

BAB 2

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Terkait dengan penelitian yang dilakukan penulis, maka penelitian terdahulu sangat penting sebagai dasar atau acuan yang dijadikan sebagai data pendukung. Berikut ini akan dijelaskan mengenai beberapa penelitian terdahulu yang berkaitan dengan penelitian yang dilakukan penulis :

Saragih (2013) yang mengambil tentang Analisis Keparahan Korban Kecelakaan Lalu Lintas di Surabaya Tahun 2012, Analisis Statistik Log Linear dan Logistik. Penelitian ini memiliki tujuan tahu akan pengaruh yang memfaktori tingkat fatalitas dari korban kecelakaan lalu lintas, tahu akan pengaruh karakteristik kecelakaan dan mengetahui sebesar apa dampak dari kecelakaan. Penelitian yang menggunakan variabel jenis kelamin, umur, pekerjaan, jenis tabrakan, peran pengemudi dan korban, waktu/hari.

Hanafi (2014) mengambil judul Analisis Regresi Logistik Ordinal dalam Identifikasi Faktor-faktor yang Berpengaruh terhadap Keparahan Korban Kecelakaan Lalu Lintas di Kabupaten Sleman. Dengan mengetahui dan identifikasi profil kecelakaan terlihat dampak mempengaruhi pada resiko fatalitas korban kecelakaan dengan kaitannya peluang resiko fatalitas berdasar faktor yang menjadi bukti memiliki pengaruh kecelakaan Tahun 2013 di Sleman.

Muhammad Syaeful Fajar (2015) dalam penelitiannya yang berjudul Analisis Kecelakaan Lalu Lintas Jalan Raya Di Kota Semarang Menggunakan Metode *K-Means Clustering*. Penelitian tersebut dilakukan untuk mengidentifikasi faktor-faktor yang berpengaruh terhadap kecelakaan, memprediksi waktu kecelakaan, memberikan alternatif penanganan yang tepat terkait kecelakaan yang terjadi di kota Semarang menggunakan metode algoritma *K-Means clustering*.

Qilbaaini Effendi M. (2015) dalam penelitiannya yang memilih Rancang Bangun Sistem Informasi Klasifikasi Status Gunung Berapi Menggunakan Metode *Naïve Bayes Classifier*. Penelitian tersebut dilakukan untuk membangun sebuah sistem informasi dengan pemfituran pengelolaan data status gunung berapi yang ada di Indonesia dan menghasilkan sesuatu informasi status dari gunung berapi tersebut. Metode algoritma *Naïve Bayes Classifier* disini penulis gunakan sebagai metode yang menjalankan pengolahan data numerik dan kategori dengan menerapkan keberlanjutan proses kemudian dalam menggunakan algoritma dengan proses pengolahan peluang yang dihitung setiap situasi dan kondisi gunung berapi dan pencarian nilai terbesar hasil dari perhitungan peluang diklasifikasikan hasilnya.

Muhammad Aidil (2019) dalam penelitiannya yang berjudul Analisis Daerah Rawan Kecelakaan Lalu Lintas Di Kabupaten Musi Rawas. Penelitian tersebut dilakukan untuk mengetahui faktor-faktor penyebab kecelakaan lalu lintas yang sering terjadi di Kabupaten Musi Rawas, mengetahui daerah rawan kecelakaan di Kabupaten Musi Rawas dan mengevaluasinya juga mengetahui karakteristik kecelakaan di Kabupaten Musi Rawas.

Dari beberapa penelitian diatas maka penulis melakukan penelitian serupa mengenai kecelakaan lalu lintas. Hal yang membedakan dengan penelitian sebelumnya adalah metode yang digunakan untuk menganalisa kasus kecelakaan. Yakni dengan menggunakan metode *Naïve Bayes Classifier* . Data sekunder yang digunakan dalam penelitian ini juga berbeda berdasarkan wilayah pengambilan data, yakni di Kota Ponorogo, Jawa Timur. Data didapat dari Polres Ponorogo untuk tahun 2020-2021.

2.2 Data Mining

Data mining merupakan suatu proses pengerukan atau pengumpulan informasi penting dari suatu data yang besar. Data mining adalah pengolahan dalam teknik statistika/matematika, *Artificial Intelligence*, pembelajaran mesin agar dapat mengidentifikasi dan mengekstrakti data informasi yang berguna dan bermanfaat terutama dalam pengetahuan yang berhubungan dari segala sumber database besar. (Turban, 2005).

Data mining berfungsi sebagai kemudahan pengguna dalam mendapat informasi yang dibutuhkan dan berguna untuk pengembangan pengguna. Berikutnya empat fungsi data mining yaitu:

a. Prediksi (*Prediction*)

Proses ini untuk memilih pola dari data sumber dengan memakai variable satu untuk memprediksi variable yang lain atau tidak bisa diidentifikasi jenis dan nilainya.

b. Deskripsi (*Description*)

Proses ini ditentukannya masing-masing karakteristik penting oleh suatu data dalam sumber data tertentu. Jika membentuk pola yang ulang dan muncul nilai maka karakteristik dari suatu data akan dapat diidentifikasi.

c. Klasifikasi (*Classification*)

Proses ini ditemukannya *function* atau model untuk memunculkan gambaran kelas dari sumber konsep data. Hasilnya untuk dideskripsikannya sumber data terpenting hingga dapat diramalkan arah karakteristik suatu data kedepan.

d. Asosiasi (*Association*)

Proses ini berguna untuk ditemukannya suatu hubungan diantara nilai atribut pada masing-masing kumpulan data.

Knowledge Discovery in Database (KDD) adalah beberapa kumpulan proses untuk menemukan pengetahuan yang bermanfaat dari data. Adapun kumpulan proses dalam KDD adalah seperti berikut :

- a. *Penyeleksian data (Data selection)*
Merupakan operasi dasar seperti penghapusan atribut yang tidak perlukan, membuang duplikasi data, dan juga memeriksa apakah data yang diambil konsisten atau tidak
- b. *Integrasi data (Data integration)*
Merupakan penggabungan beberapa data dari berbagai database ke dalam satu database baru.
- c. *Pemilihan data (Data selection)*
Merupakan proses penyeleksian data dari beberapa kumpulan data set dimana penyeleksian atau pemilihan data perlu dilakukan guna pemilihan dan penyeleksian sumber data berdasarkan relevansi terhadap analisis kemudian ditambahkan ke dalam kumpulan data utama
- d. *Transformasi data (Data transformation)*
Merupakan suatu proses dimana memasukkan data hasil proses ke dalam bentuk mining prosedur dengan memilih data dengan cara transformasi data dan gresi data.
- e. *Penambangan data (Data mining)*
Merupakan proses pencarian pola, juga merupakan proses yang paling penting untuk dilakukan berbagai teknik yang diaplikasikannya untuk mengekstrak berbagai pola-pola potensial untuk mendapatkan data yang sangat berguna.
- f. *Evaluasi pola (Pattern evaluation)*
Dalam tahap ini mencakup banyak aspek. Pada proses ini mencari pola-pola unik yang ditemukan pada tahap sebelumnya dengan mengidentifikasi berdasarkan pengukuran atau perhitungan
- g. *Presentasi pengetahuan (Knowledge presentation)*

Proses ini merupakan yang terakhir menampilkan visualisasi untuk digunakan pengguna dan menginterpretasi hasil tampilan dari data mining.

2.3 Naïve Bayes

Naïve Bayes adalah satu dari beberapa metode dengan inductive learning terefektif dan terefisien dalam pengolahan machine learning juga data mining. Kinerja produk dari *Naïve Bayes* menghasilkan pemrosesan dan pengklasifikasian dengan pengansumsian berdasarkan kemandirian setiap atribut. Pengansumsian berdasarkan kemandirian setiap atribut ini tidak sering berlaku pada data yang nyata, namun walaupun Pengansumsian kemandirian atribut berusaha di lalui dengan kinerja pengelompokan *Naïve Bayes* ini termasuk ke tingkat yang tinggi, pembuktian mengenai tingkat ini dari berberapa jurnal asli.

Klasifikasi *Naïve Bayes* digambarkan ketika ada maupun tidak kriteria pada suatu data yang ditentukan baik di dalam kelas maupun yang tidak ada hubungan dengan kriteria dengan lain.(Norfriansyah, 2016). Pemetodean *Naïve Bayes Classifier* ini menarik kesimpulan dengan dasar teorema klasifikasi Bayesian dalam menggambar masing-masing data atribut atau kriteria dijadikan dalam menentukan ketentuan keputusan bersifat tidak terikat (*independence*) dengan begitu meskipun ada maupun tidak ada data atribut atau kriteria tidak berpengaruh dan berhubungan pada adanya data atribut yang lain. Algoritma *Naïve Bayes* berjalan pada dua model data untuk alur prediksinya yaitu data set dan data tes. Data set adalah dasar dalam data tes yang digunakan menemukan peluang dikemudiannya. Namun data tes dalam pengujian data, data muncul terbentuk diawalnya sebelum peluang terprediksi.

Menurut Effrida dan Fricles satu dari keuntungan menggunakan *Naïve Bayes* dilakukan penelitian pemetodean ini dibutuhkannya data training minimal yang diolah terbentuk perkiraan kriteria dan data atribut yang dibutuhkan ketika

menjalankan memprosesan kelompok dan pemetodean ini baik didalam banyaknya kondisi masalah kehidupan nyata. Teorema *Naïve Bayes* dirumuskan sebagai berikut (Bustami, 2014) :

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

Keterangan :

A : Data dengan class yang belum diketahui

B : Hipotesis data merupakan suatu class spesifik

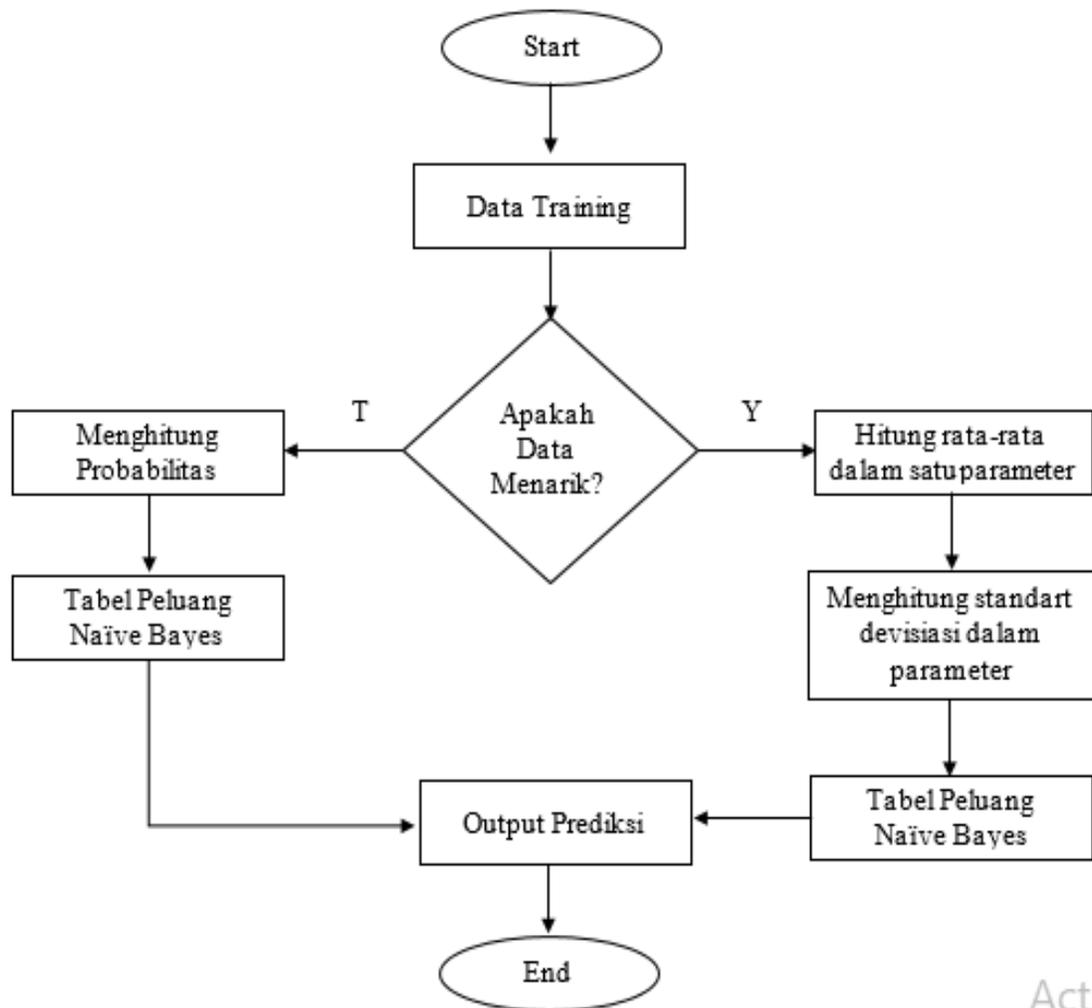
$P(B|A)$: Probabilitas dari hipotesis B berdasar kondisi A (Posterior Probabilitas)

$P(B)$: Probabilitas dari hipotesis B (Prior Probabilitas)

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

$P(A)$: Probabilitas A

Data yang dapat diolah di metode *Naïve Bayes Classifier* ada dua macam yakni data yang berjenis kategori dan datayang berjenis numeric. Didalam tahapan data perhitungannya dapat dilihat pada gambar 2.1 (Effendi, 2015).



Gambar 2.1 Tahapan Algoritma *Naïve Bayes*

Keterangan dari flowchart diatas adalah sebagai berikut :

- a. Program di mulai
- b. membaca data latih atau training
- c. Menghitung dan probabilitas, jika datanya numerik maka :
 - 1) Menghitung nilai mean atau rata-rata dan standart deviasi dari masing-masing parameter yang datanya berupa numerik.

- 2) Menghitung nilai probabilitik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut
- d. Mendapatkan nilai dalam tabel mean, standart deviasi dan pronanilitas
 - e. Hasilnya berupa prediksi
 - f. Program selesai

Adapun kelebihan dan kekurangan dari *Naïve Bayes* adalah sebagai berikut

:

- a. Kelebihan
 - 1) Hanya memerlukan pengkodean sederhana
 - 2) Menangani kuantitatif dan diskrit
 - 3) Lebih cepat dalam perhitungan
 - 4) Kokoh untuk titik *noise* yang diisolasi
 - 5) Hanya membutuhkan sejumlah kecil data training (latih) untuk memperkirakan parameter yang diperlukan untuk klarifikasi
 - 6) Menangani nilai yang hilang dengan mengabaikan instansi selama perhitungan estimasi peluang
- b. Kekurangan
 - 1) Banyak ilmuwan yang mengkritisi teorema ini karena pada teori ini satu probabilitas saja tidak bias mengukur seberapa dalam tingkat keakuratannya
 - 2) Tidak berlaku jika probabilitas kodisionalnya nol, apabila nol maka probabilitas prediksi akan bernilai nol juga
 - 3) Mengasumsikan variable bebas

2.4 R Studio

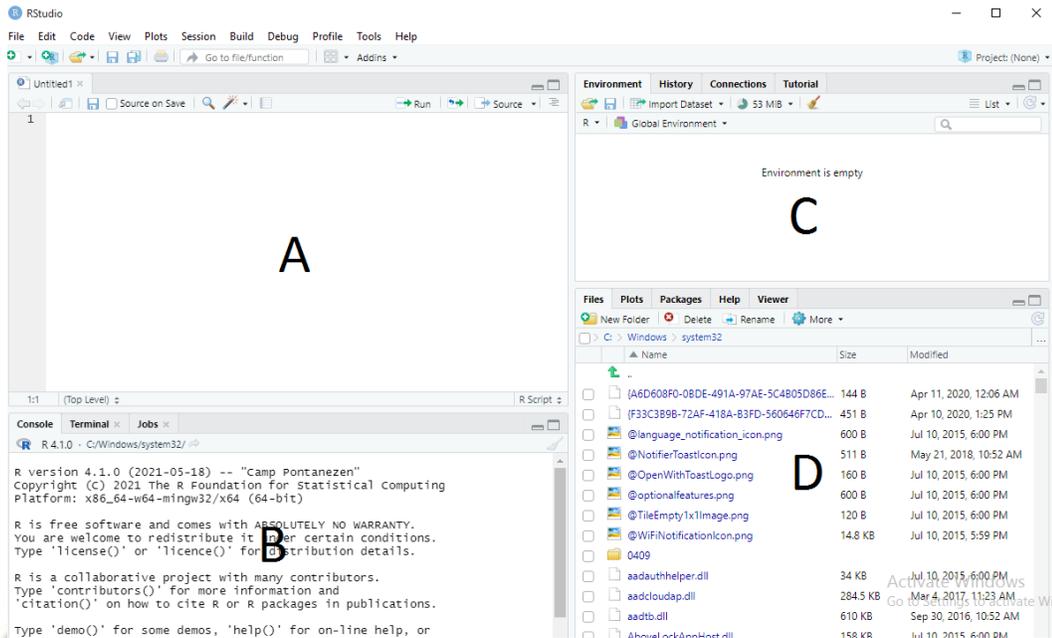
R Studio adalah bagian dari IDE (*Integrated Development Environment*) bersifat tidak umum dalam bahasa pemrograman R. Beberapa opsi dalam menyunting Bahasa R bertujuan sebagai kemudahan dalam menggunakan R, namun pada umumnya banyak penggunaan oleh user yaitu R Studio (Budiaji, 2019). Beberapa pemfituran dari kemampuan R Studio antara lain *R Console*, *code sompletion*, *syntax highlighting*.



Gambar 2.2 Logo RStudio

(Sumber:<https://www.rstudio.com/>)

R Studio umumnya oleh user dimanfaatkan yang berhubungan dengan akademik pembelajaran dan perindustrian.. R Studio pada Tahun 2013 telah memiliki tingkat urutan pertama memenangi hasil pemungutan suara KD Nuggets 2013 dengan mencantumkan salah satu “Top Language for Analytics”, Data mining, dan data ilmiah. Hal yang ditampilkan oleh monitor layar R Studio dibagi menjadi 4 bagian, yaitu sebagai berikut:



Gambar 2.3 Tampilan Awal Jendela RStudio

Keterangan :

A : *Code Script*

Adalah dimana user menulis code R, menyunting, dan melakukan penyimpanan code. Menjalankan code script bisa diproses RUN menunjukkan hasil proses yang akan ditampilkan di layar Console.

B : *Console*

Adalah bagian fitur R yang user lakukan lihat hasil keluaran langsung dan error message atas code sript user tuliskan dan dilakukan proses penjalanan dari *Code Script*.

C : *Global Environment*

Adalah bagian user melihat layar dengan tampilan pengertian masing-masing variabel dan data atribut melalui proses penyimpanan di bagian *Random Access Memory*. User dapat mengolah tampilan dengan

memilih salah satu variabel dengan nama yang diinginkan kemudian ditampilkan datanya.

D : Menu File

Yaitu menu yang digunakan untuk memanggil file yang sudah ada atau untuk membuat menu file yang baru.

2.5 Bahasa R

R adalah satu dari banyak jenis pengolahan perangkat lunak tidak berbayar yang dimanfaatkan perhitungan secara statistika maupun gambar (Faisal, 2016). Bahasa R ini digunakan untuk membuat program untuk awal mulanya diumumkan berkembang Tahun 1993 hasil dari ahli statistika Ross Ihaka bersama rekannya Robert Gentleman yang bertempat di New Zealand dengan menempuh pendidikan di Auckland University (Putra, 2018). Perangkat lunak R digunakan untuk membuat program maupun system lunak perhitungan yang hasilnya dimunculkan tanpa syarat yang kaitannya dengan perhitungan statistik dan penggambaran statistik yang akhirnya dilanjutkan pengembangan di R Core Team.



Gambar 2.4 Logo Bahasa R

(Sumber :<https://www.r-project.org/>)

2.6 Kecelakaan Lalu Lintas

2.6.1 Pengertian Kecelakaan Lalu Lintas

Kecelakaan lalu lintas adalah suatu peristiwa di Jalan yang tidak disangka dan tidak disengaja melibatkan kendaraan dengan atau tanpa pemakai Jalan lainnya, mengakibatkan korban manusia atau kerugian harta benda (PP RI No 43 Tahun 1993).

Definisi lain tentang kecelakaan lalu lintas adalah kejadian di mana sebuah kendaraan bermotor tabrakan dengan benda lain dan menyebabkan kerusakan. Kadang kecelakaan ini dapat mengakibatkan luka-luka atau kematian manusia atau binatang (WHO, 2004).

2.6.2 Bentuk-Bentuk Kecelakaan Lalu Lintas

Berdasarkan Undang-Undang Republik Indonesia Nomor 22 Tahun 2009 tentang Lalu Lintas dan Angkutan Jalan. Jenis kecelakaan lalu lintas dibagi menjadi tiga bagian, yaitu:

1. Kecelakaan lalu lintas ringan, yaitu merupakan kecelakaan yang mengakibatkan kerusakan kendaraan dan/atau barang.
2. Kecelakaan lalu lintas sedang, yaitu merupakan kecelakaan yang mengakibatkan luka ringan dan kerusakan kendaraan dan/atau barang.
3. Kecelakaan lalu lintas berat, yaitu merupakan kecelakaan yang mengakibatkan korban meninggal dunia atau luka berat.

2.6.3 Faktor Penyebab Terjadinya Kecelakaan Lalu Lintas

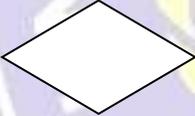
Kejadian Kecelakaan lalu lintas berhubungan dengan 3 faktor, Yaitu manusia, kendaraan, dan lingkungan fisik. Kejadian kecelakaan yang berhubungan dengan faktor manusia sebagian besar disebabkan oleh

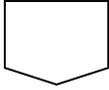
pengendara yang lengah. Faktor yang berhubungan dengan kecelakaan terbanyak kedua adalah faktor lingkungan fisik. Kecelakaan yang berhubungan dengan faktor lingkungan fisik sebagian besar disebabkan akibat jalan menikung. Faktor yang berhubungan dengan kecelakaan ketiga yaitu faktor kendaraan. Kecelakaan yang berhubungan dengan faktor kendaraan sebagian besar disebabkan oleh ban mengalami selip. Jika ketiga faktor tersebut dijumlahkan, maka jumlah persentasenya akan melebihi 100%, hal ini karena sebuah kecelakaan dapat disebabkan oleh lebih dari satu faktor penyebab (*multiple causes*) dan merupakan interaksi dari ketiga faktor tersebut yaitu manusia, kendaraan, dan lingkungan fisik. (Marsaid, 2018).

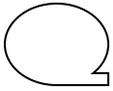
2.7 Flowchart

Flowchart yang sering dinamakan bagan mengalir merupakan kumpulan dari bentuk symbol tertentu berupa bangun datar yang memiliki aliran atau alur berdasarkan suatu pemikiran yang runtut di dalam perancangan segala jenis program. Kemudian Murdock dan Hernan (2018) memberikan landasan teori bahwa, Flowchart satu dari banyaknya proses dan runtutan pemikiran dengan pembuatan yang mengacu pada logis urutan sistematis dari kondisi pemrosesan ataupun suatu gambaran dalam bangun datar terhadap step to step akan runtutan peristiwa atau proses yang akan dilakukan di dalam program, user yang merupakan programmer maupun ahli analisa memanfaatkan flowchart dalam pemecahan masalah yang dibagi menjadi bagian-bagian kecil. Gambaran yang menjadi symbol sering digunakan di dalam flowchart antara lain:

Tabel 2.1 Simbol Flowchart

Simbol	Nama	Fungsi
	Proses / <i>Processing</i>	Digunakan untuk mempresentasikan sebuah operasi
	Input / Output	Digunakan untuk mempresentasikan input dan output data
	Terminal	Sebagai awal (<i>Start</i>) dan akhir (<i>End</i>) dari flowchart
	<i>Decision</i> (kotak keputusan)	Digunakan untuk mengambil sebuah keputusan yang sesuai dengan kondisi yang dipenuhi yaitu benar/salah
	<i>Preparation</i>	Untuk pemberian harga awal
	<i>Connector</i> / Penghubung	Digunakan untuk menghubungkan proses dalam flowchart yang masih berada pada halaman yang sama

	<p><i>On Page Connector</i></p>	<p>Untuk menggabungkan halaman kerja dari bagian flowchart yang berada pada halaman berbeda</p>
	<p><i>Flowline</i></p>	<p>Menunjukkan alur kerja</p>
	<p><i>Predefined Process</i> (Sub Program)</p>	<p>Digunakan untuk menjalankan sub bagian atau program</p>
	<p>Simbol Dokumen</p>	<p>Menatakan input berasal dari dokumen dalam bentuk kertas atau output dicetak ke kertas</p>
	<p><i>On Line Storage</i></p>	<p>Input atau output yang menggunakan penyimpanan akses langsung</p>
	<p>Manual Input</p>	<p>Simbol untuk pemasukan data secara manual</p>
	<p><i>Display</i></p>	<p>Digunakan untuk menampilkan output</p>
	<p><i>Manual Operation</i></p>	<p>Digunakan untuk menampilkan operasi yang tidak dilakukan oleh komputer (manual)</p>

	<i>Punched Card</i>	Input atau output yang menggunakan kartu
	<i>Magnetic Tape</i>	Input atau output yang menggunakan pita magnetik

