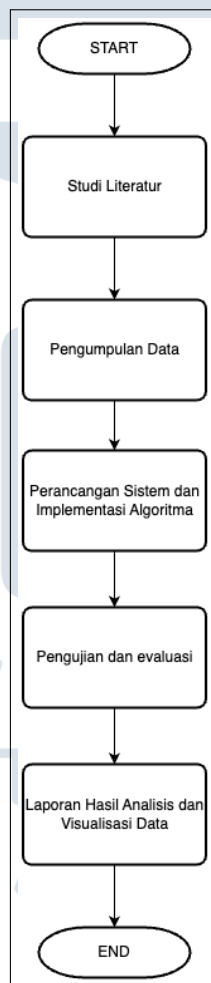


BAB 3 METODOLOGI PENELITIAN

Dalam penelitian yang berjudul "Analisis Sentimen Masyarakat Terhadap Pemandangan Ibu Kota Indonesia Menggunakan Algoritma *Support Vector Machine* dan Seleksi Fitur *Chi Square*", terdapat tahapan penting dilalui untuk merancang dan membangun sistem analisis tersebut, termasuk:

3.1 Alur Penelitian

Langkah-langkah yang dilakukan agar penelitian ini dapat tersentralisasikan adalah sebagai berikut. Gambar 3.1 merupakan gambar diagram alur mengenai langkah dan metodologi dalam melaksanakan penelitian ini.



Gambar 3.1. Diagram alur mengenai langkah-langkah penelitian yang dilakukan.

3.1.1 Studi Literatur

Studi literatur dilakukan untuk memperoleh pemahaman komprehensif terkait penelitian yang dilakukan. Fokus utama studi literatur adalah pada informasi mengenai perkembangan pemindahan Ibu Kota Negara dan dampaknya terhadap masyarakat Indonesia. Selain itu, dilakukan juga pembelajaran mendalam tentang teknik analisis sentimen dan komponen-komponennya, termasuk cara pengambilan data dari platform YouTube, implementasi algoritma *Support Vector Machine*, penerapan seleksi fitur *Chi Square*, serta metode text preprocessing untuk data berbahasa Indonesia. Studi literatur ini menjadi fondasi penting dalam memahami kerangka teoritis dan praktis yang diperlukan untuk penelitian.

3.1.2 Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data yang bersumber dari kolom komentar pada platform YouTube, khususnya pada video-video yang membahas tentang pemindahan Ibu Kota Negara. Pengumpulan data difokuskan pada video-video dengan tingkat engagement tinggi dari kanal-kanal berita resmi dan terpercaya yang telah terverifikasi. Proses pengambilan data menggunakan YouTube API, yang memungkinkan pengumpulan komentar secara sistematis dan terstruktur. Data yang dikumpulkan dibatasi pada komentar dalam Bahasa Indonesia untuk memastikan relevansi dengan konteks penelitian.

3.1.3 Perancangan Sistem dan Implementasi Algoritma

Perancangan sistem dilakukan setelah seluruh data yang diperlukan telah terkumpul. Tahap ini dimulai dengan preprocessing data komentar yang mencakup cleaning, case folding, tokenizing, stemming, dan normalisasi. Selanjutnya, dilakukan implementasi seleksi fitur menggunakan *Chi Square* untuk mengidentifikasi fitur-fitur yang paling relevan dalam analisis sentimen. Sistem kemudian dirancang untuk mengintegrasikan hasil seleksi fitur dengan algoritma *Support Vector Machine* untuk klasifikasi sentimen. Perancangan sistem juga mencakup pembuatan mekanisme evaluasi yang akan digunakan untuk mengukur performa model.

3.1.4 Pengujian dan Evaluasi

Setelah model berhasil dikembangkan, tahap selanjutnya adalah melakukan pengujian dan evaluasi kinerja model. Pengujian dilakukan dengan membandingkan performa algoritma *Support Vector Machine* tanpa seleksi fitur dan dengan seleksi fitur *Chi Square*. Evaluasi kinerja model menggunakan berbagai metrik standar dalam machine learning, termasuk accuracy, precision, recall, dan F1-score. Hasil evaluasi ini akan memberikan gambaran komprehensif tentang efektivitas penggunaan seleksi fitur *Chi Square* dalam meningkatkan performa klasifikasi sentimen.

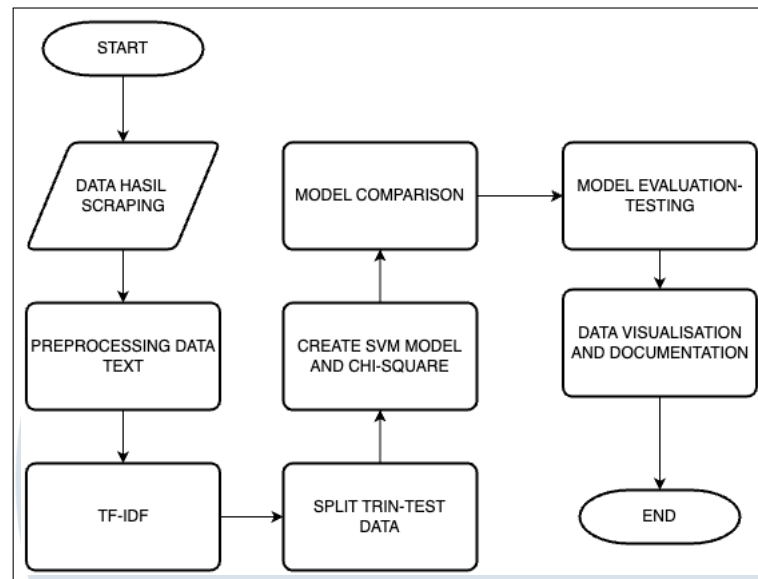
3.1.5 Laporan Hasil Analisis Dan Visualisasi Data

Tahap laporan hasil analisis dilakukan secara sistematis dan menyeluruh, mencakup seluruh aspek penelitian dari awal hingga akhir. Proses ini meliputi pencatatan detailed mengenai metodologi yang digunakan, hasil analisis yang diperoleh, serta temuan-temuan penting selama penelitian. Laporan hasil analisis juga mencakup visualisasi perbandingan performa model dan interpretasi hasilnya. Seluruh informasi ini kemudian disusun menjadi laporan penelitian yang komprehensif, yang dapat memberikan manfaat dan informasi bagi penelitian-penelitian selanjutnya di bidang analisis sentimen.

3.2 Rancangan Sistem

Perancangan sistem pada penelitian ini dapat digambarkan menggunakan *flowchart* yang tertera pada Gambar 3.2.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A



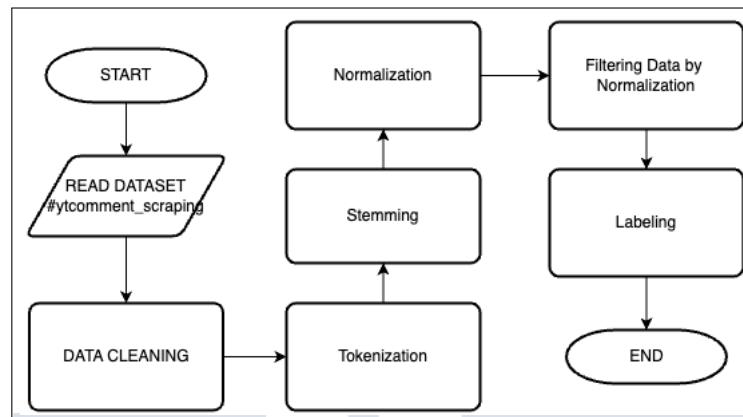
Gambar 3.2. Diagram alur mengenai rancangan sistem yang dilakukan dalam penelitian

3.2.1 Pengumpulan Data

Data diambil dari komentar-komentar yang terdapat pada kolom komentar video mengenai pemindahan Ibu Kota Negara yang diunggah oleh kanal-kanal berita terpercaya seperti Kompas TV, Metro TV, CNN Indonesia, dan CNBC Indonesia. Pengambilan data melibatkan *library* dari Python yang bernama *google-api-python-client* untuk mengakses YouTube API. Data-data yang berhasil dikumpulkan akan ditransformasikan ke bentuk .csv sebagai *dataset* untuk proses analisis selanjutnya.

3.2.2 Text Pre-processing

Pada tahap ini akan dilakukan serangkaian operasi untuk mengolah data sehingga didapatkan data final yang terstruktur dan efisien. Beberapa operasi yang dilakukan dalam *text pre-processing* dapat terlihat pada *flowchart* pada Gambar 3.3.



Gambar 3.3. *flowchart* Tahapan *Text Pre-processing*

1. *Cleaning*, merupakan operasi yang dilakukan dengan menghilangkan semua karakter selain alfabet. 2. *Tokenization*, merupakan operasi yang memisahkan rangkaian kata dalam kalimat menjadi unit kata. 3. *Stemming*, merupakan operasi untuk menemukan kata dasar atau baku pada kata berimbuhan. 4. *Normalization*, merupakan operasi yang digunakan untuk mengembalikan unit kata singkatan dan unit kata yang tidak baku menjadi unit kata baku. 5. *Labeling*, merupakan proses memberikan label positif atau negatif kepada komentar-komentar yang terdapat pada *dataset*. 6. *Stopword Removal*, merupakan operasi yang bertujuan untuk menghapus unit kata yang tidak bermakna seperti kumpulan kata hubung.

3.2.3 *Term Frequency Inverse Document Frequency (TF-IDF)*

Term Frequency Inverse Document Frequency (TF-IDF) diterapkan setelah proses pelabelan data selesai. Proses ini menggunakan kalkulasi statistik untuk memberikan pembobotan pada setiap kata, sehingga dapat diketahui tingkat signifikansi kata tersebut dalam suatu dokumen.

3.2.4 *Train-Test Split Data*

Data yang telah melalui proses seleksi fitur kemudian dibagi menjadi data *training* dan testing. Pembagian dilakukan dengan rasio 80:20, dimana 80% data digunakan untuk melatih model dan 20% sisanya digunakan untuk pengujian. Pembagian data dilakukan secara acak namun tetap mempertahankan proporsi kelas sentimen positif dan negatif.

3.2.5 Penerapan *Chi-Square* Feature Selection

Setelah data melalui tahap preprocessing, langkah selanjutnya adalah melakukan seleksi fitur menggunakan metode *Chi-Square*. Penerapan *Chi-Square* Feature Selection dilakukan untuk mengevaluasi dan memilih fitur-fitur yang paling relevan dalam menganalisis sentimen terkait pemindahan ibukota negara. Proses ini akan memberikan bobot nilai pada setiap kata dalam *dataset* untuk menentukan tingkat kepentingan kata tersebut dalam proses klasifikasi. *Chi-Square* Feature Selection diterapkan menggunakan python *library* sklearn dengan fungsi `SelectKBest` dan `chi2` sebagai metode *scoring*. Proses ini membantu mengurangi dimensi fitur dan meningkatkan efisiensi komputasi dengan tetap mempertahankan informasi yang paling signifikan untuk klasifikasi sentimen.

3.2.6 Penerapan Model *Support Vector Machine*

Model *Support Vector Machine* dikembangkan menggunakan *library* scikit-learn dengan parameter yang telah ditentukan. Implementasi SVM dilakukan dengan menggunakan kernel linear dan parameter C yang akan dioptimasi. Model akan dilatih menggunakan data *training* yang telah melalui proses *preprocessing* dan seleksi fitur *Chi Square*.

3.2.7 Pengujian Model dan Evaluasi

Evaluasi model dilakukan menggunakan confusion matrix untuk mendapatkan nilai *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil evaluasi akan dibandingkan antara model SVM tanpa seleksi fitur dan model SVM dengan seleksi fitur *Chi Square*.

3.2.8 Visualisasi Hasil

Hasil analisis sentimen akan divisualisasikan menggunakan berbagai teknik visualisasi data seperti pie chart untuk distribusi sentimen dan word cloud untuk kata-kata yang sering muncul. Visualisasi ini akan membantu dalam interpretasi hasil dan pemahaman pola sentimen masyarakat terhadap pemindahan ibu kota.