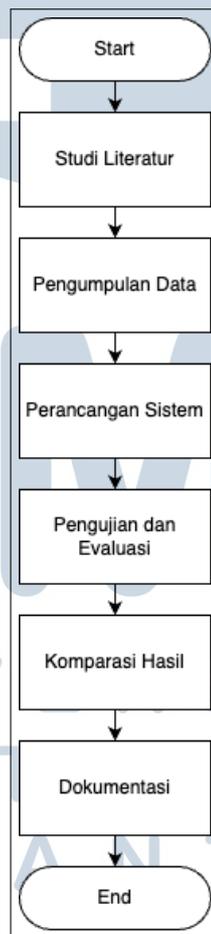


BAB 3 METODOLOGI PENELITIAN

Dalam penelitian bertajuk ”Perbandingan Hasil Analisis Sentimen terhadap Rangkaian Debat Capres-Cawapres Menggunakan Algoritma SVM”, beberapa tahapan penting dilalui untuk merancang dan membangun sistem analisis tersebut, termasuk:

3.1 Alur Penelitian

Langkah-langkah yang dilakukan pada penelitian ini digambarkan menggunakan *flowchart*. Gambar 3.1 merupakan gambar diagram alur mengenai langkah dan metodologi pada penelitian ini.



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

3.1.1 Studi Literatur

Studi literatur dilakukan guna memperoleh informasi terkait penelitian yang dilakukan. Hal ini difokuskan pada informasi terkait rangkaian Debat Capres Cawapres 2024 dan teori-teori yang dapat menambah pemahaman dalam melakukan analisis sentimen seperti cara *data scraping* dari kolom komentar YouTube, mempelajari model *Support Vector Machine*, TF-IDF, klasifikasi, *Text-Processing* untuk memproses data dan *Confusion Matrix* dalam mengukur performa model.

3.1.2 Pengumpulan Data

Pada tahap ini akan dilakukan pengumpulan data yang bersumber dari kolom komentar pada platform Youtube yang memiliki konten video Live Debat Capres Cawapres dengan jumlah *views* tertinggi dari setiap debatannya. Dengan demikian maka untuk debat pertama diwakili oleh video milik Metro TV, debat kedua yang diwakili oleh video milik Official iNews, debat ketiga diwakili dengan video milik tvOneNews, debat keempat kembali diwakili oleh video lain milik Metro TV dan debat kelima yang diwakili oleh kanal Youtube milik Najwa Shihab. Pengumpulan komentar dari Youtube ini akan menggunakan *library* dari Python yang bernama *google-api-python-client*.

3.1.3 Perancangan Sistem

Perancangan sistem dilakukan setelah seluruh data yang diperlukan telah didapatkan. Data yang telah berhasil didapatkan tersebut akan melewati operasi-operasi yang merupakan bagian dari *text-preprocessing*. Dengan dilakukannya hal tersebut maka akan dihasilkan data yang bersih dan mudah dianalisa. Pada data tersebut akan diterapkan pemisahan data *training* dan *testing*. Model akan dilatih menggunakan data *training* guna mempelajari pola-pola dalam data tersebut.

3.1.4 Pengujian dan Evaluasi

Setelah model berhasil dilatih dengan data *training*, selanjutnya tahap pengujian dilakukan menggunakan data *testing* yang telah disiapkan. Evaluasi model akan mengacu pada *Confusion Matrix* yang menghasilkan *precision* yang menjelaskan tingkat ketepatan prediksi positif, *recall* yang menyatakan tingkat

keberhasilan dalam mendeteksi kelas tertentu, *f1-score* yang menggabungkan *precision* dan *recall* dalam satu metrik, dan *accuracy* yang menyatakan tingkat kesesuaian antara prediksi model dan data aktual.

3.1.5 Komparasi Hasil

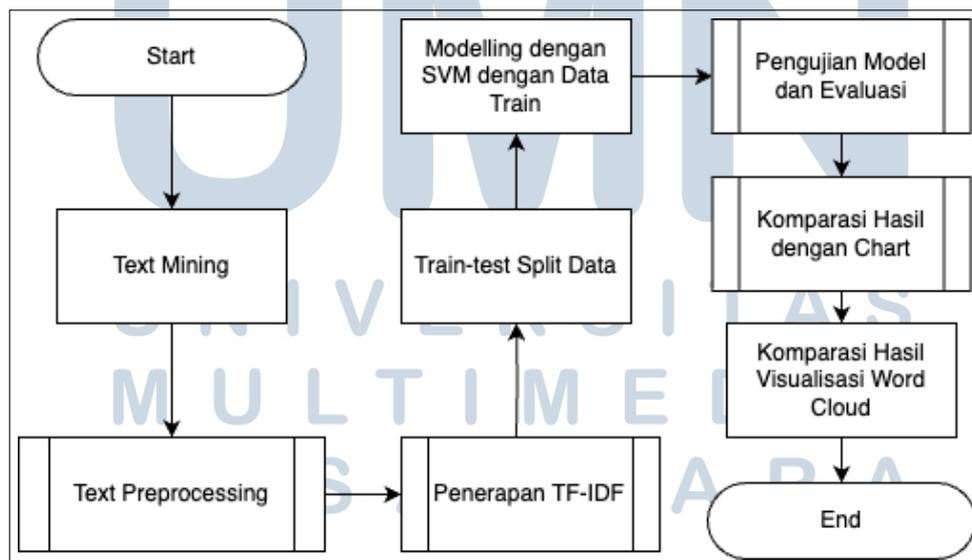
Hasil analisis sentimen yang telah berhasil didapatkan dari masing-masing dataset pengujian akan dikomparasikan satu sama lain. Hal ini bertujuan untuk menemukan dinamika opini masyarakat selama pergelaran Debat yang pertama hingga yang kelima.

3.1.6 Dokumentasi

Mendokumentasikan seluruh penelitian, desain, dan proses penelitian dari awal hingga akhir. Pendokumentasian dilakukan secara bertahap, mulai dari pendahuluan hingga kesimpulan dan saran, untuk memberikan manfaat dan informasi bagi penelitian selanjutnya.

3.2 Rancangan Sistem

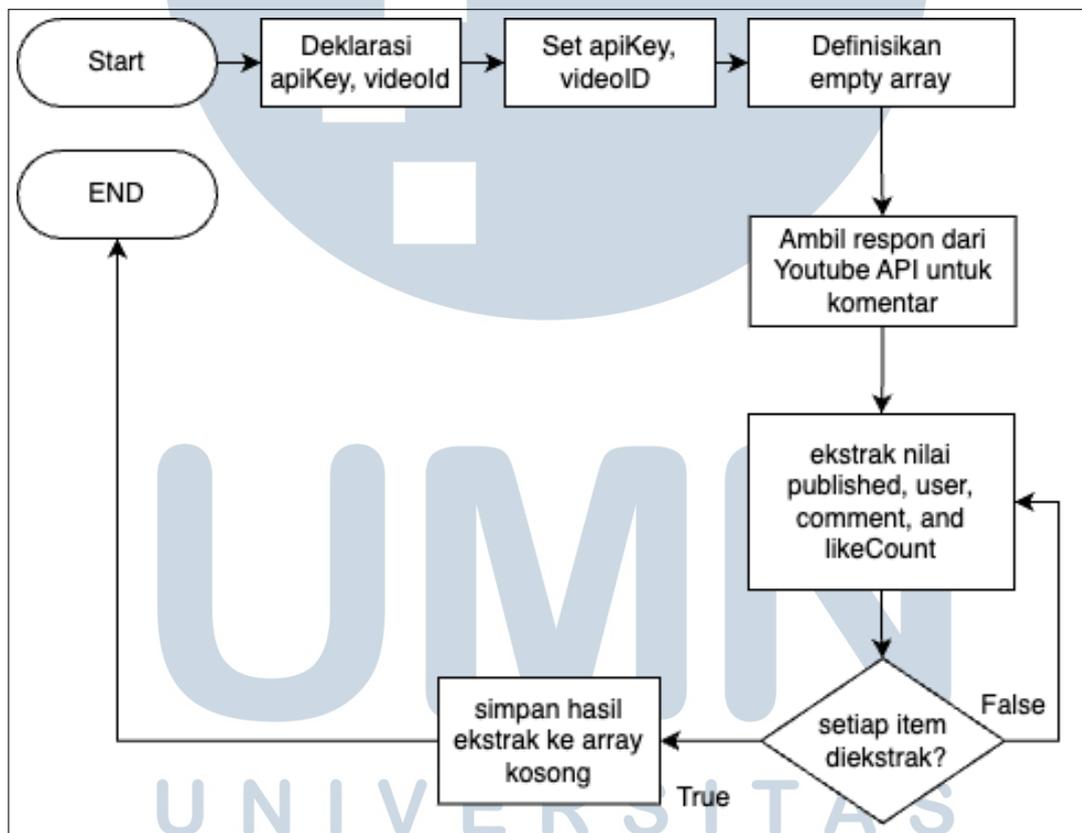
Perancangan sistem pada penelitian ini dapat digambarkan menggunakan *flowchart* yang tertera pada Gambar 3.2.



Gambar 3.2. Diagram alur mengenai langkah-langkah yang dilakukan dalam penelitian

3.2.1 Pengumpulan Data

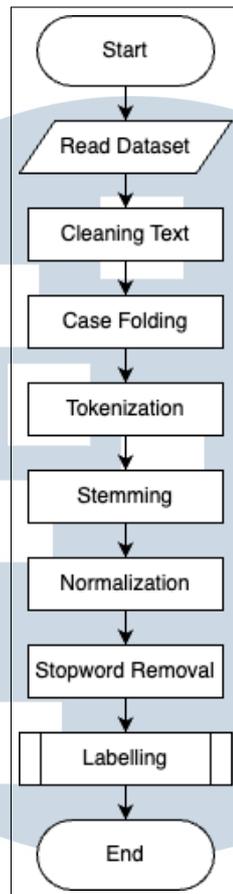
Data diambil dari komentar-komentar yang terdapat pada kolom komentar video mengenai Live Debat Capres-Cawapres dari yang pertama hingga yang kelima yang masing-masing diwakili oleh video milik Metro TV untuk debat pertama dan keempat, Official iNews untuk debat kedua, tvOneNews untuk debat ketiga dan video milik kanal Youtube Najwa Shihab untuk debat yang kelima. Pengambilan data melibatkan *library* dari Python yang bernama *google-api-python-client*. Data-data yang telah berhasil didapatkan akan ditransformasikan ke bentuk *.xlsx* sebagai dataset. Proses secara keseluruhan dalam melakukan pengumpulan data ini dapat dilihat melalui Gambar 3.3 dalam bentuk diagram alur.



Gambar 3.3. Diagram alur mengenai *text-preprocessing*

3.2.2 Text Pre-processing

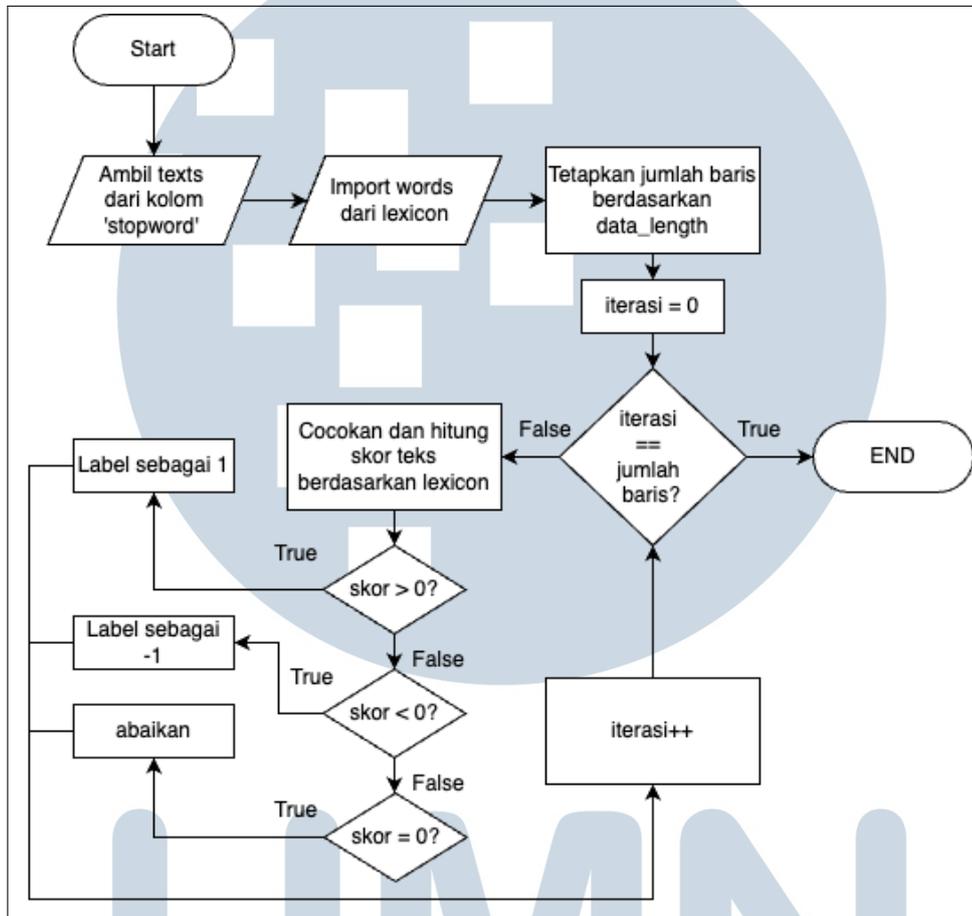
Pada tahap ini akan dilakukan serangkaian operasi untuk mengolah data sehingga didapatkan data final yang terstruktur dan efisien. Beberapa operasi yang dilakukan dalam *text pre-processing* dapat terlihat pada *flowchart* pada Gambar 3.4.



Gambar 3.4. Diagram alur mengenai *text-preprocessing*

1. *Cleaning*, merupakan operasi yang dilakukan dengan menghilangkan semua karakter selain alfabet.
2. *Case folding*, merupakan operasi yang menyeragamkan semua kata supaya tidak menggunakan huruf kapital.
3. *Tokenization*, merupakan operasi yang memisahkan rangkaian kata dalam kalimat menjadi unit kata.
4. *Stemming*, merupakan operasi untuk menemukan kata dasar atau baku pada kata berimbuhan.
5. *Normalization*, merupakan operasi yang digunakan untuk mengembalikan unit kata singkatan dan unit kata yang tidak menjadi unit kata baku.
6. *Stopword Removal*, merupakan operasi yang bertujuan untuk menghapus unit kata yang tidak bermakna seperti kumpulan kata hubung.

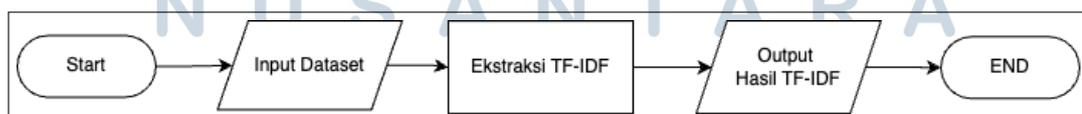
7. *Labeling*, merupakan proses memberikan label positif atau negatif kepada kalimat-kalimat yang terdapat pada dataset. Proses *labeling* secara rinci digambarkan melalui Gambar 3.5.



Gambar 3.5. Diagram alur mengenai proses pelabelan menggunakan *library Lexicon*

3.2.3 Penerapan TF-IDF

Setelah masing-masing data memiliki label, penerapan TF-IDF menjadi langkah selanjutnya. Pada tahap ini, data akan diberikan bobot nilai yang melibatkan perhitungan statistik untuk mengetahui seberapa penting suatu kata dalam dokumen. Proses pembobotan ini dapat dilihat melalui Gambar 3.6.



Gambar 3.6. Diagram alur mengenai penerapan TF-IDF untuk pembobotan

3.2.4 Train-Test Split Data

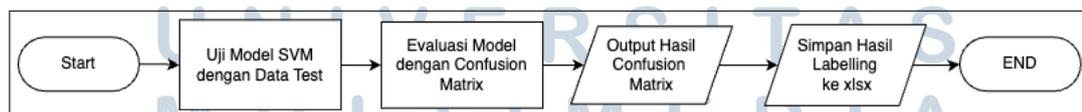
Dalam konteks *machine learning*, *Train-Test Split Data* merupakan teknik untuk memisahkan dataset menjadi data pelatihan dan data pengujian. Sebanyak 80% data akan digunakan sebagai data pelatihan dan 20% sisanya akan digunakan sebagai data pengujian. Tahap ini dilakukan untuk menilai kinerja model yang akan dibuat.

3.2.5 Penerapan Model Support Vector Machine

Dalam penelitian ini, model yang digunakan mengadopsi algoritma *Support Vector Machine* (SVM). Model akan dilatih menggunakan dataset *train* yang telah disiapkan melalui proses *text-preprocessing* dan pembobotan *TF-IDF*. Implementasi model dilakukan dengan menggunakan bahasa pemrograman Python.

3.2.6 Pengujian Model dan Evaluasi

Setelah model dilatih menggunakan dataset *train* yang telah disiapkan melalui proses *text preprocessing* dan pembobotan *TF-IDF*, langkah selanjutnya adalah menguji dan mengevaluasi kinerja model. Evaluasi dilakukan menggunakan data pengujian dari dataset yang sama, kemudian beberapa metrik kinerja dapat dihitung menggunakan *Confusion Matrix* yang memberikan gambaran tentang performa model dalam melakukan klasifikasi, termasuk *accuracy*, *precision*, *recall*, dan *f1-score*. Metrik-metrik ini akan memberikan pemahaman yang lebih baik tentang kinerja model SVM yang sudah dilatih dalam melakukan analisis sentimen pada data pengujian. Implementasi pengujian model dan evaluasi ini dapat terlihat pada Gambar 3.7 yang hingga akhirnya hasil label yang dikeluarkan model akan disimpan ke dalam excel untuk melihat dinamika opini masyarakat.



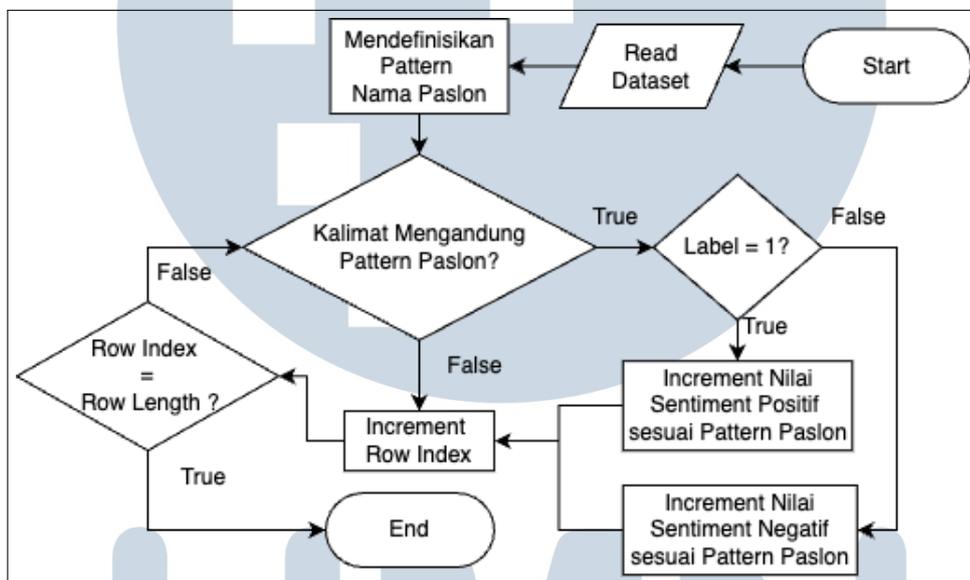
Gambar 3.7. Diagram alur proses evaluasi dan membuat *confusion matrix*

3.2.7 Komparasi Hasil dengan Chart

Tahap komparasi hasil dengan *chart* akan dilakukan ketika model telah mencapai tingkat akurasi yang baik. Pada tahap ini akan dilakukan komparasi total

sentimen positif dan negatif terhadap masing-masing kandidat disetiap debatnya. Hasil perhitungan ini akan ditampilkan melalui *chart*. Penggunaan visualisasi *chart* sangat berguna untuk melihat perubahan dinamika terhadap setiap Capres-Cawapres dalam gelaran debat yang dilakukan.

Gambar 3.8 merupakan alur proses untuk menghitung jumlah sentimen positif ataupun negatif yang tertuju pada masing-masing paslon di suatu dataset yang mewakili satu gelaran debat. Hasil perhitungan ini akan ditampilkan sebagai *line-chart* sebagai parameter dalam menilai dan menyimpulkan dinamika opini dari masyarakat yang terjadi selama gelaran Debat Capres Cawapres 2024.



Gambar 3.8. Diagram alur proses menghitung total sentimen positif dan negatif

3.2.8 Komparasi Hasil Visualisasi Word Cloud

Komparasi hasil juga dilakukan menggunakan *word cloud* untuk menampilkan kata yang paling sering muncul dalam sentimen positif maupun negatif dengan cara penekanan visual yang lebih besar [29]. Visualisasi ini membantu masyarakat untuk memahami tren atau fokus pembahasan yang sedang terjadi pada suatu gelaran debat.