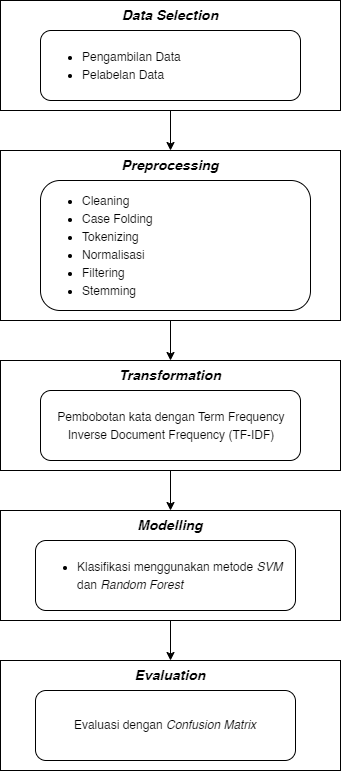
# BAB III METODE PENELITIAN

Metodologi yang digunakan yaitu *Knowledge Discovery in Database* (KDD) dengan dua pemodelan, pertama *SVM* dan yang kedua *Random Forest*. Dapat dilihat pada gambar dibawahini.



## *Data Selection*

Pada tahap ini berisi pengambilan data dan pelabelan data. Pengumpulan data dalam penelitian ini dengan cara *web* *scaping,* denganmenggunakan library *google-play-scrapper*. Dari proses *web scraping* tersebut, didapatkan data sebanyak 2.787 data. Kemudian data tersebut diklasifikasi ke dalam 2 *class* yaitu Positif dan Negatif berdasarkan nilai rating yang diberikan pengguna. Dimana nilai rating 1-3 dilabelkan sebagai kelas negatif sedangkan 4-5 dilabelkan sebagai kelas positif.

## *Preprocessing*

Preprocessing terdapat beberapa tahapan diantaranya:

* 1. *Cleaning* digunakan untuk proses membersihkan data yang tidak memiliki pengaruh pada proses data mining.
  2. *Case* *Folding* bertujuan untuk merubah semua huruf kapital dalam dokumen menjadi huruf kecil.
  3. *Tokenizing* digunakan untuk memotong kalimat menjadi sebuah kata berdasarkan spasi.
  4. Normalisasi digunakan untuk memperbaiki penulisan kata yang tidak sesuai menjadi sesuai dengan kamus KBBI.
  5. *Filtering* digunakan untuk menghilangkan kata - kata yang tidak bermakna atau kurang penting dengan topik penelitian.
  6. *Stemming* digunakan untuk menghilangkan kata imbuhan dengan menggunakan library sastrawi.

## *Transformation*

Tahapan ini memberikan bobot setiap kata untuk melihat seberapa jauh hubungan kata terhadap dokumen. Bobot dihitung dengan menggunakan *Term Frequency Inverse Document Frequency* (TF-IDF).

## *Modeling*

Pada tahap ini pemodelan menggunakan metode klasifikasi *SVM* dan *Random Forest*, data yang telah melewati tahap *preprocessing* akan dibagi terlebih dahulu menjadi 4 skenario yaitu data *training* dan *testing* dengan perbandingan 90:10, 80:20, 70:30, 60:40*.*

## Evaluasi

Setelah dilakukan proses *modeling* menggunakan metode *SVM* dan *Random Forest*, maka akan dilakukan evaluasi peforma model menggunakan *confusion matrix*. berdasarkan hasil dari pengujian tersebut maka dapat disimpulkan metode mana yang memiliki peforma lebih baik dalam menganalisa sentimen ulasan pengguna aplikas M-Paspor