction Symbols* (Simbol penghubung/alur).
- b. *Processing Symbols* (Simbol proses).
- c. *Input-output Symbols* (Simbol input-output).

Dalam tabel 2.1., tabel 2.2. dan tabel 2.3. akan menunjukkan simbol-simbol dalam masing-masing kelompok tersebut.

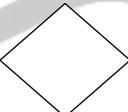
Tabel 2.1. *Flow Direction Symbols*

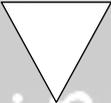
No	Simbol	Keterangan
1		<p>Simbol arus/<i>flow</i></p> <p>Untuk mengatur jalannya arus suatu proses</p>

2		Simbol <i>communication link</i> Untuk menyatakan bahwa adanya transisi suatu data/informasi dari satu lokasi ke lokasi lainnya
3		Simbol <i>connector</i> Untuk menyatakan sambungan dari satu proses ke proses lainnya dalam halaman/lembar yang sama
4		Simbol <i>Offline Connector</i> Untuk menyatakan sambungan dari satu proses ke proses lainnya dalam halaman/lembar yang berbeda

Sumber : Al-Bahra (2012)

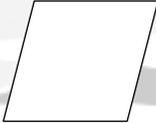
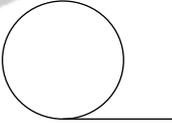
Tabel 2.2. Processing Symbols

No	Simbol	Keterangan
1		Simbol <i>Process</i> Untuk menyatakan pengolahan yang akan dilakukan oleh komputer
2		Simbol <i>Manual</i> Untuk menyatakan suatu tindakan (proses) yang tidak dilakukan oleh komputer (manual)
3		Simbol <i>Decision/Logika</i> Untuk menunjukkan suatu kondisi tertentu yang akan menghasilkan dua kemungkinan jawaban, ya/tidak
4		Simbol <i>Predefined Proses</i> Untuk menyatakan penyediaan tempat penyimpanan suatu pengolahan untuk memberi harga awal
5		Simbol <i>Terminal</i> Untuk menyatakan permulaan atau akhir suatu perangkat lunak

6		<p>Simbol <i>Keying Operation</i></p> <p>Untuk menyatakan segala jenis operasi yang diproses dengan menggunakan suatu mesin yang mempunyai <i>keyboard</i></p>
7		<p>Simbol <i>Off-line Storage</i></p> <p>Untuk menunjukkan bahwa data dalam simbol ini akan disimpan ke suatu media tertentu</p>
8		<p>Simbol <i>Manual Input</i></p> <p>Untuk memasukkan data secara manual dengan menggunakan <i>online keyboard</i></p>

Sumber : Al-Bahra (2012)

Tabel 2.3. Input-Output Symbols

No	Simbol	Keterangan
1		<p>Simbol <i>Input-Output</i></p> <p>Untuk menyatakan proses <i>input</i> dan <i>output</i> tanpa tergantung dengan jenis peralatannya</p>
2		<p>Simbol <i>Punched Card</i></p> <p>Untuk menyatakan <i>input</i> berasal dari kartu atau <i>output</i> ditulis ke kartu</p>
3		<p>Simbol <i>connector</i></p> <p>Untuk menyatakan <i>input</i> berasal dari pita <i>magnetic</i> atau <i>output</i> disimpan ke pita <i>magnetic</i></p>
4		<p>Simbol <i>Disk Storage</i></p> <p>Untuk menyatakan <i>input</i> berasal dari <i>disk</i> atau <i>output</i> disimpan ke <i>disk</i></p>
5		<p>Simbol <i>Document</i></p> <p>Untuk mencetak laporan ke printer</p>

6		<p>Simbol <i>Display</i></p> <p>Untuk menyatakan peralatan <i>output</i> yang digunakan berupa layar (video, komputer)</p>
---	---	--

Sumber : Al-Bahra (2012)

