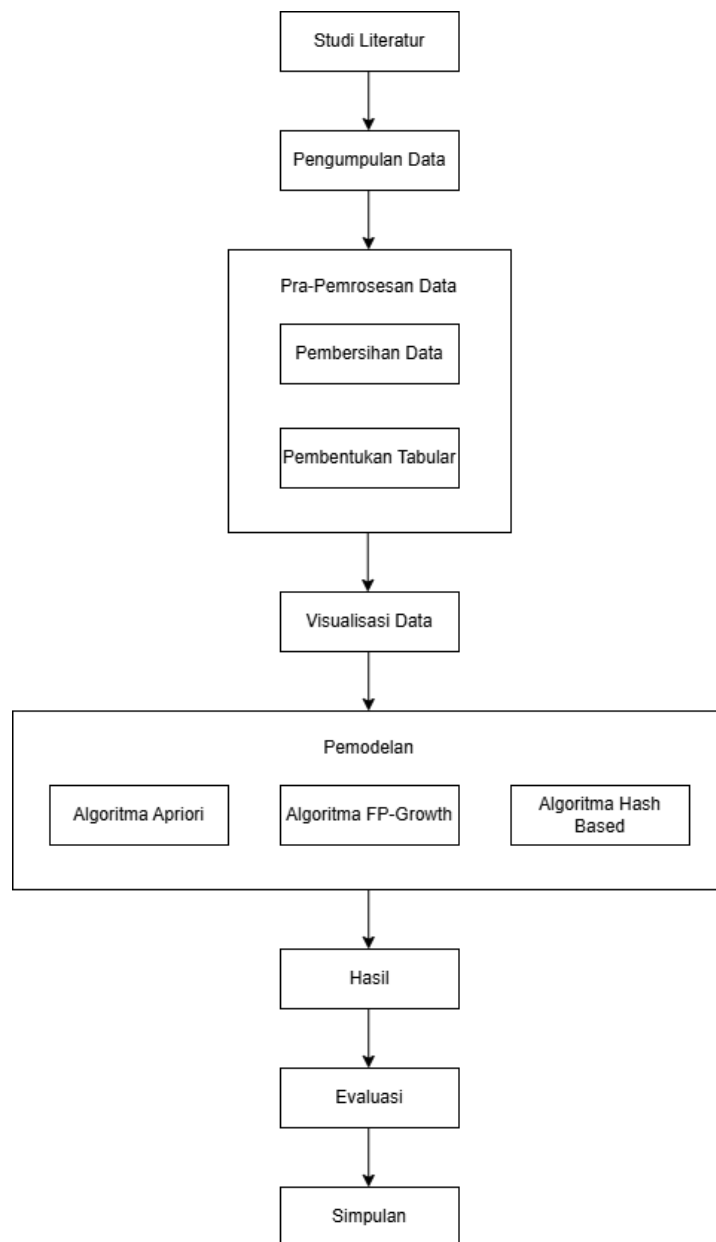


BAB III

METODOLOGI PENELITIAN

Tahapan penelitian yang digunakan untuk mendukung penelitian ini dapat dilihat pada Gambar 3.1.



Gambar 3.1 Tahapan Penelitian

3.1 Studi Literatur

Studi literatur dilakukan dengan melakukan eksplorasi jurnal penelitian dan pekerjaan terkait. Pencarian selanjutnya adalah memilih paper yang relevan yaitu mengenai perbandingan algoritma pada *Market Basket Analysis*. Sehingga hasilnya dapat menciptakan lingkungan *State of The Art* yang mendukung jalannya penelitian ini.

3.2 Pengumpulan Data

Penelitian ini menggunakan metode pengumpulan data yang berasal dari situs web Kaggle yang diunggah oleh Peter Andersen pada tahun 2023. Dataset yang digunakan sebanyak 20.003.

3.3 Pra-Pemrosesan Data

Pra-Pemrosesan data dilakukan untuk membersihkan data agar siap untuk pemodelan data. Tahapan yang dilakukan adalah sebagai berikut:

3.3.1 Pembersihan Data

Tahapan ini dilakukan untuk menghapus atau menghilangkan data yang duplikat atau sama sehingga dapat menghasilkan data yang bersih, dan dapat diandalkan untuk analisis atau pengambilan keputusan.

3.3.2 Pembentukan Tabular

Pada tahapan ini dilakukan untuk membersihkan data yang tidak digunakan dalam pencarian data, yang digunakan untuk mencari aturan asosiasi adalah data obat dan nomor transaksi selanjutnya data tersebut diubah dalam bentuk

tabulah dengan berisikan binary 1 untuk item yang terjual pada data transaksi dan 0 untuk item yang tidak terjual pada data transaksi.

3.4 Visualisasi Data

Visualisasikan data merupakan proses menggambarkan data yang telah dikumpulkan dalam bentuk *chart* atau grafik, sehingga data lebih mudah dipahami dan memperkuat hasil yang diperoleh dari algoritma tersebut. Visualisasi memungkinkan untuk menampilkan data dalam bentuk yang lebih mudah dipahami dan memungkinkan untuk menemukan pola atau hubungan yang mungkin tidak terlihat pada data yang tidak diolah.

3.5 Pemodelan

Pada penelitian ini dilakukan perbandingan algoritma dengan menggunakan *Market Basket Analysis* untuk mencari pola transaksi obat. Perbandingan akan dilakukan dengan menggunakan tiga algoritma yaitu Algoritma Apriori, FP-Growth, dan Hash Based.

3.5.1 Algoritma Apriori

a) Analisis pola frekuensi tertinggi

Tahap ini mencari kombinasi item yang memenuhi nilai *support*. Nilai *support* suatu item dapat diperoleh menggunakan persamaan (1). Sementara itu, nilai *support* untuk dua item diperoleh menggunakan persamaan (2).

b) Pembentukan aturan asosiasi

Setelah semua pola frekuensi tinggi ditemukan, kemudian dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung nilai *confidence* aturan asosiasi $A \rightarrow B$. Nilai *confidence* dari aturan $A \rightarrow B$ diperoleh dari persamaan (4).

3.5.2 Algoritma FP-Growth

Pada pencarian asosiasi, teknik *FP-Growth* merupakan perkembangan dari algoritma Apriori. Kedua metode ini sama-sama mencari nilai *confidence* dan *support*, namun *FP-Growth* hanya membutuhkan satu iterasi dan terdiri dari tiga tahap: pembentukan *pattern conditional base*, pembentukan *FP-Tree* berdasarkan kondisi, dan pencarian itemset yang sering muncul dengan menggunakan TID (*Transaction Identity*).

3.5.3 Algoritma Hash Based

a) Proses Pengolahan Data

Tahapan paling awal adalah pengolahan data. Tahap ini bertujuan untuk memperoleh kombinasi dari setiap item dengan menggunakan aturan asosiatif.

b) Pencarian kombinasi *itemset mining*

Dilakukan selanjutnya adalah melakukan pengolahan *hashing* pada kandidat disetiap 1-itemset untuk memasukkan itemset kedalam tabel *hash* menggunakan persamaan (7).

3.6 Hasil

Pada tahapan ini diperoleh hasil dari algoritma Apriori, FP-Growth, dan Hash Based berdasarkan parameter yang dilakukan pada penelitian ini yaitu Aturan asosiasi, Waktu komputasi, Tingkat akurasi, *Lift ratio*, dan *Memory usage*.

3.7 Evaluasi

Untuk mendapatkan hasil yang akurat dalam penelitian ini, perlu dilakukan pengujian. Proses pengujian ini dilakukan dengan menganalisis dan mengevaluasi hasil dari proses ketiga metode tersebut. Dalam hal ini dilakukan evaluasi menggunakan parameter-parameter yang telah ditentukan:

3.7.1 Aturan Asosiasi

Aturan asosiasi merupakan pola hubungan antara item dalam suatu data transaksi. Dilakukan pencarian aturan asosiasi kombinasi item yang sering muncul (*frequent itemsets*) dan membentuk aturan berdasarkan nilai *support* dan *confidence* pada algoritma Apriori dan FP-Growth, dan Hash Based.

3.7.2 Waktu Komputasi

Waktu Komputasi merupakan lama waktu yang dibutuhkan oleh algoritma Apriori, FP-Growth, dan Hash Based untuk menemukan *frequent itemsets* dan membentuk aturan asosiasi berdasarkan nilai *support* dan *confidence*.

3.7.3 Tingkat Akurasi

Tingkat Akurasi merujuk pada seberapa tepat dan relevan aturan asosiasi yang dihasilkan dalam menggambarkan pola sebenarnya pada data. Rumus untuk menghitung akurasi seperti pada persamaan (8) (Maulidiya & Jananto, 2020).

$$\frac{\Sigma \text{Support Algoritma A}}{\Sigma \text{Support Algoritma B}} \quad (8)$$

3.7.4 Lift Ratio

Lift Ratio adalah suatu ukuran untuk mengetahui kekuatan aturan asosiasi (*association rules*) yang telah terbentuk. Nilai *Lift ratio* digunakan sebagai penentu apakah aturan asosiasi valid atau tidak valid. Rumus untuk menghitung *Lift ratio* seperti pada persamaan (5).

3.7.5 Memory Usage

Memory Usage menampilkan ukuran seberapa banyak memori (RAM) yang digunakan oleh algoritma Apriori, FP-Growth, dan Hash Based pada saat proses komputasi dijalankan.

3.8 Simpulan

Pada tahapan ini menyajikan simpulan berdasarkan tahapan hasil dan evaluasi yang diperoleh pada perbandingan algoritma Apriori, FP-Growth dan Hash Based dengan parameter Aturan asosiasi, Waktu komputasi, Tingkat akurasi, *Lift ratio*, dan *Memory usage*.

3.8.1 Perbandingan Algoritma

Melakukan perbandingan hasil algoritma Apriori, FP-Growth, dan Hash Based berdasarkan parameter yang telah dilakukan.

3.8.2 Perbandingan Parameter dengan Penelitian Sebelumnya

Melakukan perbandingan parameter yang dilakukan pada penelitian ini dengan penelitian sebelumnya.

3.8.3 Perbandingan Perhitungan Tiap Parameter

Menkonversikan hasil perbandingan parameter yang sudah diperoleh sebelumnya, ke dalam bentuk (%).