

BAB 3 METODOLOGI PENELITIAN

3.1 Spesifikasi Sistem

Dalam pengerjaan penelitian, diperlukan sistem pendukung yang meliputi *software* dan *hardware* yang memadai. Berikut merupakan rincian dari sistem yang digunakan:

- *Software*:
 - *Operating System*: Windows 10
 - *IDE*: Google Colaboratory
 - *Others*: Google Drive, Google Sheets
- *Hardware*: Spesifikasi sistem yang digunakan untuk penelitian ini adalah sebuah laptop dengan *processor* Intel Core i7-7700HQ yang bergerak di kecepatan 2.80 GHz, memori RAM sebesar 16 GB, dan NVIDIA GeForce GTX 1050 sebagai kartu grafisnya.

Metodologi penelitian menjelaskan metode-metode yang digunakan saat jalannya penelitian.

3.2 Metodologi Penelitian

Dalam proses penelitian implementasi algoritma *Support Vector Machine* atau SVM untuk prediksi hasil pemilu terdapat beberapa proses yang dapat diuraikan sebagai berikut:

1. Studi Literatur

Studi literatur dilaksanakan untuk memahami teori dibalik penelitian yang dijalankan. Dalam penelitian ini dipelajari teori mengenai pemilu, algoritma SVM dan pendukungnya, *sentiment analysis*, *data preprocessing*, TF-IDF, serta teori dibalik prediksi yang dilakukan yang berkaitan dengan *information processing*.

2. Pengumpulan Data

Pengumpulan data menjadi langkah pertama dari penelitian ini. Data dalam penelitian ini diambil dari komentar di video YouTube dengan API yang disediakan oleh Google sendiri. Video yang digunakan merupakan video yang berkaitan dengan pemilu dari tanggal 3 sampai dengan 11 Januari 2024. Komentar disaring berdasarkan kata kunci berupa nama dari masing masing calon presiden.

3. Perancangan sistem

Perancangan ini dapat dilakukan setelah data terkumpul dan tersimpan. Di tahapan ini, data yang sudah terkumpul dilakukan *preprocessing*.

4. Evaluasi model

Dari tahapan sebelumnya didapatkan sebuah model klasifikasi. Model tersebut dievaluasi dengan *classification report* dan juga *confusion matrix*.

5. Prediksi

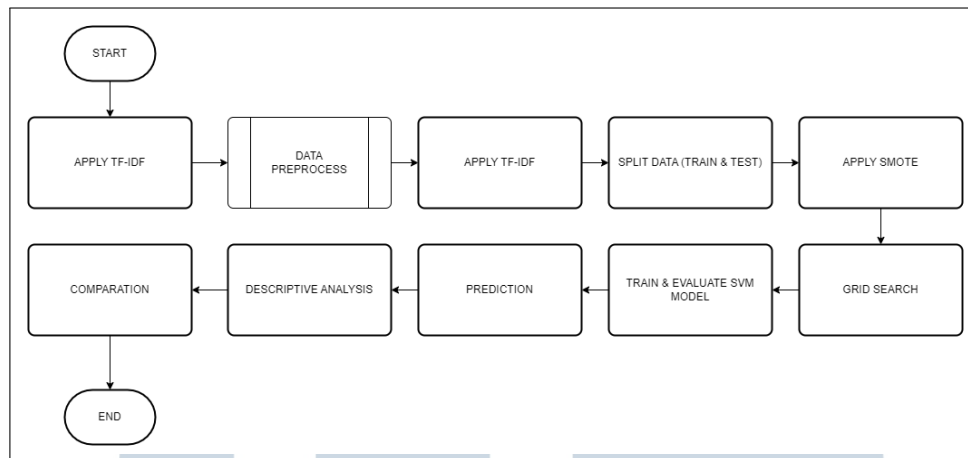
Pada tahap ini, dilakukan tabulasi untuk dilakukannya prediksi. Data yang digunakan adalah data *test* dari model yang sudah dievaluasi.

3.3 Perancangan

Dalam tahap ini, akan digambarkan garis besar dari perancangan dengan *flowchart*. Selain itu, ada pun uraian dari tiap proses.

3.3.1 Garis besar perancangan

Garis besar perancangan merupakan tahapan-tahapan dalam perancangan yang digambarkan secara umum. Tahapan yang dimaksud dimulai dari pengumpulan data hingga prediksi.



Gambar 3.1. Flowchart garis besar perancangan

Gambar 3.1 merupakan *flowchart* dari garis besar perancangan. Perancangan dimulai dari pengumpulan data dan diakhiri dengan perbandingan hasil prediksi dengan survei LSI Denny JA.

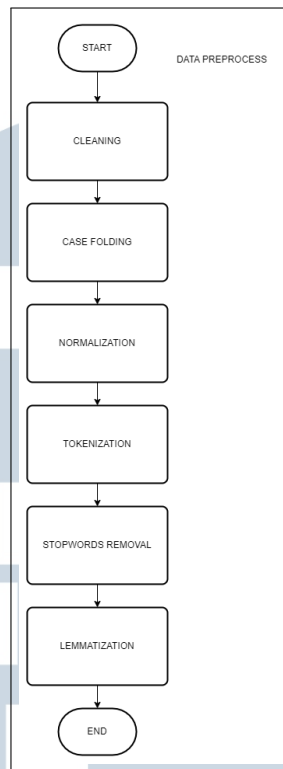
Namun, sebelum masuk ke pengumpulan data, diperlukan untuk melakukan *import* beberapa *library* yang digunakan.

3.3.2 Teknik pengumpulan data

Setelah melakukan *import library* yang digunakan, tahapan selanjutnya adalah pengumpulan data. Data yang digunakan berasal dari kolom komentar di beberapa video YouTube yang diunggah di jangka waktu 3 hingga 11 Januari 2024. Data dikumpulkan dengan cara *crawling* dengan menggunakan API yang disediakan oleh Google. Dari setiap video akan dicari komentar dengan masing-masing *keyword*, lalu data dari setiap video tersebut akan disatukan ke dalam satu *file* per masing-masing *keyword*.

3.3.3 Data preprocessing

Berikut adalah beberapa teknik *data preprocessing* yang digunakan dalam penelitian ini yang digambarkan dalam *flowchart*.



Gambar 3.2. Flowchart data preprocess

Gambar 3.2 merupakan *flowchart* yang menggambarkan proses *preprocessing* yang dilakukan. Proses tersebut meliputi *cleaning*, *case folding*, *normalization*, *tokenization*, *stopwords removal* dan *lemmatization*. Selain proses tersebut, dilakukan juga *labelling*. *Labelling* bertujuan untuk memberi label pada setiap baris data untuk menentukan sentimen. Sentimen yang dimaksud adalah 'positif', 'negatif', ataupun 'netral'. *Labelling* dapat dilakukan dengan *library* yang sudah tersedia, seperti TextBlob dan VADER.

3.3.4 Sentiment Analysis dengan SVM

Setelah proses *data preprocessing*, *sentiment analysis* dapat dilakukan. *Sentiment analysis* akan dilakukan dengan SVM, maka diperlukan juga konversi kata-kata menjadi vektor dengan menggunakan TF-IDF.

A Ekstraksi Fitur

Ekstraksi fitur menggunakan TF-IDF (*Term Frequency - Inverse Document Frequency*) karena *dataset* yang digunakan cukup besar dan TF-IDF dikatakan

dapat tetap efektif di *dataset* yang besar. Tujuannya adalah mengubah kata-kata menjadi sebuah vektor sehingga dapat digunakan pada saat *sentiment analysis* dengan SVM.

B Pembagian data latih

Dari data yang sudah melalui tahap *preprocessing*, akan dibagi menjadi data latih dan juga data uji. Pembagian data ini dilakukan dengan membagi 80% dari data menjadi data latih dan 20% data menjadi data uji.

C SMOTE

Data latih dari hasil pembagian di tahapan sebelumnya akan diterapkan SMOTE. Tujuan penerapan SMOTE ini adalah untuk menyeimbangkan data karena data yang didapatkan tidak seimbang.

D Grid Search

Sebelum pelatihan dan evaluasi model, dilakukan *hyperparameter tuning* dengan *grid search*. *Grid search* dilakukan dengan menggunakan *library* GridSearchCV dari scikit-learn. Parameter yang dicari adalah *C*, *gamma* dan *kernel*. Dilakukan pencarian untuk data sebelum dan sesudah SMOTE untuk menguji lebih dalam.

E Pelatihan model SVM dan Evaluasi model SVM

Nilai vektor dari proses TF-IDF yang telah diterapkan SMOTE akan dimasukkan ke dalam algoritma SVM sebagai *input*. Algoritma akan menghasilkan klasifikasi sentimen positif, negatif atau netral. Dari data yang digunakan, akan digunakan 80% sebagai data latih dan 20% sebagai data uji. Setelah model dibentuk akan dilakukan evaluasi dengan menguji *accuracy*, *precision* dan *recall* dalam *classification report* dan juga *confusion matrix*.

3.3.5 Prediksi Pemilih

Dari data yang ada, didapatkan opini mengenai pasangan calon presiden dari pengguna YouTube. Dapat dilakukan sebuah prediksi pasangan calon presiden

mana yang dipilih oleh masing-masing pengguna YouTube dengan melakukan deduksi berdasarkan perbandingan jumlah komentar positif dan komentar negatif kepada pasangan calon. Deduksi tersebut didasari oleh *heuristic information processing*. Kasus-kasus yang dapat terjadi adalah sebagai berikut.

- Mayoritas komentar dari Pengguna 1 mengarah positif ke calon A, maka diprediksi pengguna akan memilih calon A
- Mayoritas komentar dari Pengguna 2 mengarah negatif ke calon A, namun tidak ada komentar positif tentang calon lain, maka pengguna tersebut diprediksi tidak akan memilih calon A
- Mayoritas komentar dari Pengguna 3 mengarah negatif ke calon A dan ada komentar yang bersifat positif ke arah calon C. Maka, Pengguna 3 diprediksi memilih calon C.
- Komentar dari Pengguna 4 bersifat netral ke calon A, B dan C, maka pengguna tersebut tidak dapat diprediksi pilihannya
- Komentar dari Pengguna 5 bersifat negatif ke calon A, B dan C, maka pengguna tersebut tidak dapat diprediksi pilihannya
- Komentar dari Pengguna 6 bersifat positif ke calon A, B dan C, maka pengguna tersebut tidak dapat diprediksi pilihannya

3.3.6 Perbandingan

Perbandingan yang dilakukan adalah perbandingan antara prediksi yang dilakukan dalam penelitian ini terhadap survei oleh LSI Denny JA. Survei oleh LSI Denny JA tersebut dilakukan periode waktu tertentu, yaitu 3 hingga 11 Januari 2024. Yang dibandingkan adalah prediksi di periode waktu yang sama dengan periode waktu survei tersebut. Asalkan hasil prediksi pada periode waktu yang sama dengan survei menunjukkan konsistensi dalam urutan pemenang dan memiliki kemiripan yang signifikan dengan hasil survei LSI, maka analisis sentimen dapat dianggap sebagai alternatif yang valid untuk survei tradisional. Namun, jika hanya urutan pemenangnya yang sama, prediksi yang diusulkan hanya dapat membantu survei tradisional tanpa menggantikannya.