

Hak Cipta Dilindungi Undang-Undang

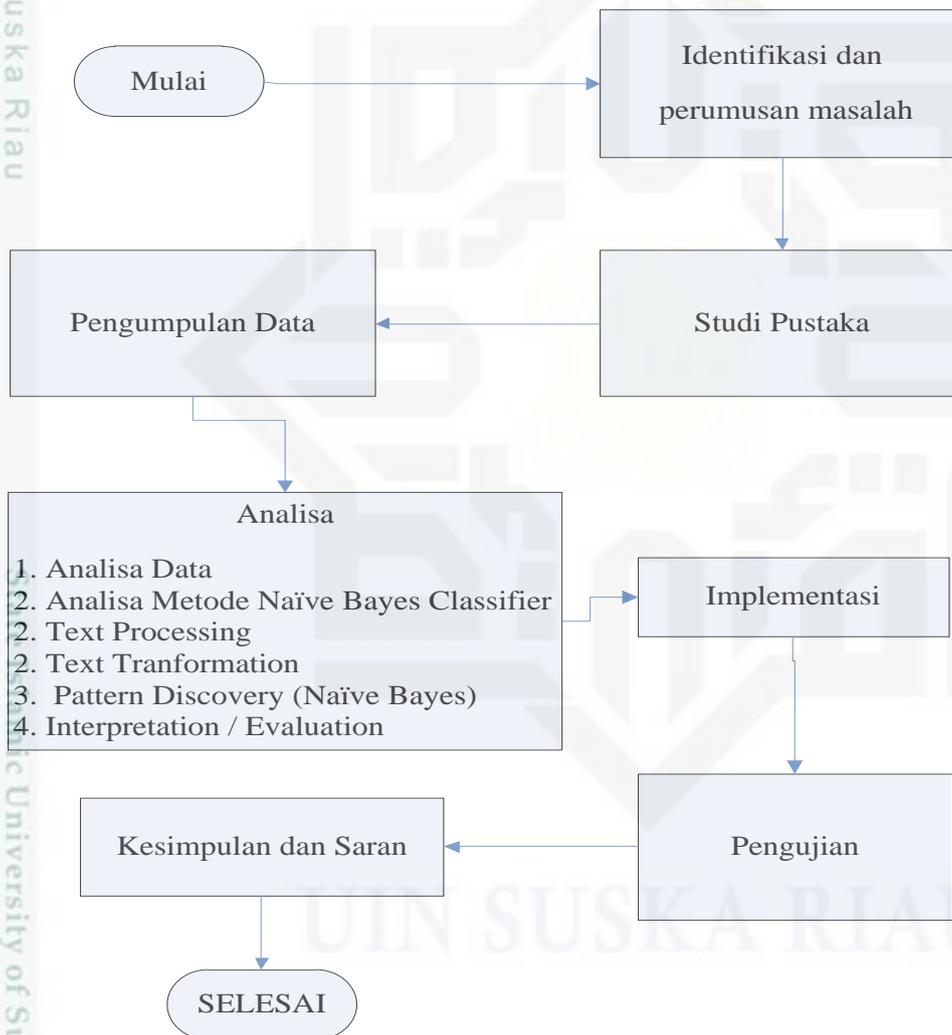
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## BAB III METODOLOGI PENELITIAN

Metodologi penelitian merupakan rangkaian dari langkah-langkah yang diterapkan dalam penelitian, secara umum dan khusus langkah-langkah tertera pada gambar *flowchart* 3.1 :



**Gambar 3.1 Tahapan Penelitian**

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

### 3.1 Identifikasi dan Perumusan Masalah

Tahap ini merupakan langkah awal dari metodologi penelitian dengan mencari dan mempelajari masalah berdasarkan topik yang akan diteliti. Setelah tahap identifikasi masalah, maka dilakukan perumusan masalah dalam menentukan ruang lingkup, latar belakang dan bagaimana mencari solusi dalam masalah tersebut. Berdasarkan latar belakang yang telah dikemukakan dalam penelitian ini, maka dirumuskan permasalahan yaitu bagaimana menerapkan metode Naïve Bayes Classifier untuk megklasifikasi sentimen masyarakat terhadap rokok.

### 3.2 Studi Pustaka

Pada tahap ini melakukan analisa dan mengutip informasi terkait melalui referensi tertentu. Beberapa hal yang dijadikan referensi pada penelitian ini diantaranya adalah jurnal internasional maupun jurnal nasional, buku, artikel dan Skripsi.

### 3.3 Pengumpulan Data

Tahap ini merupakan tahap dimana data yang dibutuhkan dikumpulkan untuk dapat diolah. Adapun sumber data, metode pengumpulan data dan jumlah data adalah sebagai berikut :

- a. Sumber data sentiment terhadap rokok diunduh pada situs *Twitter* menggunakan *Twitter API*.
- b. Dari seluruh data yang telah diunduh, data yang digunakan hanya dalam bentuk kalimat bahasa Indonesia. Dan dilakukan pelabelan manual positif atau negatif pada tiap kalimat.
- c. Data yang digunakan hanya 1000 data *tweet* yang dibagi dalam 500 data *tweet* positif dan 500 data *tweet* negatif.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## 3.1 Analisa

Analisa sistem yang diterapkan pada penelitian adalah menggunakan metode *Text Mining*. Beberapa proses yang terlibat pada tahap analisa menggunakan *Text Mining* diantaranya adalah :

### 3.2.1 Analisa Data

Pada tahapan ini akan dilakukan analisa terhadap data yang digunakan. Pada penelitian ini data yang digunakan adalah data Twitter yang mengandung sentimen terhadap objek penelitian yaitu Rokok. Data *tweet* dikumpulkan menggunakan Twitter API dan data yang digunakan adalah data *tweet* yang mengandung kata kunci rokok, harga rokok naik, *mention* dan *hash tag* (#) Rokok Mahal, #Rokok, #KenaikanHargaRokok, #saverokok, #antirokok. Setelah data *tweet* terkumpul tahap selanjutnya adalah pemberian label pada masing masing *tweet* yang berhasil dikumpulkan.

Dalam penelitian ini klasifikasi sentimen *tweet* terhadap rokok dibagi menjadi dua kelas sentimen yang digunakan yaitu sentimen positif dan sentimen negatif. Adapun beberapa tipe data yang menjadi kriteria pada penelitian ini adalah (Effendy, 2016):

#### 1. Positif

Kalimat disebut positif jika mengandung arti positif terhadap, atau akan memungkinkan menggiring para pembaca membentuk pendapat positif terhadap suatu kalimat tertentu. Seperti informasi yang berguna, kabar yang menyenangkan, pujian, dukungan, prestasi, terobosan terbaru, kejadian yang penting, keberhasilan dan keindahan.

#### 2. Negatif

Kalimat dikategorikan negatif jika mengandung arti negatif terhadap, atau akan dimungkinkan menggiring para pembaca membentuk pendapat negatif terhadap suatu objek tertentu. Berisi tentang kekecewaan masyarakat terhadap

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

rokok yang ditandai dengan adanya sindiran, kata-kata kotor, penolakan, sanggahan, bantahan dan celaan.

Jumlah data yang digunakan 1000 data *tweet*, data terdiri dari 500 data *tweet* berlabel positif dan 500 data *tweet* berlabel negatif. Selanjutnya data dilakukan proses *preprocessing* untuk membuat teks menjadi bersih.

### 3.2.2 Text Processing / Pemrosesan teks

Terdapat enam langkah proses dalam pemrosesan teks seperti dalam gambar 2.1 , yaitu:

1. *Cleaning* yaitu membersihkan setiap *tweet* dari kata-kata yang tidak diperlukan.
2. *Case folding* yaitu proses penyeragaman teks dengan mengubah setiap *tweet* menjadi huruf kecil semua atau huruf besar semua.
3. *Tokenizing* yaitu pemecahan kata pada tiap paragraf atau tiap kalimat.
4. *Spelling normalization* yaitu proses transformasi kata tidak baku pada data *tweet* menjadi kata baku.
5. *Filtering* adalah tahap mengambil kata-kata penting dari hasil token. Biasanya tahap ini menggunakan algoritma *stop-list* (membuang kata-kata kurang penting) atau *word-list* (menyimpan kata penting).
6. Proses *stemming* yaitu memotong imbuhan-imbuhan yang terdapat pada kata yang telah *tokenizing*.

### 3.2.3 Text Transformation

Tahap ini merupakan proses lanjutan dari proses *text preprocessing*, yaitu proses transformasi teks. Proses ini menyeleksi kata-kata yang berkaitan dengan dokumen dengan cara memberi bobot pada kata yang terpilih. Pada penelitian ini metode pembobotan yang digunakan adalah TF. Kata yang telah dipreprocessing hanya dihitung jumlah kata nya saja dikarenakan pada algoritma NBC melakukan klasifikasi dengan menghitung probabilitas kata.berikut. Perhitungannya menggunakan persamaan 2.1.

### 3.2.4 *Pattern Discovery* (Pengenalan Pola)

Tahap ini adalah tahap pengenalan pola pada data yang telah diproses dapat dilihat pada persamaan 2.2, dan berperan mendapatkan informasi dari data tersebut. Pada penelitian ini metode yang digunakan adalah algoritma klasifikasi *naive bayes classifier*.

### 3.2.5 Interpretation & Evaluation

Tahap ini bertujuan memberikan keputusan dari proses apakah data sesuai dengan kebutuhan sistem atau hasil keputusan akan digunakan kembali sebagai data inputan untuk tahap berikutnya.

## 3.2 Implementasi

Tahap ini merupakan tahapan manifestasi dari tahapan-tahapan sebelumnya. Pada tahapan ini algoritma akan diterapkan pada data penelitian yang sesungguhnya.

## 3.3 Pengujian

Pengujian akurasi dilakukan dengan melakukan proses validasi dan evaluasi. Proses validasi menggunakan *k-fold cross-validation*, dengan  $k = 10$  dalam mengklasifikasi sentimen terhadap rokok. dan untuk menghitung tingkat akurasi kebenaran proses klasifikasi digunakan data *confusion matrix*. Dapat dilihat pada persamaan 2.8.

## 3.4 Kesimpulan dan Saran

Bagian kesimpulan merupakan tahap penentuan hasil pengujian yang telah dilakukan. Hal tersebut bertujuan untuk mengetahui apakah penelitian yang dilakukan telah memenuhi kriteria yang sesuai dengan kebutuhan serta dapat dioperasikan dengan baik sehingga dapat bermanfaat. Pada bagian saran berisi kemungkinan pengembangan yang dapat dilakukan terhadap penelitian tersebut