

## **BAB II**

### **LANDASAN TEORI**

#### **2.1. Penelitian Terdahulu**

Berdasarkan penelitian yang dilakukan Irsyad Djamaludin dan Agus Nursikuwagus yang membahas pola pembelian yang ada perusahaan *Roseberry* yang nantinya akan membantu owner untuk membuat suatu keputusan bisnis. Pada penelitian ini metode yang digunakan adalah data mining dengan algoritma Apriori, yang bertujuan untuk menggali informasi baru dan menemukan pola atau aturan tertentu dari data transaksi penjualan yang tersimpan dalam database. Proses penelitian dimulai dengan melakukan ekstraksi data dari laporan penjualan bulan April 2017 yang ada di perusahaan *Roseberry*, sebuah perusahaan rumahan di bidang industri makanan. Data transaksi penjualan tersebut sebelumnya hanya digunakan sebagai arsip dan laporan penjualan saja, tanpa dimanfaatkan lebih lanjut. Penelitian ini menetapkan minimal support sebesar 30% dari jumlah transaksi. Hasil dari proses tersebut adalah itemset 1, 2, dan 3 yang memiliki support yang memenuhi batas minimal dan setelah itu, dilakukan perhitungan confidence untuk menentukan association rule yang relevan. Batas minimal confidence yang ditetapkan dalam penelitian ini adalah 70%. Hasil perhitungan confidence menghasilkan beberapa association rule yang menunjukkan keterkaitan antara produk. Tujuan dari penelitian ini untuk membantu pemilik perusahaan dalam pengambilan keputusan bisnis dengan memahami pola pembelian konsumen

yang ada. [4] Persamaan penelitian yang dilakukan penulis dengan penelitian yang dilakukan oleh Irsyad Djamaludin dan Agus Nursikuwagus yaitu sama menggunakan metode algoritma *apriori*, sedangkan perbedaannya yaitu pada hasil akhirnya berupa website dengan bahasa php sedangkan penulis berupa website dengan bahasa *python* dengan framework *streamlit*.

Penelitian berikutnya yaitu yang telah dilakukan oleh Esis Srikanti dkk, yang membahas tentang peminjaman buku di perpustakaan FST UIN SUSKA Riau menggunakan *algoritma apriori* dengan tujuan untuk mencari aturan asosiasi yang ada pada data peminjam buku. Pada penelitian tersebut Esis Srikanti dkk, mengolah data dari data transaksi peminjaman buku dari bulan Januari 2018 sampai dengan April 2018. Pada tahap praproses data dilakukan tiga buah aktifitas, yaitu: (1) kategorisasi buku; (2) pembersihan data; dan (3) transformasi data. Penerapan Algoritma Apriori dilakukan dengan menggunakan tools Weka 3.8. *Support* anteseden minimal yang digunakan adalah 10% dan *confidence* minimal yang digunakan adalah 60%. Tujuan dari penelitian Esis Srikanti dkk, digunakan untuk mengatur tata letak buku dan rekomendasi peminjaman buku bagi mahasiswa. [5] Persamaan penelitian yang dilakukan penulis dengan penelitian yang dilakukan oleh Esis Srikanti dkk yaitu sama menggunakan algoritma *apriori* dan dilakukan pada perpustakaan, sedangkan perbedaannya yaitu pada pengolahannya menggunakan tools Weka 3.8 sedangkan penulis menggunakan tools *jupyter notebook*.

Dari kedua penelitian tersebut membahas penelitian yang menggunakan algoritma apriori untuk mencari aturan asosiasi dari data yang telah diolah. Pada penelitian pertama, Irsyad Djamiludin dan Agus Nursikuwagus mengolah data transaksi penjualan yang ada di perusahaan *Roseberry*, sedangkan pada penelitian kedua, Esis Srikanti dkk mengolah data transaksi peminjaman buku di perpustakaan FST UIN SUSKA Riau. Kedua penelitian tersebut juga memiliki tujuan yang berbeda. Tujuan penelitian pertama adalah untuk membantu pemilik perusahaan dalam pengambilan keputusan bisnis dengan memahami pola pembelian konsumen yang ada. Sementara itu, tujuan penelitian kedua adalah untuk mengatur tata letak buku dan rekomendasi peminjaman buku bagi mahasiswa. Dalam kedua penelitian tersebut juga digunakan minimum *support* dan minimum *confidence* sebagai parameter untuk mencari aturan asosiasi dari data yang diolah. Namun, nilai minimum *support* dan minimum *confidence* yang digunakan dalam kedua penelitian berbeda-beda, pada penelitian pertama menggunakan *support* 30% dan nilai minimum *confidence* 70% sedangkan penelitian kedua yaitu *support* 10% dan *confidence* minimum yang digunakan adalah 60%.

## **2.2. Teori Pendukung**

Teori pendukung yaitu dasar teori dalam penelitian yang dapat digunakan sebagai acuan dalam melakukan penelitian oleh penulis.

### **2.2.2. Data Mining**

Menurut Hermawati, data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (mechine learning) untuk menganalisis dan mengekstraksi pengetahuan (knowledge) secara otomatis. [6]

Menurut Kusrini dkk, data mining didefenisikan sebagai sebuah proses untuk menemukan hubungan, pola dan trend baru yang bermakna dengan menyaring data yang sangat besar, yang tersimpan dalam penyimpanan, menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. [7]

Data mining didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar. [8]

Berdasarkan dari pengertian di atas dapat disimpulkan bahwa data mining merupakan proses otomatis menggunakan teknik pembelajaran mesin, pengenalan pola, statistik, dan matematika untuk menemukan pola, hubungan, dan tren baru yang bermakna dalam data yang besar dan kompleks dengan tujuan mendapatkan pengetahuan yang berharga.

#### **2.2.2.1. Operasi Data Mining**

Operasi data mining dapat dibedakan menjadi dua berdasarkan sifatnya, yaitu (1) berbasis prediksi (prediction-driven) yang digunakan untuk menjawab

pertanyaan seperti "apa" dan berkaitan dengan sesuatu yang remang-remang atau transparan. Jenis operasi ini digunakan untuk validasi hipotesis, querying dan pelaporan (misalnya: spreadsheet dan pivot table), OLAP (Online Analytic Processing), serta analisis statistik. (2) Berbasis penemuan (discovery-driven) yang bersifat transparan dan digunakan untuk menjawab pertanyaan "mengapa?". Jenis operasi ini digunakan untuk analisis data eksplorasi, pemodelan prediktif, segmentasi basis data, analisis keterkaitan (link analysis), dan deteksi anomali. [9]

#### **2.2.2.2. Teknik Data Mining**

Berikut kelompok data mining berdasarkan tugas yang dapat dilakukan sebagai teknik dalam data mining sebagai berikut.

a. Deskripsi

Penulis biasanya berusaha untuk menemukan cara untuk menjelaskan pola dan tren yang tersembunyi dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menentukan fakta bahwa orang yang kurang profesional cenderung lebih sedikit didukung dalam 14 pemilihan presiden. Deskripsi dari pola dan tren sering memberikan penjelasan yang mungkin terkait dengan pola atau tren tersebut.

b. Estimasi

Estimasi memiliki kemiripan dengan klasifikasi, namun variabel target yang dituju lebih berorientasi pada nilai numerik daripada kategori. Sebagai contoh, estimasi tekanan *systolic* dari pasien di rumah sakit dapat

dilakukan berdasarkan faktor-faktor seperti usia pasien, jenis kelamin, indeks berat badan, dan level sodium dalam darah. Beberapa metode data mining yang digunakan dalam estimasi antara lain *Regression Linear*, *Neural Network*, dan *Support Vector*.

c. Prediksi

Prediksi memiliki kemiripan dengan estimasi dan klasifikasi, namun hasilnya menunjukkan sesuatu yang belum terjadi, seperti kemungkinan terjadi di masa depan. Contohnya, kita dapat melakukan prediksi harga beras tiga bulan ke depan. Beberapa metode data mining yang digunakan dalam prediksi antara lain *Regression Linear*, *Neural Network*, *Support Vector*.

d. Klasifikasi

Dalam klasifikasi variabel, tujuan yang ingin dicapai bersifat kategorik. Sebagai contoh, kita ingin mengelompokkan pendapatan ke dalam tiga kelas, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah. Beberapa metode data mining yang digunakan dalam klasifikasi antara lain *Naive Bayes*, *K-Nearest*, *C4.5*, *ID3*, *CART*, *Linear Discriminant Analysis*, dan *Logistic Regression*.

e. *Clustering*

*Clustering* merupakan teknik yang lebih berfokus pada pengelompokan *record*, pengamatan, atau kasus ke dalam kelas-kelas yang memiliki kemiripan. Suatu cluster adalah kumpulan *record* yang saling

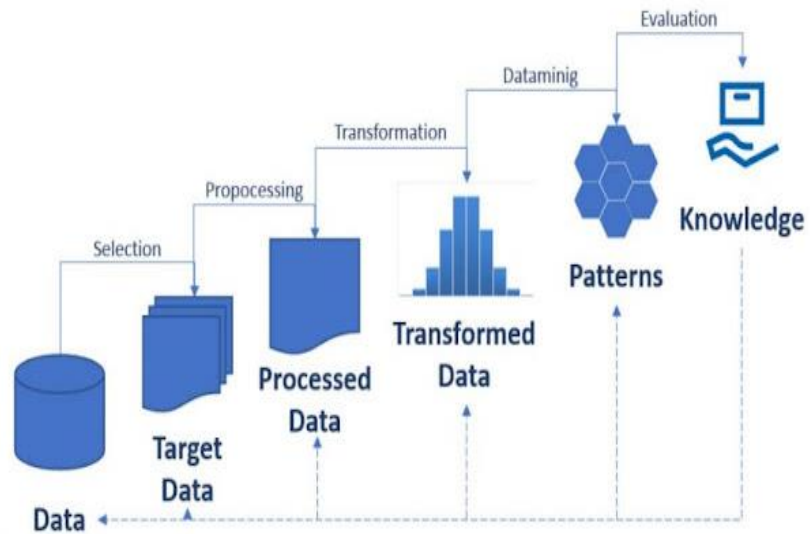
mirip dan berbeda dengan rekaman-rekaman di dalam *cluster* lainnya. Sebagai contoh, dalam tujuan audit akuntansi, dapat dilakukan segmentasi perilaku keuangan ke dalam kategori-kategori yang mencurigakan. Beberapa metode data mining yang digunakan dalam *clustering* antara lain *K-Means*, *K-Medoids*, *Self-Organizing Map (SOM)*, dan *Fuzzy C-Means*.

f. Asosiasi

Mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada waktu tertentu adalah tujuan dari pendekatan asosiasi. Pendekatan ini berfokus pada analisis keranjang pasar yang mengidentifikasi pola pembelian produk yang sering terjadi bersamaan. Beberapa metode data mining yang digunakan dalam pendekatan asosiasi antara lain *FP-Growth*, *A Priori*, *Coefficient of Correlation*, dan *Chi Square*. [6]

### **2.2.2.3. Knowledge Discovery In Database (KDD)**

*Knowledge Discovery in Database (KDD)* merupakan sebuah proses dimana informasi yang bermanfaat dan pola yang terdapat pada data dapat ditemukan. Proses ini dilakukan pada basis data yang berukuran besar dan sebelumnya tidak diketahui keberadaannya, namun memiliki potensi yang sangat bermanfaat.



**Gambar 2. 1 Proses KDD Data Mining**

(Sumber : <https://miqbal.staff.telkomuniversity.ac.id/belajar-data-mining/>)

Tahapan yang di dalam *Knowledge Discovery in Database* (KDD) yaitu sebagai berikut :

1. Data Selection

Dalam proses ini, dilakukan pemilihan set data, pembuatan set data target, atau pemfokusan pada subset variabel (contohnya, sampel data) di mana penemuan akan dilakukan. Hasil seleksi disimpan dalam file terpisah dari basis data operasional.

2. *Pre-Processing*

*Pre-Processing* dilakukan untuk menghilangkan data yang tidak konsisten dan mengurangi noise, serta mengatasi duplikasi dan



kesalahan data. Selain itu, dapat juga menambahkan data eksternal yang relevan untuk memperkaya informasi yang ada.

### 3. *Transformation*

Proses ini melibatkan transformasi atau penggabungan data ke dalam bentuk yang lebih cocok untuk dilakukan proses mining. Hal ini dilakukan dengan melakukan agregasi atau peringkasan data.

### 4. Data Mining

Proses Data Mining merupakan upaya untuk menemukan pola atau informasi penting dalam dataset yang dipilih dengan menggunakan teknik, metode, atau algoritma tertentu yang sesuai dengan tujuan dari keseluruhan proses KDD.

### 5. *Evaluation*

Proses ini bertujuan untuk menerjemahkan pola atau informasi yang ditemukan melalui Data Mining dan mengevaluasi apakah sesuai atau bertentangan dengan fakta atau hipotesis sebelumnya. Hasil dari pengetahuan yang diperoleh dari pola-pola tersebut kemudian disajikan dalam bentuk visualisasi. [7]

#### **2.2.1. Perpustakaan**

Perpustakaan merupakan tempat yang memiliki banyak koleksi buku pengetahuan, majalah dan media lainnya yang ditata atau diatur dengan cara tertentu agar mudah dimanfaatkan oleh para pengguna secara tepat dan baik. [10]

Walaupun dapat diartikan sebagai koleksi pribadi perseorangan, namun perpustakaan lebih umum dikenal sebagai sebuah koleksi besar yang dibiayai dan dioperasikan oleh sebuah kota atau institusi, dan dimanfaatkan oleh masyarakat yang rata-rata tidak mampu membeli sekian banyak buku atas biaya sendiri.

[11]

### 2.2.3. Association Rules

Association rule merupakan metode yang bertujuan mencari pola yang sering muncul di antara banyak transaksi, dimana setiap transaksi terdiri dari beberapa item sehingga metode ini akan mendukung rekomendasi transaksi yang terjadi. [2]

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap :

1. Analisis pola frekuensi tinggi

Dalam tahapan ini mencari kombinasi item yang memenuhi syarat minimum dari nilai support dalam database. Nilai support sebuah item diperoleh dengan rumus berikut:

$$\text{Support (A)} = \frac{\text{Jumlah Transaksi Mengandung A}}{\text{Total Transaksi}}$$

Sementara nilai support dari 2 item diperoleh dengan rumus berikut:

$$\text{Support (A} \cap \text{B)} = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Total Transaksi}}$$

## 2. Pembentukan aturan asosiasi

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiatif yang memenuhi syarat minimum untuk confidence dengan menghitung confidence aturan asosiatif  $A \rightarrow B$  Nilai confidence dari aturan  $A \rightarrow B$  diperoleh dari rumus berikut:

$$\text{Confidence } P(B|A) = \frac{\text{Jumlah Transaksi Mengandung A dan B}}{\text{Jumlah Transaksi Mengandung A}}$$

### 2.2.4. Algoritma Apriori

Algoritma apriori merupakan algoritma yang paling terkenal untuk menemukan pola frekuensi tinggi. algoritma apriori dibagi menjadi beberapa tahap yang disebut narasi atau pass. [12]

Algoritma Apriori menggunakan pengetahuan mengenai frekuensi dari item set yang telah diketahui sebelumnya, untuk dijadikan informasi pada iterasi selanjutnya. Algoritma Apriori memperhatikan nilai minimum support dalam menentukan kandidat-kandidat yang mungkin muncul.

#### 1. Pembentukan Kandidat Item set

Proses awal dalam algoritma apriori adalah pembentukan kandidat itemset atau himpunan item yang muncul bersama-sama dengan frekuensi yang signifikan. Kandidat k-itemset dibentuk dari kombinasi k-1 itemset yang telah ditemukan pada iterasi sebelumnya. Dalam algoritma apriori, dilakukan pemangkasan kandidat k-itemset dengan

mengecek subsetnya yang berisi k-1 item, apakah termasuk dalam pola frekuensi tinggi dengan panjang k-1. Jika tidak termasuk, maka kandidat k-itemset tersebut akan dihapus dari daftar.

## 2. Penghitungan support dari tiap kandidat k-itemset

Dalam algoritma Apriori, dukungan atau support untuk setiap kandidat k-itemset diperoleh melalui pemindaian atau scanning database untuk menghitung berapa kali transaksi yang memuat semua item dalam kandidat k-itemset tersebut. Proses ini merupakan fitur khas dari algoritma Apriori, dimana perhitungan dilakukan pada seluruh database untuk setiap k-itemset yang terpanjang.

## 3. Tetapkan pola frekuensi tinggi

Pola frekuensi tinggi yang memuat k item atau kitemset ditetapkan dari kandidat k-itemset yang supportnya lebih besar dari minimum support.

## 4. Bila tidak didapat pola frekuensi tinggi baru maka seluruh proses dihentikan. Bila tidak, maka k ditambah satu dan kembali bagian 1. [13]

### **2.3 Piranti Pendukung**

Piranti pendukung merupakan perangkat atau alat yang digunakan oleh peneliti untuk melaksanakan sebuah penelitian Sehingga peneliti dapat lebih mudah dalam menghasilkan data dan hasil penelitian yang lebih akurat.

### 2.3.1. Python

*Python* merupakan bahasa pemrograman yang termasuk dalam kategori berlevel tinggi. Guido Van Rossum, seorang programmer asal Belanda, menciptakan *python* sebagai sebuah program. Namun, *Python* didesain dengan cara yang memudahkan dalam hal pembelajaran dan pemahaman. Keunggulan program *python* adalah mudah dipelajari, dapat menjalankan fungsi kompleks dalam program dengan mudah, membutuhkan sedikit kode, dan mampu mengubah program yang kompleks menjadi lebih mudah. Meskipun demikian, *python* juga memiliki beberapa kekurangan, seperti kinerja yang lambat, kurang mendukung untuk *Android* dan *IOS*, keterbatasan dalam mengakses basis data, serta kurang cocok untuk tugas *multi-core/multi-processor*. [14]

### 2.3.2. Anaconda

*Aconda* merupakan distribusi *python* untuk memproses data dengan skala yang lebih besar, *predictive analytics*, dan *scientific computing*. *Anaconda* sudah terdiri dari beberapa paket seperti *Numpy*, *Scipy*, *matplotlib*, *pandas*, *IPython*, *Jupyter Notebook*, dan *scikitlearn*. *Anaconda* tersedia untuk semua sistem operasi, merupakan solusi yang sangat mudah dan sangat disarankan karena tidak perlu menginstal paket *python* satu-persatu. [15]

### **2.3.3. Jupyter Notebook**

*Jupyter Notebook* merupakan sebuah alat populer untuk melakukan manipulasi data pada bahasa pemrograman *Python*. *Jupyter Notebook* ini memungkinkan interaksi antara kode dengan output dalam satu dokumen. Dokumen yang dihasilkan memiliki format file dengan ekstensi "ipynb" dan dibuat menggunakan aplikasi *Jupyter Notebook*. Di dalam dokumen tersebut terdapat kode komputer serta elemen teks yang kaya seperti paragraf, rumus matematika, gambar, dan tautan. Sebelumnya, *Jupyter Notebook* dikenal sebagai *IPython Notebook* dan dalam waktu dekat akan berkembang menjadi *Jupyter Lab*. [16]

### **2.3.4. Website**

Menurut William Jonathan dan Sri Lestari *Website* adalah kumpulan dari halaman-halaman situs, yang terkumpul dalam sebuah domain atau subdomain, yang tempatnya berada di dalam *World Wide Web* (WWW) di dalam internet. Sebuah halaman web biasanya berupa dokumen yang ditulis dalam format HTML (*Hyper Text Markup Language*), yang selalu bisa diakses melalui HTTP, yaitu sebuah protokol yang menyampaikan informasi dari server *website* untuk ditampilkan kepada para pemakai melalui *web browser*. *Website* atau situs dapat juga diartikan sebagai kumpulan halaman yang menampilkan informasi data teks, data gambar diam atau gerak, data animasi, suara, video dan atau gabungan dari semuanya, baik yang bersifat statis maupun dinamis yang membentuk satu

rangkaian bangunan yang saling terkait dimana masing-masing dihubungkan dengan jaringan-jaringan halaman (*hyperlink*). [17]

Menurut Rayanti Veronica Siregar dkk *website* adalah kumpulan dari halaman *web* yang saling berhubungan yang berisi berbagai informasi dalam bentuk teks, suara, gambar, video, dan lainnya yang berada pada suatu domain di internet yang ditulis dalam format *HTML* serta bisa diakses melalui *HTTP*. *HTTP* merupakan suatu protokol yang digunakan untuk menyampaikan berbagai informasi dari server *website* yang kemudian ditampilkan kepada para *user* atau pemakai melalui *web browser*. [18]

Berdasarkan dari pendapat diatas *website* merupakan kumpulan halaman *web* yang saling terkait yang berisi berbagai informasi dalam bentuk teks, gambar, suara, dan video. *Website* tersebut ditulis dalam format *HTML* dan dapat diakses melalui protokol *HTTP*. *Website* atau situs terdiri dari halaman-halaman situs yang terangkum dalam sebuah domain atau subdomain, yang tempatnya berada di dalam *World Wide Web (WWW)* di dalam internet. Setiap halaman *web* biasanya berupa dokumen yang ditulis dalam format *HTML* dan selalu bisa diakses melalui *HTTP*, yaitu sebuah protokol yang menyampaikan informasi dari server *website* untuk ditampilkan kepada para pemakai melalui *web browser*.

### **2.3.5. Streamlit**

Streamlit adalah kerangka kerja web yang ditujukan untuk menyebarkan model dan visualisasi dengan mudah menggunakan bahasa *Python*, yang cepat dan minimalis tetapi juga memiliki tampilan yang cukup baik serta ramah pengguna. Tersedia widget bawaan untuk masukan pengguna, seperti pengunggahan gambar, penggeser, masukan teks, dan elemen *hypertext markup language* (HTML) lain yang sudah dikenal, seperti *checkboxes* dan radio *buttons*. Setiap kali pengguna berinteraksi dengan aplikasi *Streamlit*, skrip *Python* dijalankan kembali dari atas ke bawah. Hal ini merupakan sebuah konsep penting yang perlu diingat saat mempertimbangkan berbagai status aplikasi yang akan dipilih.

*Streamlit* merupakan aplikasi yang tidak berbayar dan pengguna tidak perlu memiliki pengetahuan pengembangan *front-end* yang mahir untuk mengoperasikannya. *Streamlit* dapat dijalankan pada editor *Anaconda* serta bahasa *Python* seri 3.7 ke atas, tetapi tidak mendukung pada editor *Jupyter Notebook*, sehingga harus dikonversi ke editor *Pycharm* atau *Visual Code*. Tampilan beranda pada aplikasi *Streamlit* dapat dipisahkan menjadi dua bagian, yaitu *buttons*, untuk pemilihan menu, serta tampilan visual chart. Hal ini menyebabkan dibutuhkananya *library NumPy* serta *Pandas* untuk menampilkan grafik. Keluaran grafik sejalan dengan hasil olah data metode pembelajaran mesin menggunakan kombinasi lapisan tersembunyi LSTM dan GRU. *Buttons* berfungsi untuk memilih dataset dari kategori negara, jenis hewan, arsitektur lapisan tersembunyi, *optimizer*, serta pilihan untuk *epoch* dan prediksi dalam beberapa tahun ke depan. [19]



### 2.3.6. VsCode (*Visual Studio Code*)

Visual Studio Code adalah kode editor sumber yang dikembangkan oleh Microsoft untuk Windows, Linux dan macOS. Ini termasuk dukungan untuk debugging, kontrol git yang tertanam dan GitHub, penyorotan sintaksis, penyelesaian kode cerdas, *snippet*, dan *refactoring* kode. Ini sangat dapat disesuaikan, memungkinkan pengguna untuk mengubah tema, pintasan *keyboard*, preferensi, dan menginstal ekstensi yang menambah fungsionalitas tambahan. [20]

Muncul dengan dukungan bawaan untuk *JavaScript*, *TypeScript* dan *Node.js* dan memiliki array dengan beragam ekstensi yang tersedia untuk bahasa lain seperti *C++*, *C*, *Python* dan *PHP*. *Visual studio code* gratis dan *open source*, meskipun unduhan resmi berada dibawah *lisensi proprietary*. [21]

### 2.3.7. *Flowchart*

Pengertian *Flowchart* (Diagram Alir) atau di sebut *Flowchart* merupakan bagan (*Chart*) yang mengarahkan alir (*flow*) di dalam prosedur atau program sistem secara logika. *Flowchart* adalah cara untuk menjelaskan tahap-tahap pemecahan masalah dengan merepresentasikan simbol-simbol tertentu yang mudah dipahami, mudah digunakan dan standar. Tujuan penggunaan *flowchart* adalah untuk menggambarkan suatu tahapan penyelesaian masalah secara sederhana, terurai dan rapi dengan menggunakan simbol-simbol yang standar yang dapat di mengerti oleh programmer. [22]