

BAB II
TINJAUAN PUSTAKA

Tabel 2.1 Penelitian Terdahulu

No	Judul	Peneliti (Tahun)	Metode	Hasil Penelitian
1	Penerapan Algoritma K-Means Untuk Analisis Prestasi Akademik Mahasiswa Fakultas Ilmu Komputer Universitas Klabat	Green F Mandias, Green A Sandag, Susi, Haryanto (2017)	K-Means Clustering	Menghasilkan output 3 cluster yang berisi tentang berapa persen mahasiswa dengan nilai tinggi, sedang dan rendah
2	Penerapan data mining dalam Meningkatkan mutu pembelajaran pada Instansi perguruan tinggi menggunakan Metode k-means clustering (studi kasus di Program studi tkj akademi komunitas Solok selatan)	Koko Handoko (2016)	K-Means Clustering	Dapat menghasilkan data mutu pembelajaran yang dilihat dari cluster
3	Penerapan Data Mining Dalam Pengelompokan Data Nilai Siswa Untuk Penentuan Jurusan Siswa Pada SMA Tamora Menggunakan Algoritma K-Means Clustering	Yohanni Sahra (2018)	K-Means Clustering	dapat digunakan dalam pengelompokan data nilai siswa untuk penentuan jurusan pada SMA Tamora
4	Sistem pembagian kelas kuliah mahasiswa dengan metode k-means dan k-nearest neighbors	Gede Aditra Pradyana, Agus Aan Jiwa Permana (2018)	k-means dan k-nearest neighbors	Dari metode tersebut kemudian dikembangkan sebuah sistem berbasis web dengan menggunakan bahasa

untuk meningkatkan
kualitas pembelajaran

pemrograman
PHP dengan
framework
Laravel

Berdasarkan penelitian diatas, maka dapat ditarik kesimpulan bahwa penelitian yang saya buat dengan penelitian terdahulu ada perbedaan. peneliti terdahulu dalam pengolahan data masih manual, sedangkan penelitian yang saya buat menggunakan software Rstudio dengan antarmuka supaya lebih mudah dalam penerapan data mining.

2.1 Mutu Pembelajaran

2.1.1 Pengertian Mutu Pembelajaran

Mutu Pembelajaran adalah ukuran atau tingkat dari proses belajar mengajar yang melalui berbagai tahap seperti proses kegiatan mengajar atau proses pembelajaran, melakukan penilaian dari hasil pembelajaran tersebut serta monitoring selama proses pembelajaran, salah satu masalah dalam proses pembelajaran yang umum adalah nilai atau hasil belajar yang didapatkan siswa di sekolah yang kadang rendah – tinggi karena berbagai faktor seperti kemampuan masing – masing siswa yang berbeda, pengaruh dari teman yang memunculkan rasa semangat dan malas, serta cara mengajar dari guru tersebut [11].

Ciri dari pembelajaran adalah adanya interaksi yang berada di dalam lingkungan sekolah maupun diluar lingkungan sekolah seperti bimbingan belajar, baik dengan guru, media digital.

Dalam hal ini,pokok dalam konsep mutu pembelajaran terdiri dari beberapa inti antara lain:

1. Persamaan yang terdiri dari beberapa indikator seperti sesuai dengan karakter siswa di sekolah, keinginan masyarakat, cocok

dengan tempat tinggal di sekitar, mengikuti perkembangan di dunia pendidikan dan atau inovasi baru.

2. Pembelajaran harus memiliki tujuan mendasar seperti kegiatan belajar yang didapatkan oleh semua kalangan, materi bisa dipahami karena disesuaikan dengan usia dengan tingkat dan kesulitan.
3. Pencapaian kualitas didapatkan dari tercapainya tujuan, serta bisa disebut dengan kecocokan dalam penyesuaian.
4. Ketepatan dalam pembelajaran bisa disebut ketepatan antara waktu, biaya, dan tenaga yang digunakan dengan hasil yang diperoleh atau dapat dikatakan sebagai dengan mengerjakan sesuatu dengan benar.[12]

Kemudian penilaian dilakukan oleh guru serta dijadikan satu yang menghasilkan kesimpulan untuk mengukur tingkat pencapaian kompetensi peserta didik selama beberapa waktu atau selama disekolah, sehingga digunakan untuk membuat laporan kemajuan hasil belajar atau raport. prosesnya dilaksanakan dengan tetap, sistematis, serta terlaksana sesuai alur dengan ujian, penilaian sikap dari korseling, penilaian tugas harian dan kenaikan kelas.[13]

2.2 Prediksi

1.2.1 Pengertian Prediksi

Prediksi atau peramalan adalah suatu proses memperkirakan dengan menggunakan proses dan alat tertentu yang pada dasarnya membawa pengamat tentang sesuatu yang bisa terjadi di masa mendatang dari informasi terdahulu, bisa beberapa bulan sebelumnya, tahun bahkan informasi yang dimiliki saat ini untuk kedepan, prediksi berfungsi untuk mengukur dan mengetahui selisih jarak yang dimiliki antara sesuatu yang mungkin akan terjadi dimasa yang akan datang dengan hasil yang diperkirakan karena didalam prediksi tidak harus memberi hasil informasi

yang pasti dan kejadian yang akan terjadi, tetapi untuk mencari jawaban sedekat mungkin yang akan terjadi.

Peramalan membutuhkan data terdahulu yang tidak terbatas waktu bisa yang belum lama atau yang sudah lama serta data sekarang atau data yang dimiliki saat ini, diantara data tersebut digunakan sesuai kebutuhan dan tujuan kemudian diproses maka menghasilkan perkiraan yang hasilnya menjadi pertimbangan kedepan untuk mengetahui kemungkinan yang terjadi beberapa waktu kedepan.[14]

1.2.2 Jenis Jenis Perkiraan

A. Perkiraan Berdasarkan Waktu

Berdasarkan waktu, prediksi atau dibagi menjadi tiga jenis yaitu :

1. Perkiraan jangka panjang

Meliputi waktu melebihi 1,5 tahun. seperti memperkiraan hal-hal yang diperlukan terkait perusahaan, perencanaan modal dan alat-alat, dan perencanaan untuk kegiatan riset.

2. Perkiraan jangka menengah

Meliputi waktu antara 3 Bulan-1,5 Tahun. seperti memperkirakan bahan baku yang dibutuhkan, merencanakan pembuatan produk, dan memperkirakan jumlah karyawan yang akan direkrut.

3. Perkiraan jangka pendek

Meliputi jangka waktu kurang dari 3 Bulan. seperti memperkirakan perencanaan pembelian perabotan, dan perencanaan pembagian shift.[15]

B. Perkiraan dalam fungsi dan rencana operasi masa depan

Dalam fungsi dan rencana operasi di masa depan, perkiraan dibagi menjadi tiga jenis yaitu:

1. Perkiraan ekonomi

Memperhatikan dalam perputaran bisnis dengan perkiraan kenaikan dalam jangka waktu tertentu, kondisi keuangan, material yang

dibutuhkan untuk membangun tempat tinggal dan hal-hal perencanaan yang lain.

2. Perkiraan teknologi

Memperhatikan perkembangan teknologi dalam beberapa waktu tertentu yang dapat meluncurkan produk terbaru dan menarik perhatian masyarakat.

3. Perkiraan permintaan

Memenuhi permintaan pasar bagi konsumen. Perkiraan ini disebut perkiraan penjualan dalam mengatur dalam produksi perusahaan supaya seimbang antara keduanya serta mengurangi resiko kehabisan stock saat permintaan banyak[15]

2.3 Analisis

2.3.1 Pengertian Analisis

Analisis adalah suatu kegiatan yang terdiri dari beberapa aktifitas yang memiliki sebuah tujuan seperti, mencari, membedakan obyek satu dengan yang lain, mengelompokkan obyek atau yang terkait dengan penelitian menurut syarat kesamaan yang sudah ditentukan dan kemudian dicari hubungannya beserta hasil yang digunakan sebagai pemecahan masalah.[16]

Pengertian analisis bisa diartikan sebagai kegiatan untuk mengamati sesuatu secara mendalam dan detail serta mempelajari dari berbagai sumber penelitian yang lalu diterapkan dengan cara menyesuaikan komponen yang terpisah serta untuk dipahami lebih lanjut, karena kemampuan dalam memecahkan masalah dan juga mengelompokkan sebuah informasi yang bermanfaat menjadi komponen-komponen bisa dipahami dan mudah dijelaskan [16]

2.4 Data Mining

2.4.1 Pengertian Data Mining

Data Mining adalah proses penggalian data untuk memisahkan serta menganalisa informasi yang memiliki tujuan tertentu, data mining merupakan suatu alur untuk memahami sekumpulan data yang diamati berupa pengetahuan, untuk mengetahui karakteristik data yang dimiliki. Beberapa istilah metode yang digunakan dalam literatur data Mining antara lain : clustering, klasifikasi, kaidah asosiasi, jaringan syaraf tiruan, dan algoritma genetika.[17]

Data mining merupakan salah satu cara yang digunakan untuk mendapatkan informasi dengan menganalisa jumlah data yang masih bersifat mentah. Berbagai cara telah dikembangkan dan diterapkan untuk menambah pemahaman pengetahuan dan informasi untuk menghasilkan sebuah cara yang berguna untuk mengambil sebuah keputusan.[18]

Data mining merupakan salah satu disiplin ilmu yang digunakan untuk menetapkan sebuah informasi yang dibutuhkan dari sumber data sebagai jembatan untuk mengambil keputusan.[18]

2.5 Tahap – Tahap Data Mining

1. Pembersihan Data (Data Cleaning)

Tahap ini merupakan proses penghapusan atau menghilangkan data yang tidak sesuai dengan atribut dan tidak ada kaitannya dalam sebuah dataset. data yang dimaksud adalah data yang tidak sesuai baik fungsinya maupun salah dalam penulisan karena keberadaanya bisa mengurangi ketelitian dan tujuan dari hasil data mining tersebut, hal ini harus dilakukan karena proses data mining harus lebih akurat karena jumlah data yang akan diproses lebih banyak.[19]

2. Integrasi Data

Integrasi data merupakan hasil penggabungan dua atau lebih dari beberapa data yang berasal dari beberapa database dan dimasukkan ke

dalam database yang baru. Proses ini perlu dilakukan pada atribut-atribut yang menunjukkan entitas seperti nama siswa, nomor induk, dan nilai. Integrasi data sangat diperlukan dalam data mining karena kesalahan pada integrasi data dapat menghasilkan hasil yang tidak sesuai dengan tujuannya.

3. Transformasi Data

Transformasi data merupakan tahap pengubahan format atau konversi data ke dalam format tertentu, dalam proses data mining pengubahan format data sangat penting sebelum proses data mining.

4. Seleksi Data

Data selection atau seleksi data adalah tahap pemilihan data yang akan digunakan karena tidak semua data pada database akan digunakan dalam penelitian. Hanya data yang sesuai dengan kriteria analisis yang akan diambil dari database untuk diteliti.

5. Proses Mining

Merupakan tahapan proses yang utama. Semua tahapan sebelumnya adalah untuk mendukung tahapan proses ini. Tahapan ini adalah proses menggunakan metode dan algoritma yang ada untuk menemukan pengetahuan dari data yang ada.

6. Evaluasi Pola yang ditemukan

Merupakan tahapan dalam mengidentifikasi pola – pola atau hasil yang kemudian dievaluasi berdasarkan hasil dari proses data mining yang telah dilakukan dan memberikan kesimpulan untuk menilai hasilnya.

7. Presentasi Pengetahuan

Presentasi pengetahuan adalah tahapan visualisasi atau penyajian hasil pengetahuan yang telah didapat ke dalam deskripsi agar dipahami oleh semua orang.[19]

2.6 Teknik Data Mining

2.6.1 Klasifikasi

Teknik klasifikasi berfungsi mengelompokkan data yang didapatkan dari persamaan data asli dengan data yang digunakan sebagai sampel, serta sebagai pemrosesan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan konsep atau kelas data, untuk kepentingan tertentu.[20]

2.6.2 Regresi

Teknik ini memprediksi nilai dari suatu variabel bebas yang diberikan berdasarkan nilai dari variabel tergantung, dengan cara menghitung jumlah variabel bebas yang ditetapkan dengan variabel tergantung [20],

2.6.3 Clustering

Clustering adalah pengelompokan data yang ditentukan clusternya, di dalam cluster terdiri dari beberapa data yang mirip antara satu dengan yang lain, dan memiliki perbedaan bila dibandingkan dengan data dari cluster lain seperti jumlahnya atau model data tersebut. cluster juga bisa sebagai awalan dalam proses data mining, dengan hasil cluster yang telah ditentukan akan menjadi data input dari algoritma yang lain.

Kemudian pendekatan partisi dan clustering dengan pendekatan hirarki digunakan dalam mengembangkan metode tersebut. Dua pendekatan utama tersebut adalah clustering dengan Disamping kedua pendekatan tersebut.[20]

2.6.4 Kaidah Asosiasi

Kaidah asosiasi adalah suatu metode *data mining* yang mampu menemukan pola dan kaidah agar dapat mengelompokkan atribut satu dengan yang lainnya yang saling bergantung dari sekumpulan besar data yang sering muncul secara bersamaan serta Mendeteksi kumpulan atribut-atribut yang muncul bersamaan dalam frekuensi yang sering dan membentuk sejumlah kaidah dari kumpulan-kumpulan tersebut.[20]

2.6.5 Pencarian Pola Sekuensial

Pencarian pola sekuensial atau disebut dengan sequence mining merupakan teknik data mining dengan cara kerja mencari kejadian yang terjadi bersamaan dalam sebuah kasus[20].

2.6.6 Algoritma K-Means

Algoritma ini bekerja dalam pengelompokan data berdasarkan titik pusat cluster terdekat dengan data yang digunakan. Lalu pengelompokan data tersebut bertujuan untuk mengetahui kemiripan data dalam satu cluster dan setiap cluster berbeda beda isinya, ukuran kemiripan yang digunakan dalam klaster adalah jaraknya dan juga bisa jumlah clusternya yang kemudian menarik kesimpulan sebagai hasil.[21]

Pengelompokan data yang bisa membagi data kedalam beberapa cluster, tujuan dari teknik ini adalah menentukan titik pusat massa rata-rata (mean) dari sekupulan data kemudian menentukan cluster di awal secara acak sampai nilainya tidak berubah karena melalui beberapa tahap iterasi tergantung dari data yang digunakan.[22]

Rumus dalam algoritma Kmeans Clustering sebagai berikut :

$$D(N1, C1) = \sqrt{(N_{1AT1} - C_{1AT1})^2 + \dots + (N_{1AT9} - C_{1AT9})^2}$$

Keterangan :

D = Jarak

N1 = Nama pertama

C1 = Centroid

AT1 = Atribut pertama

AT9 = Atribut terakhir

Langkah-langkah dari metode K-Means adalah sebagai berikut :

1. Menentukan jumlah cluster yang ingin dibuat.
2. Menentukan centroid dari data angka atribut.
3. Kemudian menentukan jarak data dari centroid.
4. Lalu mengelompokan data sesuai clusternya dalam jarak terdekat
5. Kembali ke iterasi jika posisi centroid baru dengan centroid lama, tidak sama, jika sama proses perhitungan bisa dihentikan.[22]

A. kelebihan k-means

1. Bisa dilakukan dengan data angka atau angka mewakili huruf.
2. Bisa dipahami dalam satu perhitungan karena menggunakan iterasi.
3. Mendukung berbagai macam aplikasi data mining
4. Bisa menentukan cluster sendiri.
5. Menggunakan tatacara yang tidak terlalu banyak dan dapat di jelaskan kepada pemula.[23]

B. Kekurangan dari k-means

1. Sebelum algoritma di jalankan, titik K diinisialisasikan secara random sehingga pengelompokan data yang di dapatkan bisa berbeda-beda. Namun apabila nilai yang diperoleh acak untuk penginisialisasi kurang baik maka pengelompokan yang didapatkn menjadi tidak optimal.
2. Apabila terjebak dalam kasus yang biasanya di sebut dengan curse of dimensionality. Hal ini pun akan terjadi apabila salah satu data untuk melakukan pelatihan mempunyai dimensi yang sangat banyak,
3. Proses tersebut dapat dipercepat namun dibutuhkan sebuah struktur data yang lebih rumit seperti kD-tree atau hashing untuk melakukan proses tersebut.
4. Penggunaan k secara acak, maka tidak bisa menemukan sebuah cluster yang optimal. [23]

2.7 Software Yang Digunakan

2.7.1 Bahasa R dan Rstudio

Software Rstudio adalah software yang terintegrasi dengan bahasa pemrograman R serta bersifat open source dan gratis, software ini ada dua jenis yaitu RStudio Desktop, di mana program dijalankan secara lokal sebagai aplikasi desktop biasa, serta Rstudio Server yang

digunakan untuk proyek yang lebih besar software ini menggunakan bahasa pemrograman R karena berhubungan dengan analisis statistika dan grafik.

bahasa R didukung dalam output data berbentuk visual gambar grafik yang memudahkan pengguna dalam analisis karena pada dasarnya mendukung packages untuk membangun library script code dan script visualisasi yang bisa dijadikan satu dalam bentuk web. [24]

Analisis Cluster digunakan dalam teknik pengolahan variabel yang bertujuan mengelompokkan objek yang sifatnya hampir sama yang disebut cluster, data yang terintegrasi dengan Algoritma K Means Clustering dikelompokkan ke beberapa kelompok menggunakan bahasa R. Sebagai contoh pengelompokan di dalam bahasa R maka dibuat 2 penghubung jembatan yang terdiri dari total pengeluaran pelanggan dan usia mereka. Untuk meningkatkan promosi dan banyak dikenal masyarakat, tim sales menyebarkan iklan melalui sosial media.[24]

Contoh perhitungan seperti dibawah, Perusahaan yang Anda miliki merencanakan total pengeluaran dan usia pelanggan.

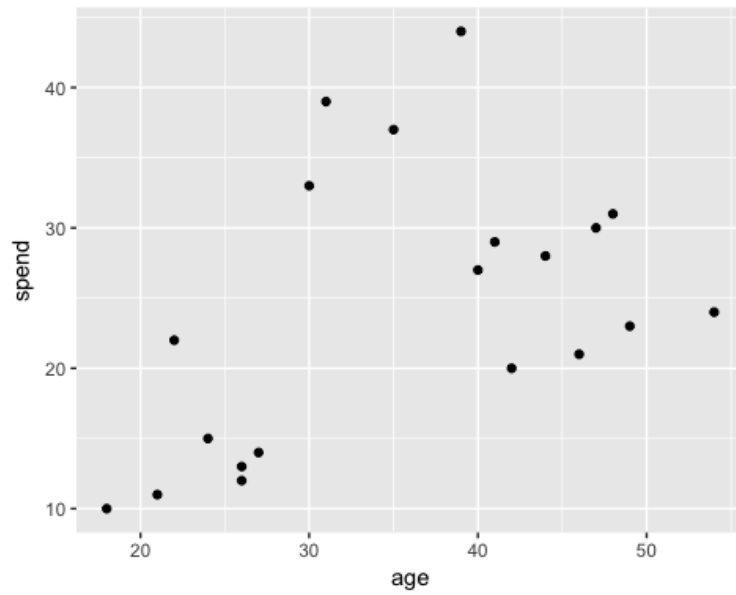
```
perpustakaan (ggplot2)

df <- data.frame (age = c (18, 21, 22, 24, 26, 26,
27, 30, 31, 35, 39, 40, 42, 42, 44, 46, 47, 48,
48, 49, 54 ),

    menghabiskan = c (10, 11, 22, 15, 12, 13, 14,
33, 39, 37, 44, 27, 29, 20, 28, 21, 30, 31, 23,
24)

)

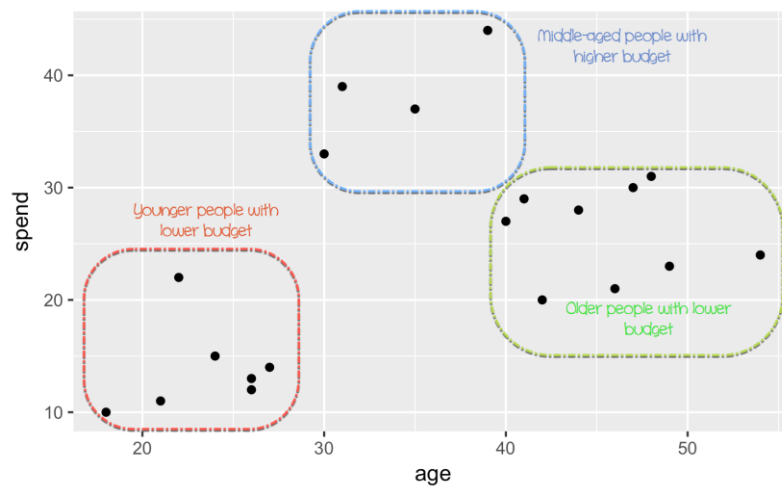
ggplot (df, aes (x = umur, y = menghabiskan)) +
    geom_point ()
```



Gambar 2.1 ada 3 kelompok pola

Keterangan dari Pola Tersebut :

1. Berwarna merah kiri paling bawah menunjukkan bahwa anak muda dengan daya beli rendah.
2. Berwarna biru atas tengah menunjukkan orang yang sudah bekerja bisa membeli lebih banyak.
3. Berwarna Hijau kanan menggambarkan orang tua dengan anggaran lebih rendah.



Gambar 2.2 hasil pengelompokan

Pada hasil tersebut posisi titik dan nilai sangat terlihat jelas dan menunjukkan perbedaannya apabila masing masing kelompok ditandai di dalam kotak, lalu masing masing kotak bisa didefinisikan. Algoritma K-means menemukan kelompok dengan meminimalkan jarak antara pengamatan berdasarkan koordinat [25].

2.7.2 Microsoft Excel

Microsoft Excel adalah sebuah program atau aplikasi yang dimiliki oleh Microsoft Corporation, berfungsi untuk mengolah data dan angka serta terdapat rumus perhitungan bawaan yang bisa muncul otomatis dari software tersebut , didalamnya terdiri dari toolbar dan menu option serta kolom dan baris yang bisa diatur sesuai keinginan untuk memasukan data serta mengeksekusi sebuah perintah didalamnya.

Kegunaan program aplikasi microsoft excel yang terpenting adalah sebagai aplikasi pengolah data dan angka. Software ini memiliki fungsi antara lain:

1. Membuat data, memfilter untuk diurutkan.
2. Melakukan perhitungan dengan rumus.
3. Membantu dalam perhitungan dalam jumlah data besar
4. Membuat grafik presentasi yang memukau.
5. Membuat laporan keuangan. [26]

2.8 Flowchart

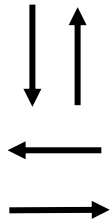


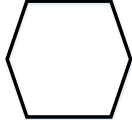
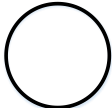

2.8.1 Pengertian Flowchart

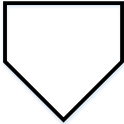


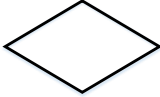
Flowchart adalah simbol simbol grafis yang dipadukan dengan aneka bentuk yang menunjukkan alur algoritma secara menyeluruh serta dipadukan dengan logika. flowchart sebagai penunjuk alur memberikan kemudahan bagi pemakai terhadap pemecahan masalah yang dimasukan di dalam diagram flowchart dibuat. pada dasarnya flowchart hanya dibuat pada saat akan membuat suatu program supaya program tersebut bisa dijalankan.

Flowchart bisa disebut sebagai alat bantu atau gambaran dalam bentuk grafik yang menunjukkan alur proses yang harus dilaksanakan



dalam menyelesaikan suatu permasalahan untuk yang bertujuan untuk mempermudah dan memperjelas langkah-langkah .[27]

Tabel 2.2 Simbol Flowchart

Simbol Flowchart	Fungsi	Simbol Flowchart	Fungsi
Flow Direction symbol		Simbol Manual Input	
	Digunakan untuk menghubungkan symbol pada flowchart satu dengan yang lain.		Digunakan Untuk menginput data
Terminator Symbol		Simbol Preparation	
	Digunakan untuk memulai dan berhenti dari suatu proses yang dilakukan oleh komputer.		Digunakan untuk menggambarkan sebuah persiapan.
Connector Symbol		Simbol Predefine Proses	
	Menggambarkan proses keluar masuk dari symbol-symbol yang dibuat.		Digunakan untuk menggambarkan jalanya sub bagian.

	<p>Connector Symbol</p> <p>Simbol keluar masuk atau penyambungan proses.</p>	<p>Simbol Display</p> <p>Digunakan untuk menyatakan output.</p>
	<p>Processing Symbol</p> <p>Digunakan untuk menampilkan pengolahan pada komputer</p>	<p>Simbol disk and On-line storage</p> <p>Digunakan untuk menyatakan input atau output dari sebuah disk atau penyimpanan.</p>
	<p>Simbol Manual Operation</p> <p>Digunakan untuk menampilkan proses yang tidak dilakukan komputer.</p>	<p>Simbol Magnetik tape unit</p> <p>Digunakan menyatakan sebuah input yang berasal dari pita magnetic</p>
	<p>Simbol Decision</p> <p>Digunakan untuk menggunakan kondisi.</p>	<p>Simbol punch card</p> <p>Menyatakan sebuah input</p>

		yang berasal dari kartu atau output ditulis ke kartu.
--	--	---

	Simbol Input-Output		Simbol Dokumen
	Digunakan untuk menyatakan proses input dan output.		Digunakan untuk menyatakan input yang berasal dari dokumen.

2.8.2 Jenis Jenis Flowchart

1. Flowchart Sistem

Flowchart ini menunjukkan alur kerja sistem yang akan dibuat maupun yang telah dibuat dengan tujuan untuk menjelaskan urutan cara kerja sistem tersebut dari awal sampai akhir.

2. Flowchart Dokumen

Flowchart ini menelusuri kemana alur data form dari sistem dan laporan sistem dari satu bagian ke bagian lainnya baik alur form dan laporan pada saat diproses, dicatat kemudian disimpan.

3. Flowchart Schematic

Flowchart Skematik mirip dengan Flowchart Sistem. Tetapi ditambahkan juga menggunakan gambar peralatan, perangkat hardware, form-form atau peralatan lain yang digunakan dalam sistem.

4. Flowchart Program

Flowchart Program menunjukkan alur yang lebih rinci tentang bagaimana setiap langkah program dalam sistem dari awal sampai akhir yang kemudian dilakukan uji pada program tersebut.

5. Flowchart Proses

Flowchart proses adalah. flowchart ini banyak dipakai di s dan juga analis sistem. Fungsi flowchart ini untuk mengetahui proses dalam suatu proses produksi di perusahaan.[28]