

BAB III METODOLOGI PENELITIAN

3.1. Metode Pengumpulan Data

Langkah-langkah yang diterapkan dalam penelitian adalah sebagai berikut:

1. Scraping Data dari Web

Untuk mengumpulkan data dalam penelitian ini, dilakukan ekstraksi data secara online dengan memanfaatkan teknik web scraping menggunakan *Google Collab Research*. Teknik web scraping ini dilakukan untuk mengambil informasi dari situs yang diinginkan dengan menggunakan bahasa pemrograman Python.

2. Dataset

Penelitian ini menggunakan sampel ulasan pengguna Aplikasi EdLink berbahasa Indonesia. Fokus penelitian dilakukan pada periode 23 Februari 2017 hingga 26 Oktober 2023. Dalam analisis, sebanyak 2063 ulasan sebelum proses preprocessing, digunakan untuk mengevaluasi sentimen pengguna terhadap aplikasi tersebut dengan menggunakan algoritma Support Vector Machine (SVM). Berikut link halaman dataset diperoleh dari ulasan pengguna aplikasi EdLink :

[https://play.google.com/store/apps/details?id=id.co.sevima.edlink
&hl=en-ID](https://play.google.com/store/apps/details?id=id.co.sevima.edlink&hl=en-ID)

3.2. Metode Pengembangan Perangkat Lunak Prototype

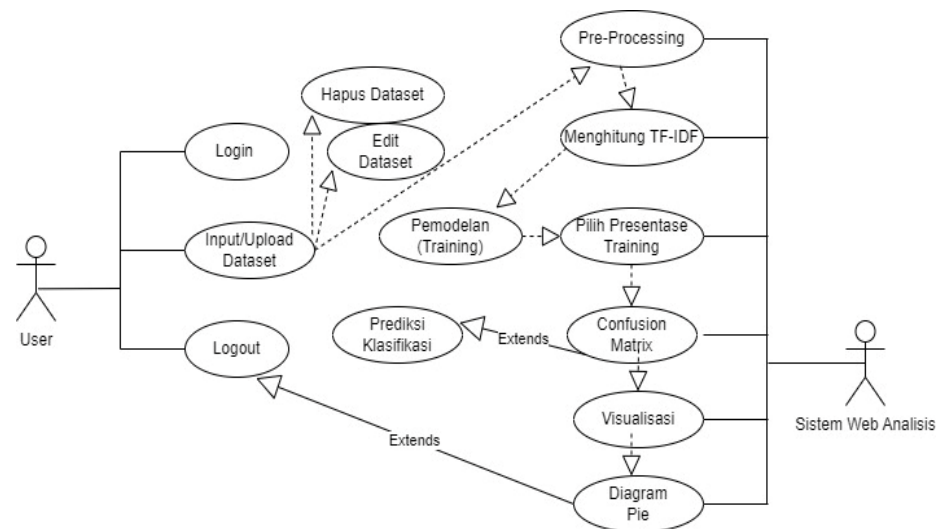
Model pengembangan perangkat lunak yang diterapkan dalam penelitian ini adalah *prototipe*. Secara ringkas, model *prototype* adalah metode pengembangan yang melibatkan pembuatan versi awal atau percontohan dari sistem yang akan dikembangkan. Berikut tahapan yang terdapat pada metode *prototype* :

3.2.1 Quick Plan

Pada tahap desain website untuk analisis sentimen dengan algoritma SVM, dimulai dengan merancang diagram UML. Ini mencakup pembuatan Use Case Diagram, Activity Diagram dan database.

a. Use Case

Use Case Diagram digunakan untuk pemodelan kegiatan pada system analisis sentimen yang akan dibuat. Gambar 3.1 berikut adalah rancang Use Case Diagram yang akan dibuat :

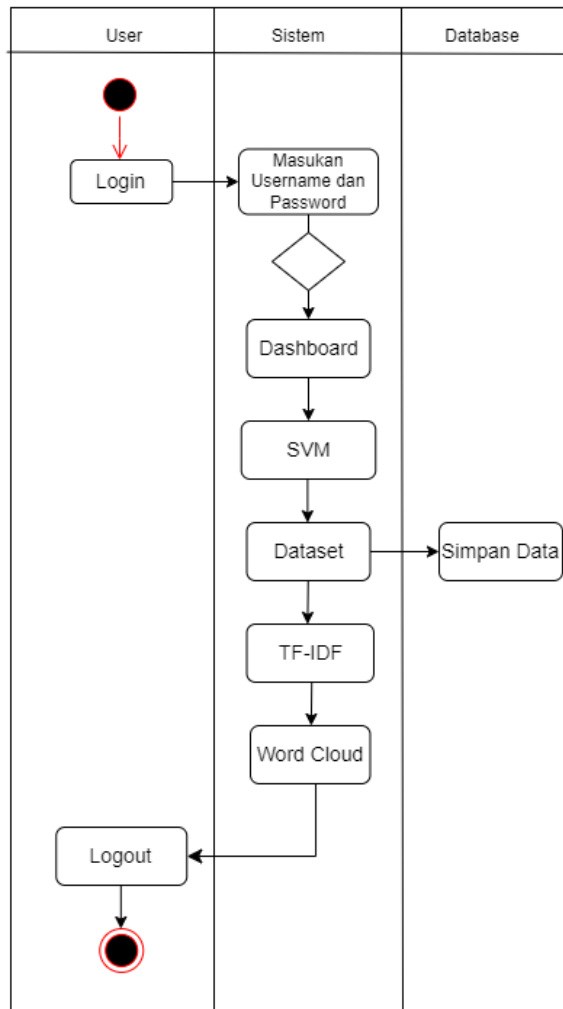


Gambar 3. 1 Use Case Diagram

Dari diagram use case yang terlampir, terlihat bahwa ketika pengguna mengakses aplikasi, proses login diperlukan sebelumnya. Di dalam menu utama, terdapat sejumlah opsi utama yang tersedia.

b. Diagram Activity

Activity Diagram digunakan untuk menggambarkan urutan langkah atau aktivitas dalam suatu sistem, menunjukkan bagaimana proses dimulai, dilakukan, dan diakhiri. Dalam konteks sistem analisis sentimen, Activity Diagram akan menampilkan langkah-langkah seperti pengumpulan data, pra-pemrosesan, ekstraksi fitur, pelatihan model, pengujian, dan evaluasi performa. Dengan diagram ini, alur kerja sistem analisis sentimen dapat dipahami dengan jelas, membantu pengembang dan pengguna dalam memahami proses secara keseluruhan.



Gambar 3. 2 Activity Diagram

c. Database

Berikut adalah struktur database yaitu sebagai berikut :

1. Tabel model_textmining

Nama Database : mlc_svm_tc_dheany

Fungsi : Berfungsi untuk menyimpan data algoritma dan model json.

Tabel 3. 1 Rancangan Struktur Tabel Model_textmining

No.	Field	Type	Length
1	algorithn	Varchar	255
2	model_json	mediumtext	

2. Tabel svm_textclassifier

Nama Database : mlc_svm_tc_dheany

Fungsi : Berfungsi untuk menyimpan id, data ulasan dan label.

Primary Key : dataset_id

Tabel 3. 2 Rancangan Struktur Tabel svm_textclassifier

No.	Field	Type	Length
1	dataset_id (primary key)	Int	11
2	response	Text	
3	label	varchar	30

3.2.2 Construction of Prototype

Pada tahap implementasi dijalankan melalui pengembangan sistem atau pembuatan kode program.

a. Preprocessing (Cleansing dan Case Folding)

Tabel 3. 3 Code Preprocessing

CODE
foreach (\$dataset as \$key) {

<pre> \$txt = mb_strtolower(\$key['response']); \$preprocessed_text = preg_replace('/[^A-Za-z 0-9]/', '', \$txt); echo \$preprocessed_text; } </pre>	
INPUT	OUTPUT
Materi bagi dosen, tugas teman terkirim edlin akun muncul??	materi bagi dosen tugas teman terkirim edlin akun muncul

b. Pembobotan TF-IDF

Tabel 3. 4 Code Pembobotan TF-IDF

CODE
<pre> <?php function calculateTF(\$word, \$document) { \$wordList = explode(" ", strtolower(\$document)); \$wordCount = array_count_values(\$wordList); \$wordFrequency = isset(\$wordCount[strtolower(\$word)]) ? \$wordCount[strtolower(\$word)] : 0; \$totalWords = count(\$wordList); return \$wordFrequency / \$totalWords; } function calculateIDF(\$word, \$documents) { \$totalDocuments = count(\$documents); \$wordOccurrence = 0; foreach (\$documents as \$document) { \$wordList = explode(" ", strtolower(\$document)); if (in_array(strtolower(\$word), \$wordList)) { \$wordOccurrence++; } } } </pre>

<pre> return \$wordOccurrence > 0 ? log(\$totalDocuments / \$wordOccurrence) : 0; \$tf = calculateTF(\$word, \$document); \$idf = calculateIDF(\$word, \$documents); return \$tf * \$idf; } </pre>	
INPUT	OUTPUT
cepat responnya	TF-IDF for document 22: Word: 'cepat', TF-IDF: 2.1769278788536 Word: 'responnya', TF-IDF: 3.6221137578017

c. Training dan Testing

Tabel 3. 5 Code Training dan Testing

CODE
<pre> // Pemilihan persentase data training if (\$this->input->post('train') !== NULL) { \$strain = \$this->input->post('train'); \$countdata = sizeof(\$dataset); \$ndatatraining = (\$strain / 100) * \$countdata; \$ndatatraining = floor(\$ndatatraining); \$newtrainingdata = []; // Pemisahan data training dan testing foreach (\$dataset as \$key) { // ... code untuk pemisahan data ... // Proses testing if (\$x > \$ndatatraining) { // ... code untuk testing SVM ... } } } </pre>

}	
OUTPUT	
Data training	
1084	jelek
372	lumayan bagus
Data testing	
1291	tehubung server
107	nuri bagus bantu ajar

d. Klasifikasi

Tabel 3. 6 Code Klasifikasi

CODE	
<pre><?php if(\$this->input->post('prediksi') !== NULL){ \$getmodel=\$this->db >get_where("model_textmining",array("algoritma"=>"svm"))- >result(); foreach(\$getmodel as \$gm){ \$models = json_decode(\$gm->model_json,true); } \$dt = preprocess_text(array_column(\$dataset,'response')); \$svm_train = \$dt[0]; \$alw = \$dt[1]; \$tokenresp = preprocess_text_feature(\$alw,\$this->input- >post('response'));</pre>	
INPUT	OUTPUT
terimakasih sangat membantu pengumpulan tugas	“Positif”

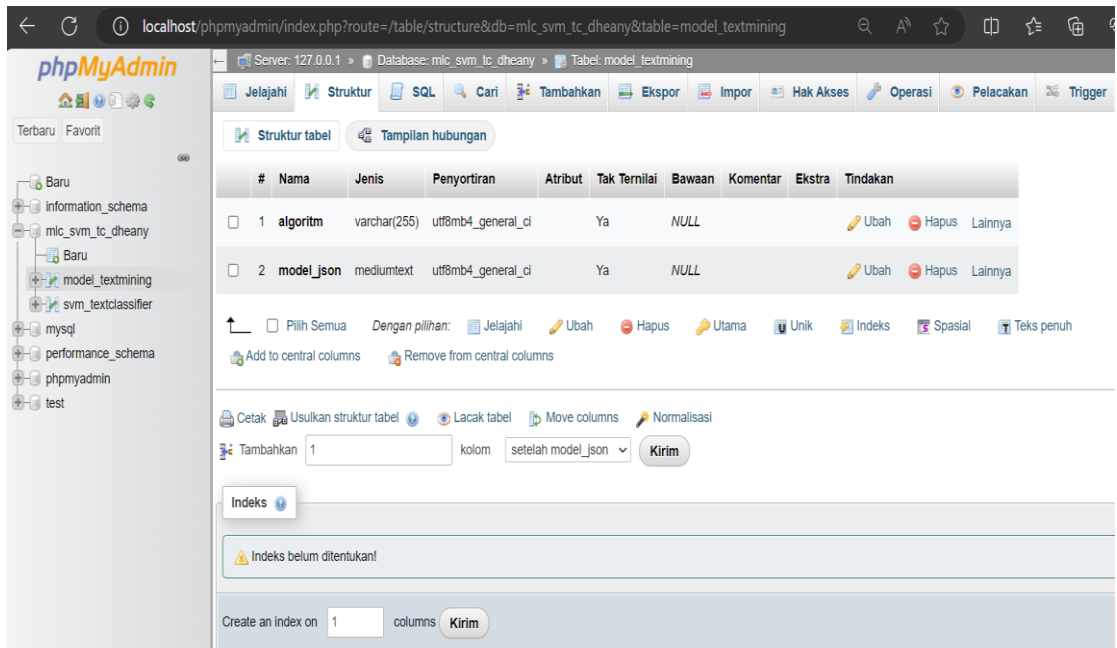
e. Word Cloud

Tabel 3. 7 Code Word Cloud

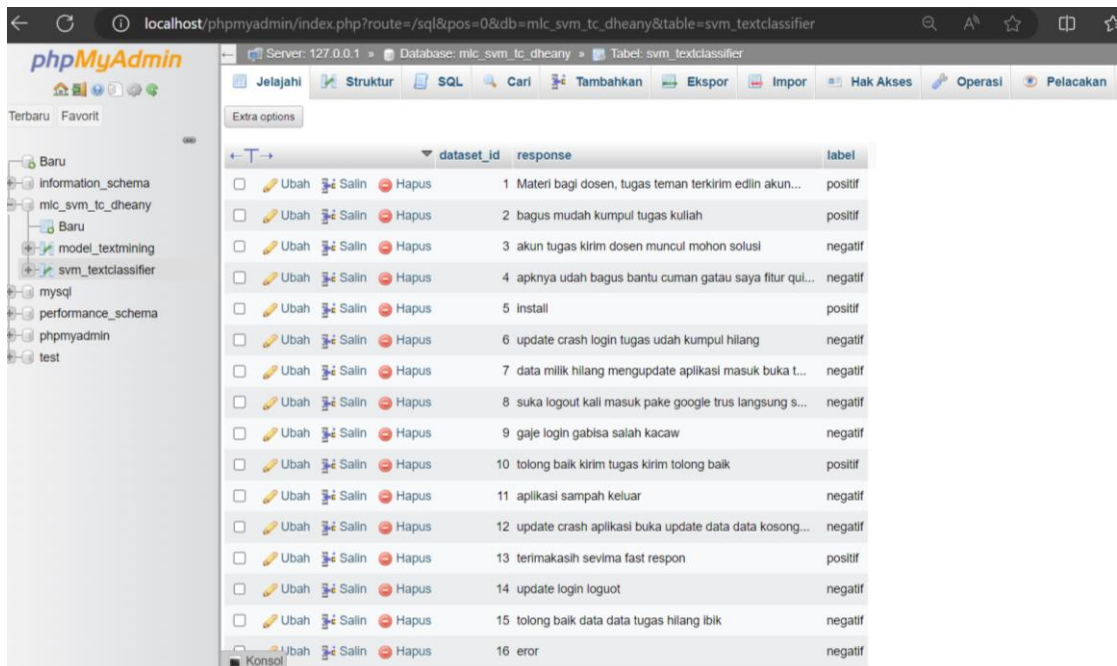
CODE
<pre> \$wordCloudHTML = '<div>'; // Adjust the width as needed \$maxFreq = max(\$wordFreq); foreach (\$wordFreq as \$word => \$freq) { \$fontSize = (\$freq / \$maxFreq) * 150 ; // Adjust the multiplier to control font size based on frequency \$wordCloudHTML .= '' . \$word . ' '; } \$wordCloudHTML .= '</div>'; echo \$wordCloudHTML; </pre>
OUTPUT
<p>aplikasi tugas bagus buka tolong update baik bantu masuk file mohon kuliah ajar mudah udah dosen banget</p>

3.2.3 Dokumentasi

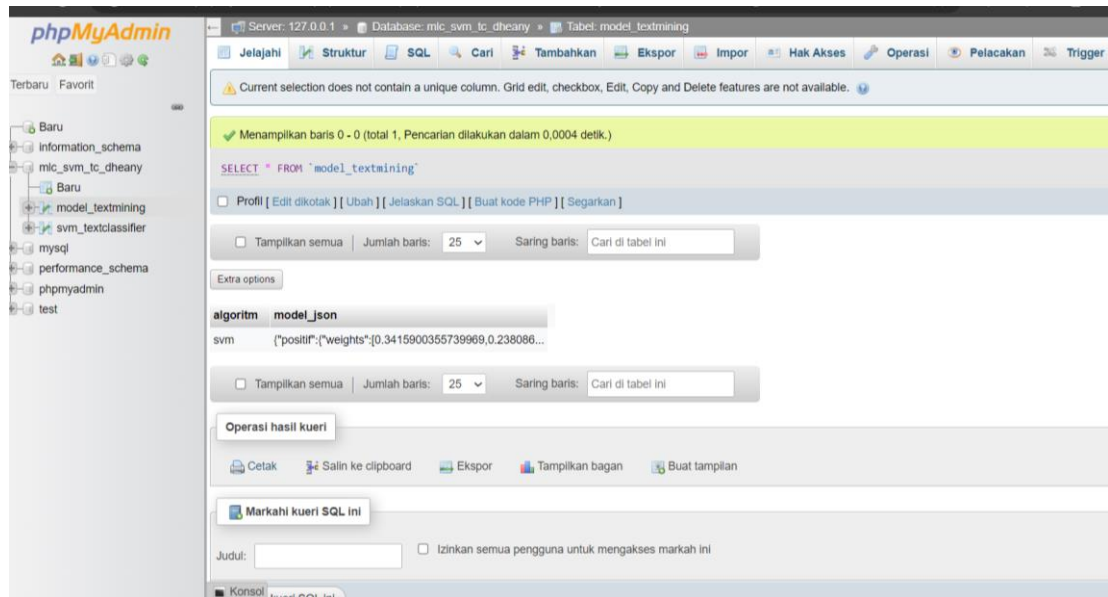
Pada langkah ini, upaya dilakukan untuk secara terperinci menampilkan rincian database.



Gambar 3. 3 Rancangan Database



Gambar 3. 4 Isi Database pada Tabel svm_textclassifier



Gambar 3. 5 Isi Database pada Tabel model_textmining

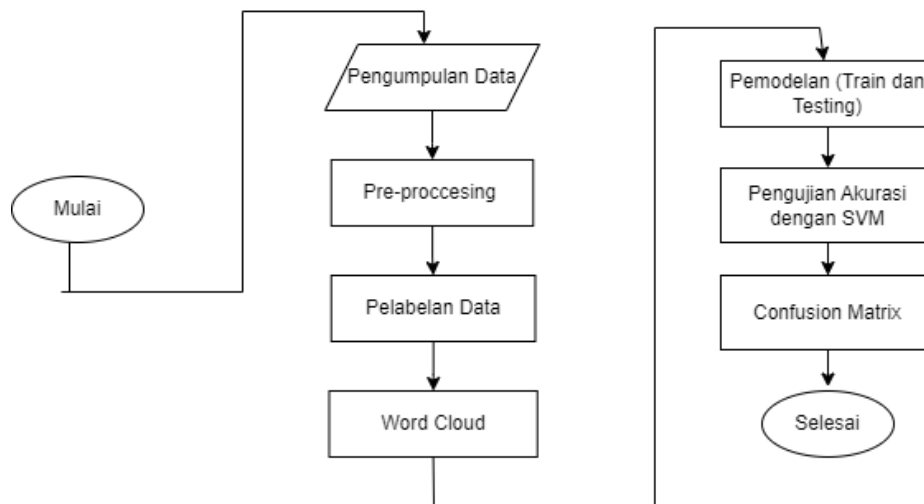
3.3 Variabel Penelitian

Dalam penelitian ini terdapat 2 macam variabel yang digunakan yaitu :

1. *Review*, yaitu isi ulasan pengguna aplikasi EdLink.
2. *Rating*, yaitu penilaian yang diberikan pengguna.

3.4 Alur Penelitian

Alur penelitian adalah konsep atau gambar dari penelitian yang akan dilakukan. Penjabaran alur penelitian tersebut dapat dilihat pada gambar 3.6 dibawah ini :



Gambar 3. 6 Alur Penelitian

Berikut adalah langkah-langkah penelitian yang dilakukan dengan fokus pada aplikasi Edlink:

- a. Mulai
- b. Pengumpulan Data : Data ulasan pengguna Edlink dikumpulkan dengan menggunakan teknik scraping data melalui Google Colab dengan bahasa pemrograman Python.
- c. Preprocessing Data : Data ulasan pengguna diproses dengan langkahlangkah seperti case folding, penghapusan stopwords, tokenisasi dan stemming menggunakan perangkat lunak Google Colab dengan bahasa pemrograman Python.
- d. Pelabelan Data : Pelabelan data ulasan dilakukan dengan membaginya menjadi dua kelas utama, yaitu positif dan negatif. Kriteria untuk memberikan label positif yaitu berdasarkan *rating* 4 dan 5. Kemudian, label negatif diberikan berdasarkan *rating* 1, 2 dan 3. Pelabelan juga dilakukan secara manual, kriteria untuk memberikan label positif adalah adanya kata-kata yang mengekspresikan pujian, dukungan, dan elemen positif lainnya dalam kalimat. Sebaliknya, label negatif diberikan kepada kalimat yang mengandung kata-kata hinaan, cacian, atau ejekan.
- e. Word Cloud : Tahap ini melibatkan visualisasi pada data sentimen positif dan negatif.

- f. Pembobotan Data : Pembobotan data melibatkan transformasi kata-kata dalam sebuah dokumen menjadi representasi numerik. (TF-IDF) dilakukan untuk mengevaluasi sejauh mana suatu kata muncul dalam sebuah dokumen atau dataset.
- g. Pemodelan Training dan Testing : Pada tahap Pemodelan Training dan Testing, tujuannya adalah melatih model menggunakan pembagian dataset menjadi data latih dan data uji. Hal ini bertujuan agar model dapat mempelajari pola dan fitur dari data latih tanpa menghafalnya sepenuhnya, sehingga dapat menggeneralisasi dengan baik saat diuji menggunakan data uji.
- h. Pengujian Akurasi SVM : Dalam konteks pengujian akurasi SVM, ini melibatkan penggunaan data uji yang telah ditandai (label) untuk mengukur seberapa baik model SVM dapat memprediksi kelas atau label data yang belum pernah dilihat sebelumnya. Dengan kata lain, pengujian akurasi memberikan gambaran tentang seberapa baik model SVM dapat mengklasifikasikan data yang baru.
- i. Confusion Matrix : Confusion matrix pada SVM digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan hasil prediksi dengan nilai sebenarnya dari dataset. Ini mencatat true positives, true negatives, false positives, dan false negatives. Dengan confusion matrix, kita dapat menghitung metrik evaluasi seperti akurasi, presisi, recall memberikan pemahaman yang lebih baik tentang kinerja model dalam mengklasifikasikan data.
- j. Selesai.

3.5 Analisis Kebutuhan Perangkat

3.5.1 Perangkat Keras

Perangkat keras yang digunakan dalam penelitian ini adalah sebagai berikut :

- a. Asus LAPTOP-M05P2CFL.
- b. Ram 8GB.
- c. Prosesor i3 Intel(R) Core(TM) i3-1005G1 CPU @ 1.20GHz

3.5.2 Perangkat Lunak

- a. Sistem Operasi Windows 11 Home Single Language (64 bit).
- b. Google Collab Research.
- c. Microsoft Excel.
- d. Visual Studio Code
- e. Xampp