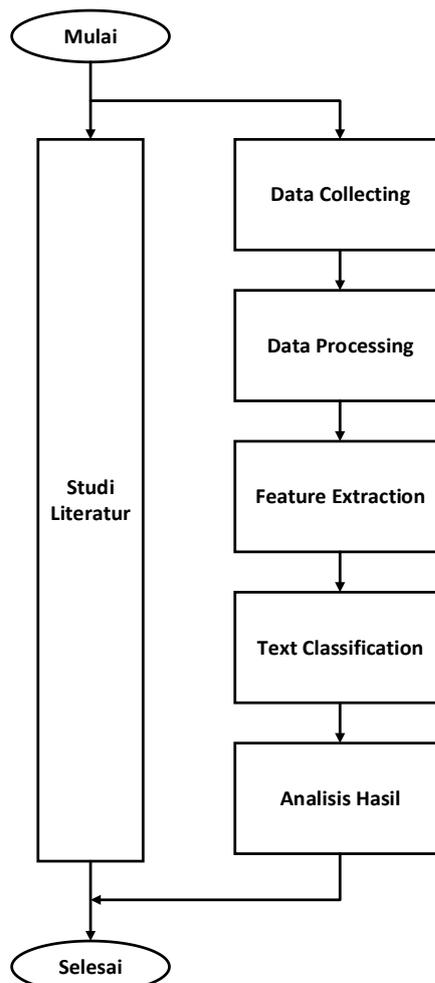


BAB III

METODOLOGI PENELITIAN

3.1 Metode Penelitian

Metode penelitian adalah cara ilmiah untuk mendapatkan data dengan tujuan dan kegunaan tertentu. Terdapat 4 kunci yang harus diperhatikan yaitu cara ilmiah, data, tujuan dan kegunaan. Cara ilmiah berarti kegiatan penelitian itu didasarkan pada ciri-ciri keilmuan yaitu rasional, empiris dan sistematis (Sugiyono, 2013). Gambar 3.1 merupakan metode penelitian yang dilakukan pada penelitian ini.



Gambar 3.1 Metode Penelitian

3.2 Tahapan Penelitian

Dalam penelitian ini terdapat tahapan-tahapan yang dilalui dalam proses penelitian. Tahapan-tahapan tersebut terdiri dari studi literatur, *data collecting*, *data processing*, *feature extraction*, *text classification* dan analisis hasil.

3.3.1 Studi Literatur

Pada Tahap ini mempelajari semua hal yang berkaitan dengan analisis sentimen dan penerapan menggunakan bahasa pemrograman *python* dari berbagai sumber pustaka berupa buku, jurnal, tugas akhir, surat kabar digital serta hasil pencarian pustaka di internet. Studi literatur juga dilakukan selama proses *data collecting*, *data processing*, *feature extraction*, *text classification* dan analisis hasil bertujuan sebagai pedoman dalam melakukan penelitian.

3.3.2 Data Collecting

Data collecting dilakukan dengan menggunakan teknik pengambilan data pada ulasan aplikasi PeduliLindungi di *Google Play Store* yang digunakan pada penelitian ini. Proses pengambilan data menggunakan bahasa pemrograman *python* dengan *library Google Play Scraper* sehingga dapat dengan mudah memperoleh data ulasan aplikasi PeduliLindungi di *Google Play Store*.

3.3.3 Data Processing

Tahap ini adalah proses untuk mengolah data yang didapat dari tahap pengambilan data dengan tujuan untuk melakukan pengontrolan ukuran daftar

kata dan diharapkan dapat meningkatkan performa pada proses klasifikasi yang dilakukan.

A. *Text Preprocessing*

Tahapan *text preprocessing* memiliki 6 proses yang dilakukan, proses tersebut adalah :

1. *Data Cleaning*

Pada tahap dilakukan pembersihan kalimat ulasan dengan menghilangkan tanda baca dari kalimat yang tidak berguna pada proses klasifikasi.

2. *Case Folding*

Pada tahap ini dilakukan proses untuk transformasi setiap karakter yang terdapat pada kalimat ulasan dengan mengubah huruf kapital menjadi huruf kecil.

3. *Spell Checking*

Pada tahap ini dilakukan proses untuk mengubah kata singkatan ataupun *typo* menjadi kata yang sesuai dengan ejaan.

4. *Tokenization*

Pada tahap ini merupakan proses untuk membagi kalimat menjadi tiap kata, dimana tahap ini akan sangat berpengaruh untuk tahap *stop removal*.

5. *Stop Removal*

Pada tahap ini merupakan proses untuk menghapus kata-kata yang terlalu umum dan juga kata yang frekuensi kemunculan katanya cukup banyak tetapi kurang berpengaruh untuk proses klasifikasi.

6. *Stemming*

Pada tahap ini dilakukan proses untuk merubah kata yang terdapat dalam data ulasan ke dalam bentuk dasarnya.

B. *Data Labeling dan Polarity*

Pada tahap ini dilakukan proses untuk *data labeling* dan *polarity* pada data yang telah di proses sebelumnya.

1. *Data Labeling*

Proses ini dilakukan untuk memberikan label sentimen untuk setiap ulasan, pemberian label ini dilakukan secara otomatis dengan menggunakan *library vaderSentiment*. Untuk label sentimen yang digunakan adalah Positif, Netral dan Negatif dengan nilai *score* digunakan sebagai acuan pelabelan, dimana nilai *score* ini diambil dari nilai *compound score*. Ketentuan pelabelan data ulasan adalah jika ulasan memiliki *compound score* $\geq 0,05$ maka ulasan diberi label positif, jika ulasan memiliki *compound score* $> -0,05$ dan $< 0,05$ maka ulasan diberi label netral dan jika ulasan memiliki *compound score* $\leq -0,05$ maka ulasan akan diberi label negatif.

2. *Polarity*

Pada tahap ini dilakukan proses untuk untuk merubah teks menjadi numerik, karena metode *machine learning* tidak dapat melakukan proses *training* dalam bentuk teks secara langsung. Ketentuan *polarity* yang digunakan adalah jika data ulasan berlabel positif maka akan diberikan nilai polaritas 1, jika data ulasan berlabel netral maka akan diberikan nilai polaritas 0 dan jika data ulasan berlabel negatif maka akan diberikan polaritas dengan nilai -1.

3.3.4 *Feature Extraction*

Pada tahap ini dilakukan dua proses yaitu *CountVectorizer* dan TF-IDF. *CountVectorizer* adalah sebuah fitur untuk mengubah teks menjadi representasi *vector* dan juga untuk menghitung frekuensi kata dalam sebuah dokumen. TF-IDF merupakan pendekatan untuk memberi bobot pada jumlah kata dengan ukuran seberapa sering kata tersebut muncul pada sebuah dokumen. Pendekatan ini digunakan untuk mengatasi permasalahan dimana jumlah kata mentah yang mengarah ke fitur terlalu membebani kata-kata yang sering muncul, hal tersebut dapat menyebabkan *suboptimal* pada beberapa algoritma klasifikasi. Untuk penerapan pada penelitian ini memanfaatkan *library Tfidfvectorizer* dengan bahasa pemrograman *python*.

3.3.5 Text Classification

Pada tahap ini dilakukan proses *data training* dan prediksi sentimen, data yang sudah diproses pada tahapan *feature extraction* akan didefinisikan menjadi variabel X dan Y. Kemudian dilakukan proses *K-Folds Cross Validation* untuk melakukan pengujian pada data yang telah diproses sebelumnya, dimana nilai k yang digunakan pada proses *K-Folds Cross Validation* adalah 5 sehingga pengujian akan dilakukan secara 5 kali pengulangan dan data yang diuji dibagi menjadi 5 bagian. Berikut merupakan tahap pemodelan klasifikasi yang dilakukan:

1. Klasifikasi menggunakan *Multinomial Naïve Bayes*
2. Klasifikasi menggunakan *Decision Tree*
3. Klasifikasi menggunakan *Multinomial Naïve Bayes* dengan *AdaBoost*
4. Klasifikasi menggunakan *Decision Tree* dengan *AdaBoost*

3.3.6 Analisis Hasil

Proses terakhir adalah analisis hasil, yaitu proses evaluasi model dan memprediksi data yang telah diklasifikasi menggunakan pemodelan yang digunakan. Pada tahap ini dilakukan perbandingan performa dari metode *Multinomial Naïve Bayes*, *Decision Tree*, *Multinomial Naïve Bayes + AdaBoost* dan *Decision Tree + AdaBoost*. Hasil evaluasi yang dilakukan berupa nilai *accuracy*, *precision*, *recall*, *F1-Score* dan *confusion matrix*. Sehingga dari hasil tersebut dapat dianalisis bahwa hasil keempat metode tersebut apakah memiliki

perbedaan hasil akurasi ataupun tidak dan apakah penerapan *AdaBoost* dapat meningkatkan nilai akurasi pada metode yang digunakan.

Selain itu, dibandingkan juga nilai akurasi berdasarkan nilai *n_estimators* pada metode *Multinomial Naïve Bayes + AdaBoost* dan *Decision Tree + AdaBoost* dengan rentang nilai *n_estimators* yang diujikan adalah 100 – 2500 dengan selisih tiap nilai *n_estimators* adalah 100. Sehingga dapat dilihat pada nilai *n_estimators* berapa metode *Multinomial Naïve Bayes + AdaBoost* dan *Decision Tree + AdaBoost* memiliki nilai akurasi terbaik.