

BAB III

METODE PENELITIAN

A. Jenis dan Pendekatan Penelitian

Jenis penelitian yang digunakan adalah studi pustaka (*library research*), yakni studi kepustakaan untuk menemukan kesenjangan teoritis dengan mengkaji literatur untuk dijadikan dasar dalam penelitian. Selanjutnya, pendekatan yang digunakan adalah kuantitatif prediktif atau inferensial. Penelitian kuantitatif prediktif yaitu penelitian yang mengumpulkan dan mengolah data angka. Penelitian prediktif adalah suatu metode penelitian yang bertujuan untuk memprediksi kecenderungan dari fenomena yang ada, yang berlangsung pada saat ini atau yang telah lalu. Penelitian ini tidak menggunakan manipulasi atau pengubahan variabel-variabel bebas, tetapi menggambarkan kondisi apa adanya.⁴³

Oleh karena itu, penelitian ini akan mengubah data teks pada review novel menjadi angka untuk memudahkan komputer dalam membacanya. Kemudian, komputer akan diberi kesempatan melatih dirinya sendiri memahami pola pada review yang bernilai negatif, netral, maupun positif. Penilaian prediktif atau estimasi sentimen pada review pembaca yang dilakukan oleh komputer ini nantinya akan dinilai persentase akurasi.

⁴³ Gidion. *Research Methodology: Penulisan Skripsi, Tesis, & Karya Ilmiah*. Yogyakarta: Mahata, 2019.

Peneliti memilih *Support Vector Machine* (SVM) karena merupakan salah satu metode yang memiliki pemodelan lebih matang dan lebih jelas secara matematis dibandingkan klasifikasi lainnya. Selain itu, sudah banyak juga penelitian terdahulu yang melakukan pembuktian bahwa SVM memiliki nilai akurasi lebih baik dibanding model *supervised learning* (komputer memahami pola data yang diberi label) yang lain sebagai algoritma klasifikasi.

B. Setting Penelitian

1. Lokasi Penelitian

Pengambilan data penelitian ini dilaksanakan di situs *Goodreads.com* untuk pengambilan data, lalu proses pengolahan menggunakan alat bantu seperti Spreadsheet (<https://docs.google.com/spreadsheets/>) dan Google Colaboratory (<https://colab.research.google.com/>).

2. Waktu Penelitian

Penelitian ini dimulai pada bulan Desember sampai Februari 2023 dari tahap persiapan penelitian, pengolahan data, penyusunan hasil dan pembahasan.

C. Populasi dan Sampel

Per-tanggal 09 Februari 2023 novel *The Midnight Library* karya Matt Haig memiliki 139.772 pemberi review novel. Keseluruhan review tersebut dijadikan sebagai populasi penelitian. Lalu peneliti menggunakan Rumus Slovin untuk mencari tahu jumlah minimal sampel yang harus diambil untuk penelitian.

Slovin Formula atau Rumus Slovin merupakan rumus yang digunakan untuk menentukan jumlah sampel atau jumlah data yang dapat dijadikan model pada suatu populasi data. Rumus Slovin digunakan untuk menentukan sampel acak dengan memanfaatkan estimasi suatu ukuran atau populasi sampel.⁴⁴

$$n = \frac{N}{1 + Ne^2}$$

Keterangan:

n : *sample size* / ukuran sampel

N : *population size* / total populasi

e : *margin of error* / toleransi error (dalam satuan persen / atau 1 bagi 100)

Berdasarkan persamaan tersebut maka dapat dihitung estimasi jumlah review untuk dijadikan sebagai sampel untuk dataset penelitian ini, yakni:

$$n = \frac{N}{1 + Ne^2}$$

$$n = \frac{139.772}{1 + (139.772)(2\%)^2}$$

$$n = \frac{139.772}{1 + (139.772)(0,02)^2}$$

$$n = \frac{139.772}{1 + (139.772)(0,0004)}$$

$$n = \frac{139.772}{1 + 55,9088}$$

$$n = \frac{139.772}{56,9088}$$

$$n = 2.456,07$$

⁴⁴ Rizki, Muhammad, Muhammad Arhami, and Huzeni Huzeni. "Perbaikan Algoritma Naive Bayes Classifier Menggunakan Teknik Laplacian Correction." *Jurnal Teknologi* 21, no. 1 (2021): 39-45.

Hasil yang didapatkan dari perhitungan jumlah sampel yang dilakukan menggunakan rumus Slovin dengan jumlah populasi sebesar 139.772 serta tingkat Toleransi Error 0.2 adalah 2.456,07 yang dibulatkan menjadi 2.456 review pembaca novel. Namun, penelitian ini mempertimbangkan bahwa jumlah tersebut merupakan estimasi untuk jumlah sampel minimal dan keinginan peneliti menurunkan persentase *margin of error* sekaligus menaikkan *confidence level* dari data. Selain itu juga agar jumlah review positif, netral, dan negatif dapat seimbang untuk memudahkan pengumpulan dan pengolahan data.

Oleh karena itu, peneliti menentukan bahwa sampel yang digunakan dalam penelitian ini adalah semua review berbahasa Inggris untuk novel *The Midnight Library* yang terdapat di situs *Goodreads* sebanyak 3.000 review yang tersebar secara seimbang, yakni terdiri dari 1.000 review negatif (bintang 1 dan 2), 1.000 review netral (bintang 3), dan 1.000 review positif (bintang 4 dan 5). Seluruh review tersebut merupakan hasil *web scraping* (mengeksrak data dari halaman web menjadi format data terstruktur yang disimpan untuk keperluan analisis).

D. Data, Sumber Data, dan Teknik Pengumpulan Data

1. Data

Data yang digali di penelitian ini adalah dua macam, yaitu data pokok dan data penunjang.

a. Data pokok

- 1) Gambaran umum review pembaca novel *The Midnight Library* pada situs *Goodreads*.

- 2) Hasil penerapan atau implementasi metode *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen data review dari pembaca novel *The Midnight Library* pada situs *Goodreads*.
- 3) Pemahaman tentang bagaimana *text mining* dapat digunakan dalam pengolahan data dan informasi di bidang perpustakaan.

b. Data penunjang

- 1) Data tentang penelitian analisis sentimen seperti video demonstrasi proses *coding*, dan *source code* pada GitHub.
- 2) Data yang didapat melalui karya-karya ilmiah misalnya buku dan jurnal pada penelitian terdahulu.

2. Sumber Data

Sumber data dalam penelitian ini meliputi:

- a. Situs web *Goodreads.com*, yakni review berbahasa Inggris buku *The Midnight Library* karya Matt Haig. Pengambilan data dilakukan dengan Google Chrome *extension* yang bernama, “Instant Data Scraper”.
- b. Dokumentasi, yaitu berkas maupun karya yang memuat data-data atau informasi yang mendukung dalam penelitian ini.

3. Teknik Pengumpulan Data

Teknik pengumpulan data pada penelitian ini adalah *text mining*, yakni mengumpulkan data dengan meng-*install extension data miner*, melakukan *web scraping*, dan menyimpan data sebagai persiapan tahapan analisis sentimen.

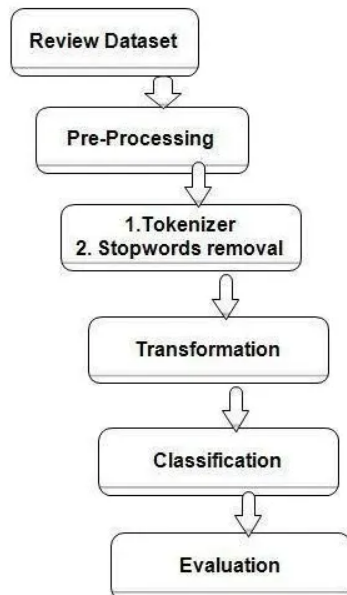
Tabel 3.1 Matrik Data, Sumber Data, dan Teknik Pengumpulan Data

No	Data	Sumber Data	Teknik Pengumpulan Data
1	Data Pokok: a. Gambaran umum review pembaca novel <i>The Midnight Library</i> pada situs <i>Goodreads</i>. b. Hasil penerapan atau implementasi metode <i>Support Vector Machine</i> (SVM) dalam mengklasifikasikan sentimen data review dari pembaca novel <i>The Midnight Library</i> pada situs <i>Goodreads</i>. c. Pemahaman tentang bagaimana <i>text mining</i> dapat digunakan dalam pengolahan data dan informasi di bidang perpustakaan.	Situs <i>Goodreads.com</i> dan hasil pengolahan data (klasifikasi review).	<i>Web scraping</i> dan hasil analisis sentimen.
2	Data Penunjang a. Referensi penelitian mengenai analisis sentimen. b. Karya ilmiah.	Video demonstrasi proses <i>coding</i> , <i>source code</i> pada GitHub, buku, dan jurnal.	dokumentasi

E. Teknik Analisis Data

Analisis data adalah proses mencari dan menyusun secara sistematis data yang diperoleh melalui hasil *web scraping* dan bahan-bahan lainnya, sehingga dapat mudah dipahami, dan temuan dapat di informasikan kepada orang lain. Peneliti akan memberi penjelasan deskriptif terhadap visualisasi data yang disajikan, sehingga dapat tergambar antara hubungan yang satu poin dengan

yang lainnya sebagai analisis sentimen pembaca yang dijabarkan pada gambar desain penelitian berikut ini (tahapan yang dilakukan oleh peneliti).



Gambar 3.1 Desain Penelitian

Penjelasan dari Gambar 3.1 adalah sebagai berikut:

1. *Review Dataset*

Review dataset dilakukan setelah mendapatkan data penelitian yang bersumber dari *Goodreads.com*, yakni review berbahasa Inggris tentang novel *The Midnight Library* karya Matt Haig. Pengambilan data dapat dilakukan dengan memanfaatkan Google Chrome *extension* yang bernama, “Instant Data Scraper”. Data review tersebut disimpan dalam file dengan format CSV (*Comma Separated Values*). File ini dapat dibaca menggunakan perangkat lunak pengolah angka seperti Spreadsheet (<https://docs.google.com/spreadsheets/>).

2. *Preprocessing* Data

Pada tahap ini dilakukan proses *preprocessing* menggunakan bahasa pemrograman Python pada Google Colaboratory (<https://colab.research.google.com/>). Proses yang dapat dilakukan antara lain adalah *tokenizer* (memecah kalimat menjadi kata untuk memudahkan pembobotan kata) dan *stopwords removal* (menghapus kata yang tidak berarti pada review).

3. *Transformation*

Setelah melewati *preprocessing*, teks akan bertransformasi menjadi potongan kata. Ini akan memudahkan pemisahan data review menjadi data latih dan data uji. Hal ini dilakukan untuk memudahkan komputer menemukan pola dari kata-kata yang sering muncul pada review.

4. Klasifikasi

Pada tahap ini, dilakukan proses pengklasifikasian menggunakan metode *Support Vector Machine* karena pada penelitian terdahulu menemukan bahwa metode ini memiliki tingkat akurasi yang baik. Namun, pada penelitian ini juga akan dicoba metode lainnya sebagai pembandingan tingkat akurasi yang dihasilkan.

5. Evaluasi

Evaluasi dilakukan dengan membandingkan persentase akurasi SVM terhadap algoritma klasifikasi lain dan menghitung jumlah review yang berhasil diklasifikasikan secara benar menggunakan rumus akurasi *confusion matrix*. Selain itu, evaluasi juga dapat dilakukan dengan memilih salah satu review secara acak dari dataset, lalu meminta komputer mengidentifikasi apakah review tersebut memiliki sentimen positif, netral, atau negatif.

F. Prosedur penelitian

Dalam penelitian ini ada beberapa tahapan yang dilalui, yaitu :

1. Tahap Pendahuluan
 - a. Melakukan penjajakan awal terhadap novel dengan membacanya.
 - b. Membuat desain proposal skripsi.
 - c. Membuat dan mengajukan proposal skripsi kepada dosen pembimbing untuk dikoreksi dan mohon persetujuan judul skripsi.
2. Tahap Persiapan
 - a. Mengadakan seminar proposal setelah disetujui.
 - b. Memperbaiki hasil proposal berdasarkan hasil seminar dan berkonsultasi dengan dosen pembimbing.
 - c. Menyiapkan pedoman untuk pengambilan dan pengolahan data.
3. Tahap Pelaksanaan
 - a. Meng-*install* Google Chrome *extension* untuk melakukan *web scraping*, dan menyimpan data.
 - b. Mempelajari *coding* yang diperlukan untuk menganalisis data.
 - c. Mengolah dan menganalisis data.
4. Tahap Penyempurnaan Penelitian
 - a. Melakukan penyusunan hasil penelitian dalam bentuk skripsi.
 - b. Berkonsultasi dengan dosen pembimbing mengenai laporan yang telah disusun serta diadakan koreksi dan perbaikan sehingga disetujui.
 - c. Siap dipresentasikan pada sidang munaqasyah untuk disempurnakan.