

BAB III METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini menggunakan jenis penelitian kuantitatif eksperimental, atau penelitian yang melakukan pengujian secara sistematis, terencana dan diawasi dilihat dengan alur pengujian yang mana dilakukan eksperimen pada implementasi metode klasifikasi SVM menggunakan *RapidMiner* untuk meninjau ulasan pengguna dari segi sentimen positif dan negatif, dan juga dengan meninjau hasil performansi sebagai evaluasi keberhasilan implementasi yang dilakukan, yaitu dengan menghitung performansi menggunakan beberapa indikator, diantaranya seperti *accuracy*, *f-1 Score*, *precision*, dan *recall*.

3.2 Objek Penelitian

Pada penelitian ini, objek penelitian yang digunakan adalah *database Google Play Store*, khususnya pada data *review* pengguna aplikasi *My Pertamina*. Untuk data yang digunakan adalah *review* terbaru pengguna aplikasi *My Pertamina* dalam Bahasa Indonesia yang berjumlah 5.000 data.

3.3 Variabel dan Definisi Operasional

Tabel 3.1 merupakan tabel yang menunjukkan penjelasan dan definisi operasional dari masing-masing variabel penelitian:

Tabel 3. 1
Variabel dan definisi operasional variabel

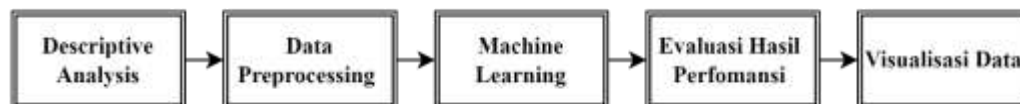
Variabel	Definisi Operasional Variabel
<i>Review</i>	Ulasan pengguna aplikasi <i>My Pertamina</i>
<i>Rating</i>	Tingkat kepuasan pengguna aplikasi <i>My Pertamina</i>
<i>Date</i>	Tanggal saat pengguna memberikan ulasan

3.4 Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan *data primer* yang mana diperoleh melalui *scraping* data dari *database Google Play Store*, khususnya pada data *review* pengguna aplikasi *My Pertamina* yang berjumlah 5.000 ulasan.

3.5 Teknik Analisis Data

Adapun untuk teknik pengolahan analisis data yang digunakan dalam penelitian ini terdiri dari beberapa tahap:



Gambar 3. 1 Teknik analisis data

Dengan penjelasan dari masing-masing bagian adalah sebagai berikut:

1. *Descriptive analysis:*

Descriptive analysis digunakan untuk menunjukkan gambaran umum dari *review* aplikasi *My Pertamina* pada *Google Play Store*.

2. *Data pre-processing:*

Data pre-processing pada penelitian ini yaitu menggunakan *microsoft excel* dan *RapidMiner* untuk melakukan pelabelan data untuk mengkategorikan ulasan pengguna aplikasi masuk ke dalam sentimen positif atau sentimen negatif.

3. *Machine learning:*

Machine learning merupakan metode klasifikasi SVM untuk melakukan *text pre-processing* yang terdiri dari proses *case folding*, *normalization*, *filtering*, *tokenizing*, *stopwords*, dan *stemming*, juga untuk mengklasifikasikan *text review* pengguna aplikasi *My Pertamina*.

4. *Evaluasi hasil perfomansi:*

Evaluasi hasil perfomansi dilakukan dengan mengevaluasi hasil analisis sentimen yang dilakukan pada *software RapidMiner* dengan metode klasifikasi SVM, ditinjau dari perhitungan nilai *accuracy* yang mengacu pada persamaan (1), *precision* yang mengacu pada persamaan (2), *recall* yang mengacu pada persamaan (3), dan *f-1 score* yang mengacu pada persamaan (4).

5. *Visualisasi data:*

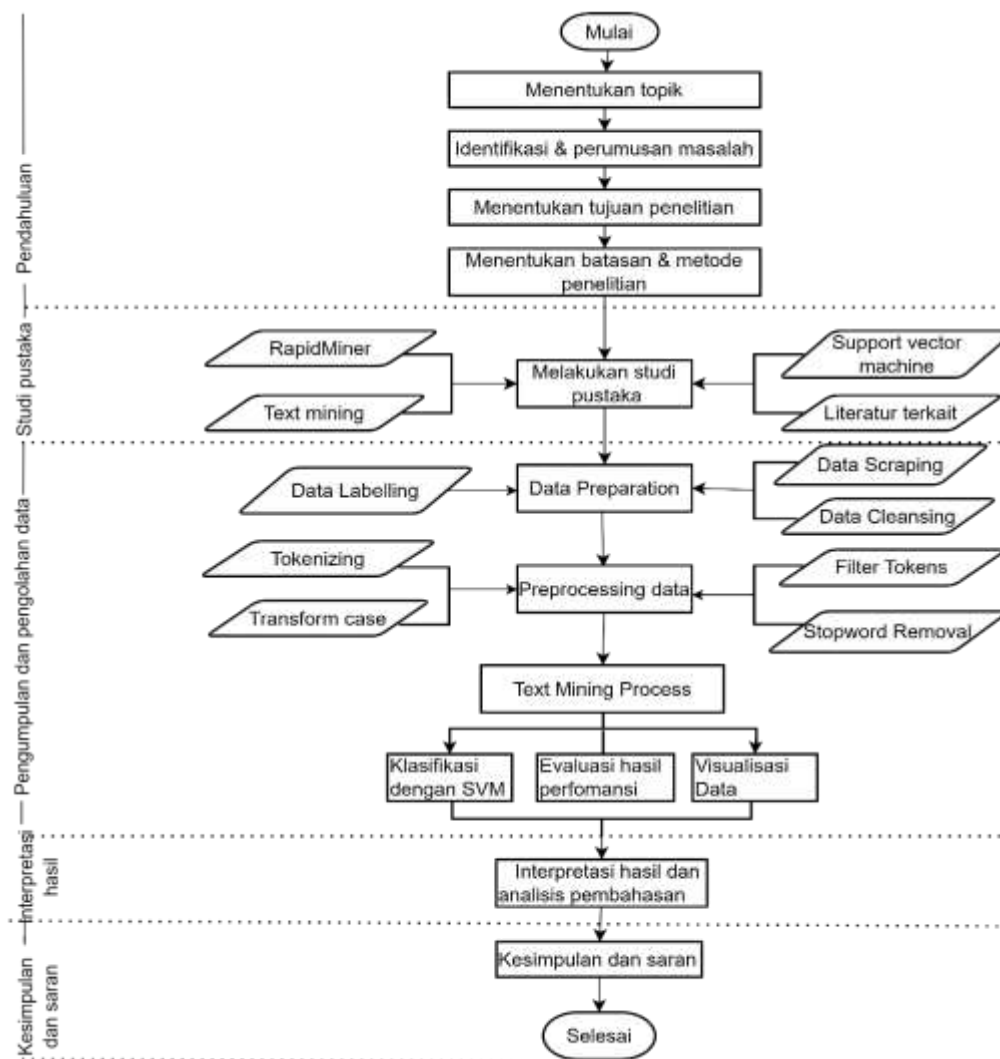
Visualisasi data pada penelitian ini akan menggunakan *word cloud* dan *column chart* untuk menggambarkan hasil analisis sentimen pada sentimen negatif dan sentimen positif.

3.6 Waktu Penelitian

Penelitian ini dilakukan mulai dari bulan Januari sampai bulan Juli tahun 2023 pada aplikasi *My Pertamina* yang terdapat di *Google Play Store*.

3.7 Diagram Alir Penelitian

Alur penelitian yang akan dilakukan pada penelitian ini dapat dilihat pada Gambar 3.2:



Gambar 3. 2 Alur penelitian

Pada Gambar 3.2, tahapan pertama yang dilakukan pada penelitian ini yaitu dengan menentukan topik, dimana topik yang dipilih adalah tentang analisis sentimen. Setelah itu, tahap selanjutnya yaitu mengidentifikasi dan merumuskan masalah. Pada penelitian ini terdapat dua rumusan masalah yang terdapat pada BAB

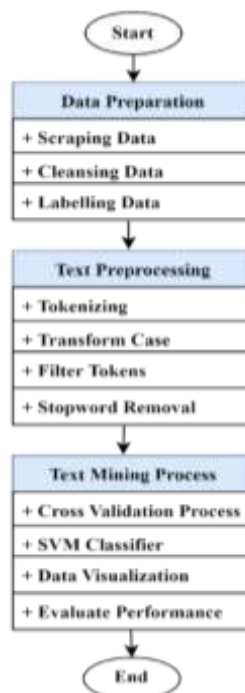
Devi Puspita Dewi, 2023
IMPLEMENTASI METODE KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) MENGGUNAKAN RAPIDMINER UNTUK ANALISIS SENTIMEN ULASAN APLIKASI MY PERTAMINA DI GOOGLE PLAY STORE

I dimana rumusan masalah yang mana menjadi dasar alasan mengapa diambil topik penelitian ini. Kemudian tahap selanjutnya yaitu menentukan tujuan, batasan, dan metode penelitian. Dimana tujuan dari penelitian ini adalah menjawab pertanyaan dari rumusan masalah yang ada, kemudian untuk manfaat penelitian yaitu sebagai harapan peneliti terhadap pemanfaatan dengan adanya penelitian ini.

Selanjutnya untuk batasan penelitian berfungsi agar penelitian ini lebih terkendali dan relevan dengan apa yang ingin dibuktikan. Kemudian setelah itu dilakukan studi pustaka mengenai literatur-literatur yang terkait dengan analisis sentimen dengan meninjau penelitian-penelitian sebelumnya dan membaca literatur yang terkait, dilanjutkan dengan melakukan pengumpulan dan pengolahan data dengan melakukan tahapan *data preparation*, *data preprocessing*, dan *text mining process*. Setelah itu, tahap selanjutnya adalah interpretasi hasil dan analisis pembahasan yang terdiri dari data dan hasil pengujian, hasil implementasi algoritma SVM, hasil visualisasi data dan pembahasan, kemudian dilanjutkan dengan tahap yang terakhir yaitu menentukan kesimpulan dan saran.

3.8 Rancangan Penelitian

Adapun rancangan penelitian yang akan dilakukan pada penelitian ini dapat dilihat pada Gambar 3.3:



Devi Puspita Dewi, 2023

IMPLEMENTASI METODE KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) MENGGUNAKAN RAPIDMINER UNTUK ANALISIS SENTIMEN ULASAN APLIKASI MY PERTAMINA DI GOOGLE PLAY STORE

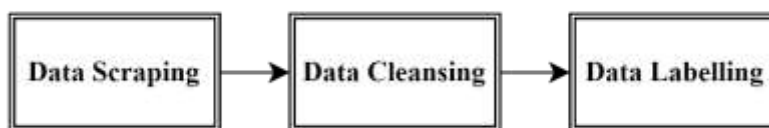
Universitas Pendidikan Indonesia | repository.upi.edu | Perpustakaan.upi.edu

Gambar 3. 3 Rancangan penelitian

Penjelasan dari masing-masing tahapan dalam rancangan penelitian yang ditunjukkan pada Gambar 3.3 adalah sebagai berikut:

3.8.1 *Data Preparation*

Dalam penelitian ini, tahap pertama yang dilakukan adalah mempersiapkan data sebelum data digunakan untuk analisis sentimen, atau biasanya disebut sebagai tahap *data preparation*. *Data preparation* ini berguna untuk mendukung keberhasilan hasil analisis dengan data yang bersih dan siap digunakan. Pada proses ini terbagi menjadi beberapa tahapan yang dijelaskan pada Gambar 3.4:

Gambar 3. 4 *Data preparation*

Penjelasan dari masing-masing bagian pada tahap *data preparation* adalah sebagai berikut:

1. *Data Scraping*

Data scraping atau yang merupakan tahap pengumpulan data dilakukan dengan menarik sejumlah data ulasan aplikasi *My Pertamina* dari *Google Play Store*. Pada penelitian ini, peneliti melakukan *scraping data* dengan jumlah penarikan data sebanyak 5.000 ulasan yang diambil dari bulan Desember 2022 sampai Maret 2023. *Tools* yang digunakan dalam melakukan *scraping data* yaitu menggunakan *Google Colab* dengan bahasa *pyhton*. Untuk *script coding* tertera pada lampiran 1.

2. *Data Cleansing*

Data Cleansing adalah proses untuk menghapus atau memperbaiki data yang rusak, duplikat, salah format, atau tidak lengkap pada suatu himpunan data. Disini peneliti menggunakan *microsoft excel* dalam melakukan *data cleansing* dengan menghapus data ulasan yang tidak beraturan atau yang merupakan data *outlier* pada penelitian ini. Dari hasil *data scraping* yang didapatkan dengan jumlah sebanyak 5.000 ulasan, terdapat beberapa ulasan yang tidak terstruktur

dan tidak diperlukan pada penelitian ini sehingga perlu dihapus agar tidak menjadi penyebab ketidaknormalan data.

3. *Data Labeling*

Tahap selanjutnya setelah dilakukan proses *data cleansing* yaitu *data labeling*. Tahap ini digunakan untuk mengkategorikan ulasan kedalam label positif dan label negatif pada *microsoft excel*. Pada tahap ini, *content* atau ulasan pengguna diidentifikasi dan dikategorikan kedalam dua kelas sentimen, yaitu sentimen positif dan sentimen negatif. Dalam melakukan pemberian label, pelabel harus mengidentifikasi dan mengkategorikan ulasan pengguna ke dalam sentimen positif dan negatif secara obyektif, teliti, dan tepat. Hal ini dikarenakan *output data labeling* ini akan digunakan sebagai *data training* pada model yang dirancang.

3.8.2 *Text Pre-processing*

Setelah tahap *data preparation* selesai, kemudian dilanjutkan dengan tahapan selanjutnya yaitu *text pre-processing*. Tahap *text pre-processing* atau pemrosesan awal teks merupakan tahap yang berfungsi untuk mengubah sebuah data menjadi lebih terstruktur dalam proses *data mining*. Dalam penelitian ini, *text pre-processing* terdiri dari beberapa tahapan diantaranya yaitu:



Gambar 3. 5 *Text Preprocessing*

1. ***Tokenizing***: proses untuk menghilangkan simbol-simbol dan memecah kalimat menjadi kata-kata,
2. ***Transform Case***: proses untuk menyamakan *case* pada suatu dokumen yang biasanya disamakan menjadi huruf kecil,
3. ***Filter Tokens***: membatasi minimal karakter dan maksimal karakter pada kata,
4. ***Stopwords Removal***: menghapus kata-kata yang tidak memiliki arti jika kata tersebut berdiri sendiri.

3.8.3 Text Mining Process

Tahap selanjutnya adalah *text mining process* yang mana terdiri dari beberapa tahapan yang akan dijelaskan pada bagian di bawah ini:



Gambar 3. 6 Text Mining Process

1. Cross Validation Process

Cross validation merupakan *operator* pada *RapidMiner* yang digunakan untuk memvalidasi data. Dengan menggunakan operator ini maka akan diperoleh hasil akurasi yang maksimal. *Cross validation* yang digunakan yaitu merupakan *20-Cross Validation* pada *RapidMiner Studio*. Dalam *model cross validation*, terdapat 4 model untuk pengambilan sampel diantara yaitu, *linear sampling*, *shuffled sampling*, *stratified sampling* dan *simple random sampling (automatic)*.

2. SVM Classifier

Dalam melakukan analisis sentimen, diperlukan suatu *algoritma* yang berfungsi untuk mengklasifikasikan kata. Pada penelitian ini digunakan *algoritma SVM* yang cara kerjanya adalah dengan menerapkan batas kategori *nonlinier* dengan mengubah pemberian ruang contoh menjadi satu yang dapat dipisahkan secara *linier* melalui pemetaan *nonlinier*. tahap *cross validation process* untuk memvalidasi data.

3. Data Visualization

Dalam merepresentasikan suatu data, dibutuhkan suatu proses visualisasi data yang mana biasanya menggunakan elemen visual seperti diagram, grafik, atau peta. Pada penelitian ini, data akan divisualisasikan dalam bentuk *word cloud* dan *column chart*. Dalam *word cloud*, data yang direpresentasikan merupakan kata yang sering muncul dalam sentimen positif dan sentimen negatif. Sedangkan *column chart* digunakan untuk menunjukkan jumlah kata yang

sering muncul, jumlah ulasan per bulan, dan jumlah *prediction label* baik untuk *data training* ataupun *data testing*.

4. *Evaluate Performance*

Dalam suatu analisis sentimen dengan menggunakan algoritma klasifikasi tertentu, setelah dilakukan perancangan model diperlukan suatu evaluasi hasil performansi agar dapat diketahui apakah model yang dirancang sudah memenuhi dan dapat dikategorikan baik. Dalam penelitian ini, terdapat beberapa indikator pengukuran untuk mengevaluasi hasil performansi, diantaranya yaitu *accuracy*, *precision*, *recall*, dan *f-1 score*