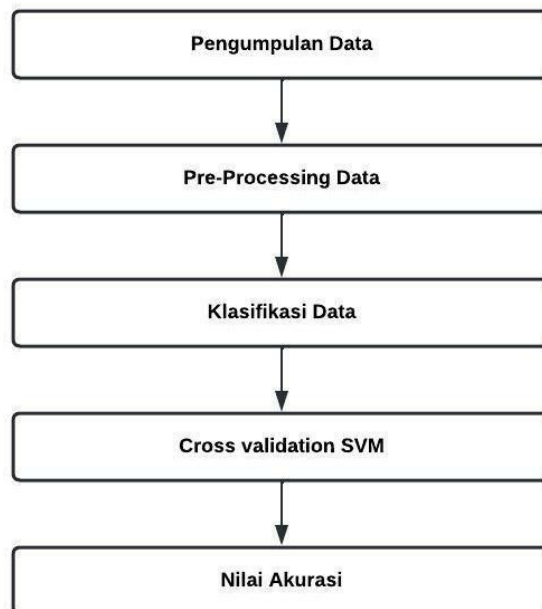


BAB III METODELOGI PENELITIAN

3.1. Metode Penelitian

Metode penelitian yang di gunakan yaitu *Support Vector Machine* adalah algoritma pembelajaran mesin yang diawasi yang dapat digunakan untuk klasifikasi dan regresi.



Gambar 3. 1 : Metode Penelitian

3.1.1. Pengumpulan Data

Proses pengumpulan data dilakukan dengan *web scraping* menggunakan *tools* data miner yang ada di google chrome. Data yang diambil merupakan ulasan komentar pada aplikasi Traveloka pada *google play store*.

3.1.2. Pre-Processing Data

Data-data yang sudah dikumpulkan menjadi data uji akan melalui tahap pre-processing melalui tahapan analisis sentimen, *text mining* serta menggunakan *tools Rapidminer* data mining. Proses ini bertujuan agar data dapat diklasifikasi dan untuk mempermudah proses analisis menggunakan algoritma SVM.

3.1.3. Klasifikasi Data

Klasifikasi data dilakukan menggunakan tools *Rapidminer* data mining. Dimana pada tahap ini dilakukan pengklasifikasian data ulasan komentar kedalam sentimen positif dan negatif untuk selanjutnya diolah menggunakan algoritma Support Vector Machine (SVM).

3.1.4. Penerapan Algoritma Support Vector Machine (SVM)

Metode *Support Vector Machine* merupakan metode yang digunakan untuk analisis sentimen pada penelitian ini. Hasil yang akan ditentukan dengan metode ini adalah nilai akurasi algoritma SVM yang melalui tahap *Cross Validation*.

3.2. Tahapan Penelitian

adapun tahapan penelitian yang dilakukan peneliti sebagai berikut :

3.2.1 Metode Pengumpulan Data

Proses pengumpulan data dilakukan dengan *web scraping* menggunakan tools data miner yang ada di google chrome. Data yang diambil merupakan ulasan komentar pada aplikasi Traveloka pada *google play store*. Jenis data yang dikumpulkan dalam tugas akhir ini adalah data sekunder. Data sekunder adalah data yang dikumpulkan melalui referensi pembacaan buku-buku yang berkaitan erat dengan kajian masalah yang diteliti, misalnya pengumpulan bahan dari skripsi, dari media internet dan jurnal serta referensi buku. Metode yang digunakan dalam proses pengumpulan data adalah sebagai berikut:

1. Observasi

Peneliti melakukan pengamatan secara langsung pada situs *google play store* untuk mempelajari ulasan persepsi pengguna aplikasi traveloka yang erat kaitannya dengan objek yang diteliti. Pengamatan ini dilakukan peneliti untuk menambah pengetahuan mengenai topik yang diangkat penulis.

2. Studi Pustaka

Pengumpulan data dilakukan dengan mempelajari dan membaca untuk mendapatkan landasan teori atau pemikiran serta data-data empiris yang bersumber dari buku, internet, maupun penelitian penelitian sebelumnya.

3.2.2. Alat dan Bahan Analisis Kebutuhan Sistem

Dalam penerapan metode *Support Vector Machine* membutuhkan alat dan bahan untuk menganalisa, dimana alat dan bahan ini sangat diperlukan sebagai analisis kebutuhan sistem. Adapun alat dan bahan yang dibutuhkan dalam penelitian ini adalah sebagai berikut:

A. Alat

Alat yang akan digunakan sebagai implementasi sebagai kebutuhan non fungsional yang merupakan apa saja yang dimiliki oleh system agar dapat berjalan, alat yang digunakan sebagai berikut:

1. Perangkat Lunak (*Software*)

Perangkat lunak yang dibutuhkan dalam analisis *data mining* menggunakan metode *Support Vector Machine* adalah sebagai berikut:

a. Sistem Operasi

Dibutuhkannya sistem operasi berupa *windows 10* dikarenakan *support* dalam pemakaian aplikasi *Rapidminer* pada *windows 64 bit* dan *windows 10* ini juga termasuk meminimalisir kendala yang tidak diinginkan.

b. Data Set

Microsoft excel digunakan sebagai alat perhitungan manual antar kelas positif dan kelas negatif sampai menjadi iterasi yang menetap.

c. Analisis Data

Rapidminer adalah aplikasi yang dimana dalam penerapan metode klusterisasi dapat membentuk hasil visualisasi juga perhitungan yang instan sehingga hasil manual pada *microsoft excel* dan hasil yang di terapkan dalam *Rapidminer* menggunakan metode *Support Vector Machine* harus relevan, sama dan sebanding.

B. Perangkat Keras (*Hardware*)

Perangkat keras yang digunakan dalam menganalisa data adalah sebagai berikut:

a. Processor Core- i3

Dibutuhkannya *Processor Core-i3* agar penggunaan dalam pemakaian aplikasi implementasi dapat berjalan dengan lancar karena jenis processor ini mampu bekerja secara multi proses pada bagian kecil.

b. Hardisk 320 GB

Hardisk 320 GB memiliki kapasitas yang cukup besar sehingga mampu dijadikan sebagai alat acuan dalam pemakaian aplikasi yang digunakan untuk analisis data tersebut.

c. RAM 4 GB

RAM 4 GB sudah cukup untuk memenuhi standar kriteria yang digunakan dalam menyimpan data set yang cukup banyak untuk penerapan dan pemakaian aplikasi analisis data dan segala macam aplikasi perhitungan seperti *microsoft excel*.

d. Keyboard dan Mouse Standar

Keyboard dan mouse standar sudah cukup mampu sebagai alat yang digunakan untuk membantu dalam penelitian ini.

C. Bahan

Bahan yang digunakan dalam penerapan metode *Support Vector Machine* yaitu dimuat dalam aplikasi *microsoft excel*, bahan yang dianalisis yaitu *data set*. Yang mana *data set* merupakan sekumpulan data tabel dengan berupa relasi antar tabel yang digunakan untuk mempresentasikan data beserta relasinya seperti *data set* yang digunakan pada penerapan *Support Vector Machine* yaitu mendeskripsikan data ulasan pengguna aplikasi dimana terdapat *field* beserta keterangan yang akan dianalisis.

Tabel 3.1 Data Set

Deskripsi	Field	Keterangan
Data Ulasan	<i>Username</i>	Nama Pengguna Aplikasi
	<i>Content</i>	Ulasan atau komentar pengguna terhadap aplikasi traveloka
	<i>Score</i>	Penilaian yang diberikan pengguna terhadap aplikasi traveloka

3.2.3 Processing Data

Adapun tahapan *processing Data* sebagai berikut :

3.2.3.1. Analisis Sentimen

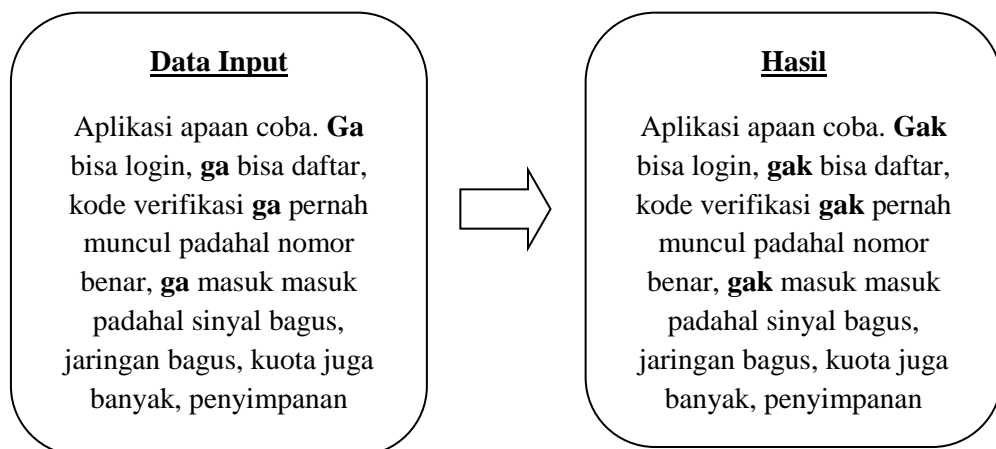
Analisis sentimen atau *opinion mining* adalah bidang ilmu data mining yang mempunyai tujuan untuk menganalisis, memahami, mengolah dan mengekstrak data tekstual yang berupa opini, sentimen, evaluasi, sikap, dan emosi terhadap suatu entitas seperti produk, servis, organisasi, individu, dan topik tertentu. Klasifikasi sentimen digunakan untuk menyelesaikan masalah klasifikasi dua kelas, positif dan negatif.

3.2.3.2. Text Mining

Teks yang dilakukan pada proses *text mining*, pada umumnya memiliki beberapa karakteristik berdimensi tinggi, terdapat *noise* pada data, dan terdapat struktur teks yang tidak baik. Untuk mendapatkan bentuk data yang siap diproses oleh data mining, maka harus dilakukan *text preprocessing* terlebih dahulu dengan cara penyeleksian kata. Setiap kata dipecah menjadi bagian yang lebih kecil sehingga mempunyai arti yang lebih sempit.

1. Spelling Normalization

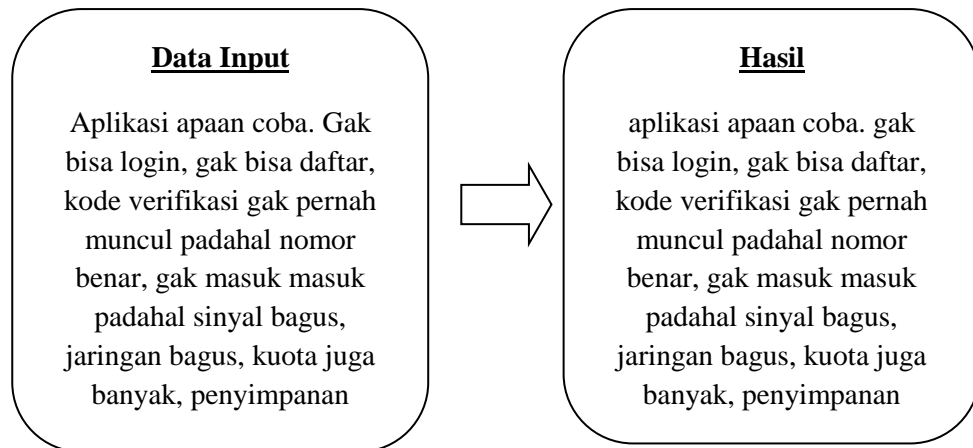
Proses perbaikan kata-kata yang masih salah dalam pengejaan atau merupakan singkatan dalam bentuk tertentu. Perbaikan kata dilakukan agar jumlah perhitungan dimensi kata tidak melebar. Perhitungan jumlah dimensi kata akan melebar jika kata yang salah eja atau kata singkatan tidak dirubah karena akan dihitung sebagai kata yang berbeda pada proses penyusunan matriks.



Gambar 3.2 Proses Spelling Normalization

2. Case Folding

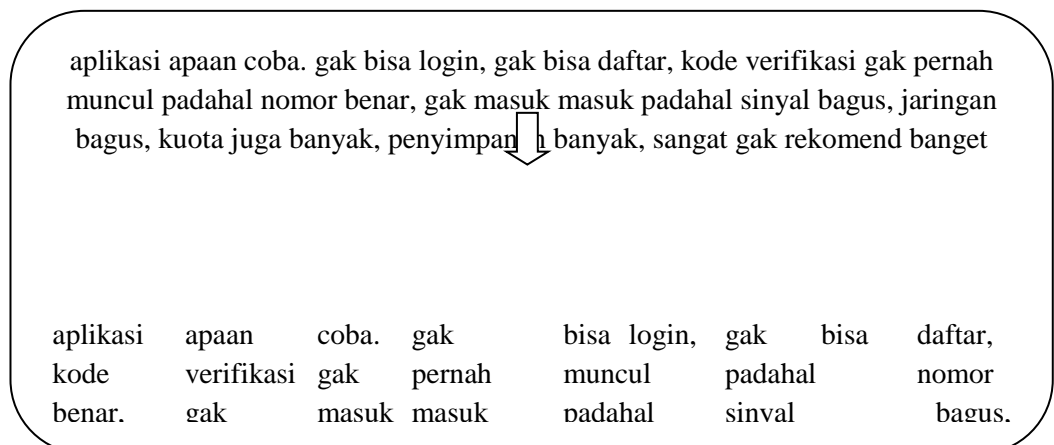
Case folding adalah mengubah semua huruf dalam dokumen menjadi satu bentuk yang sama, misalnya dokumen diubah menjadi huruf kecil atau sebaliknya dan hanya huruf „a“ sampai huruf „z“ yang diterima. Selain karakter huruf dihilangkan.



Gambar 3.3 Proses Case Folding

3. Tokenizing

Proses pemisahan teks menjadi potongan kalimat dan kata yang disebut token. Tujuan dari proses ini adalah mendapatkan potongan kata yang akan menjadi entitas yang memiliki nilai dalam matriks dokumen teks yang akan dianalisis.

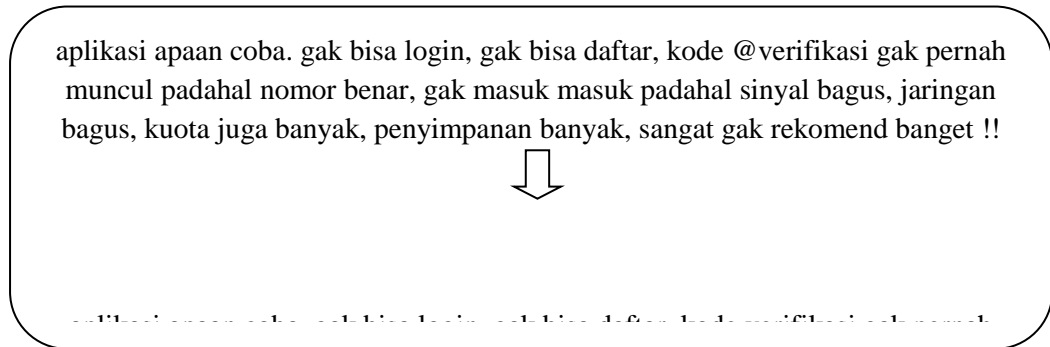


Gambar 3.3 Proses Tokenizing

4. Filtering

Proses penghilangan kata-kata serta tanda-tanda yang tidak bermakna secara signifikan, seperti *hashtag* (#), url, tanda baca tertentu (*emoticon*), dan lainnya (Megawati, 2015). Kata dan tanda baca yang nantinya tidak bernilai atau tidak

berarti akan dieliminasi seperti url, angka, tanda baca, hastag, kata hubung, kata ganti dan lainnya. Pemilihan kata yang bermakna menggunakan algoritma *StopWord* (menghilangkan kata yang kurang penting).



Gambar 3.4 Proses Filtering

3.3 Representasi Model

Setelah data melalui tahap *preprocessing*, data perlu dibuat model agar data yang masih berupa kata-kata data diolah dan dihitung. Akan dilakukan perhitungan skor berdasarkan kamus kata untuk kemudian diberi label kelas sentimen. Perhitungan skor sentimen hanya berdasarkan pada jumlah kata yang terdeteksi.

3.3.1 Perhitungan Skor Sentimen

Berdasarkan *review* pengguna “aplikasi apaan coba. gak bisa login, gak bisa daftar, kode verifikasi gak pernah muncul padahal nomor benar, gak masuk masuk padahal sinyal bagus, jaringan bagus, kuota juga banyak, penyimpanan banyak, sangat gak rekomend banget”. Terdapat 6 kata negatif dan 5 kata positif yang terdeteksi. Kata negatif yang terdeteksi adalah apaan, login, masuk, rekomend, daftar dan muncul. sedangkan lima kata positif yang terdeteksi adalah benar, bagus, jaringan, banyak dan penyimpanan. Rumus perhitungan skor sentimen yang digunakan pada proses ini adalah jumlah kata positif dikurangkan dengan jumlah kata negatif.

$$\text{Skor} = (\text{jumlah kata positif}) - (\text{jumlah kata negatif})$$

Tabel 3.2 Perhitungan Skor Sentimen

Review	Kata Positif	Kata Negatif
aplikasi apaan coba. gak bisa login , gak bisa daftar , kode verifikasi gak pernah muncul padahal nomor benar , gak masuk masuk padahal sinyal bagus, jaringan bagus , kuota juga banyak , penyimpanan banyak, sangat gak rekomend banget	benar, bagus, jaringan, banyak dan penyimp anan	apaan, login, masuk, rekomend daftar dan muncul
Jumlah	5	6
Perhitungan	Skor = 5-6 Skor = -1	

3.3.2 Pelabelan Kelas Sentimen

Analisis sentimen diawali dengan pelabelan data yang dilakukan secara otomatis dengan menghitung skor sentimen. Pada tahapan pelabelan akan dilakukan kedalam dua kelas sentimen yaitu sentimen positif, dan sentimen negatif. Data akan masuk pada kelas positif jika skor yang didapat > 0 , akan masuk pada kelas negatif jika skor yang di dapat ≤ 0 .