

BAB 3

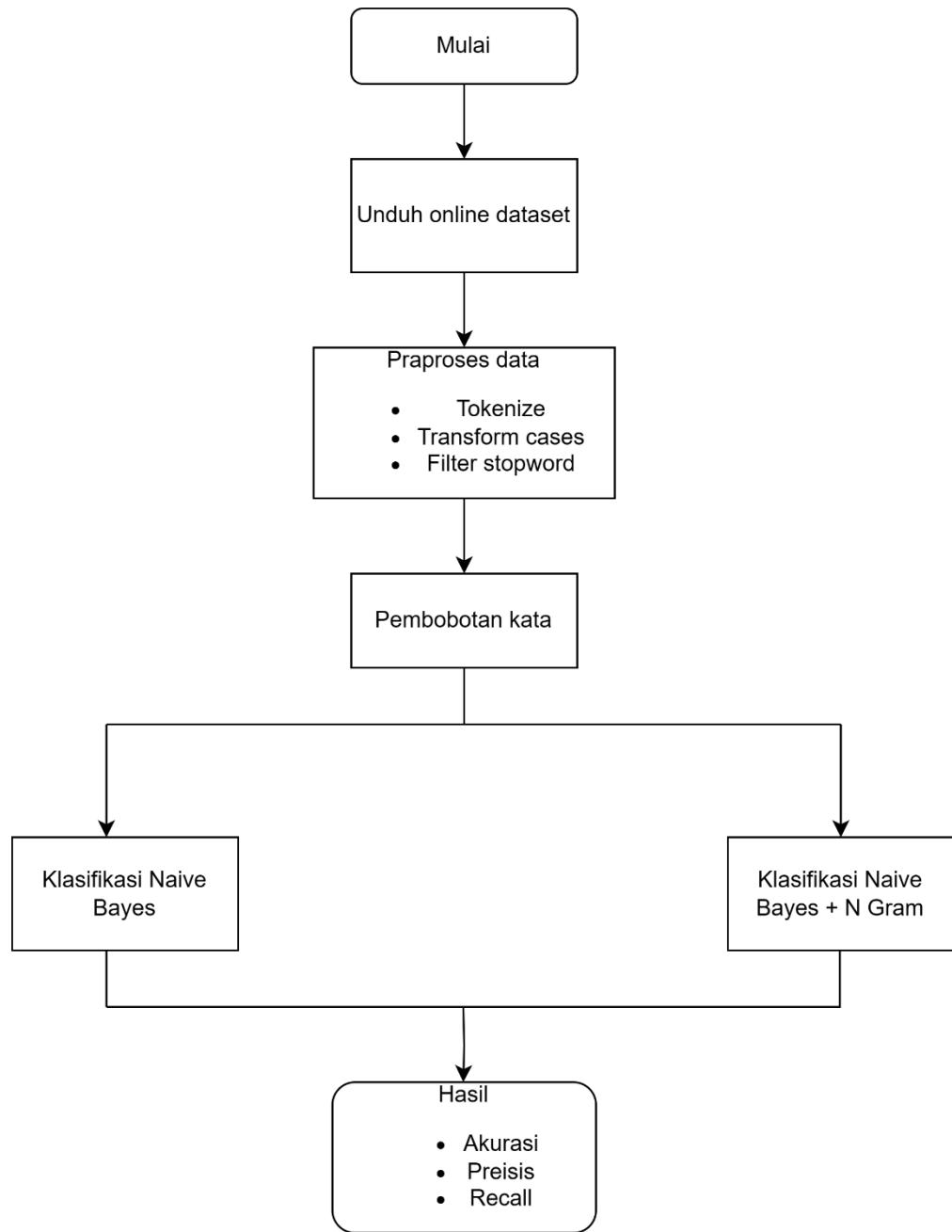
METODOLOGI

3.1 Tahapan Penelitian

Model penelitian yang digunakan dalam penelitian ini adalah model eksperimen dan evaluasi. *Tools* atau alat yang digunakan untuk eksperimen penelitian ini adalah RapidMiner versi 9.10. Tahapan penelitian ini adalah:

1. Metode yang diusulkan
2. Pengumpulan Data
3. Praproses Data
4. Pembobotan Kata
5. Pengujian algoritma *Naïve Bayes*
6. Evaluasi dan Validasi.

3.2 Metode yang diusulkan



Gambar 3.1 Metode yang diusulkan

Metode yang diusulkan untuk menganalisis sentimen publik perilaku antisosial menggunakan algoritma *Naïve Bayes* dengan fitur *n-gram* di

evaluasi *performance* atau yang biasa disebut *confusion matrix* yang akan menghasilkan akurasi, presisi, dan *recall* di validasi menggunakan *k-fold validation*.

3.3 Pengumpulan Dataset

Hasil unduh dataset publik dari *Antisocial Behaviour Public Twitter Indonesia / Kaggle* yang diunggah oleh Fitri Andri Astuti berjumlah 1.251 data.

The screenshot shows a Microsoft Excel spreadsheet titled "Non-Antisocial / Umum". The data is organized into two columns: "text" and "classname". The "text" column contains various tweets, and the "classname" column indicates the category for each tweet. The categories are color-coded: Non-Antisocial (light blue), Umum (light green), Agresivitas (light orange), Iritabilitas (light purple), and Kegagalan (light red). The first few rows of data are as follows:

text	classname
1 @kskoopill kegagalan untuk menyesuaikan diri dengan norma-norma sosial tentang perilaku yang sah	Non-Antisocial
2 Heran bar kegagalan untuk menyesuaikan diri dengan norma-norma sosial tentang perilaku yang sah	Non-Antisocial
3 @aryasyad kegagalan untuk menyesuaikan diri dengan norma-norma sosial tentang perilaku yang sah	Non-Antisocial
4 @im_mec kegagalan untuk menyesuaikan diri dengan norma-norma sosial tentang perilaku yang sah	Non-Antisocial
5 @_AnakKi kegagalan untuk menyesuaikan diri dengan norma-norma sosial tentang perilaku yang sah	Non-Antisocial
6 @_Thinks_Non-Antisocial / Umum	Umum
7 @jaenikill Non-Antisocial / Umum	Umum
8 Benar ben Non-Antisocial / Umum	Umum
10 Aku selalu Non-Antisocial / Umum	Umum
11 @iravant Non-Antisocial / Umum	Umum
12 Mungkin Non-Antisocial / Umum	Umum
13 laper sial iritabilitas dan Agresivitas	Agresivitas
14 @jaenikill kegagalan untuk menyesuaikan diri dengan norma-norma sosial tentang perilaku yang sah	Non-Antisocial
15 @_suraj Non-Antisocial / Umum	Umum
16 @_mmeill Non-Antisocial / Umum	Umum
17 @adlesyr kegagalan untuk menyesuaikan diri dengan norma-norma sosial tentang perilaku yang sah	Non-Antisocial
18 @jutsmee kegagalan untuk menyesuaikan diri dengan norma-norma sosial tentang perilaku yang sah	Non-Antisocial
19 @_viveetn Non-Antisocial / Umum	Umum
20 @dusrimi Non-Antisocial / Umum	Umum

Gambar 3.2 Dataset perilaku antisosial

Gambar 3.2 merupakan dataset perilaku antisosial. Dataset tersebut memiliki 5 nama kelas yaitu: 1) Kegagalan untuk menyesuaikan diri dengan norma-norma sosial tentang *perilaku* yang sah, 2) Iritabilitas dan agresivitas, 3) Pengabaian yang gegabah untuk keamanan, 4) Kurangnya penyesalan, 5) Non-antisosial atau Umum.

3.4 Praproses Data

Pada tahap *praproses* data yang dilakukan dalam penelitian ini ada 3 langkah, yaitu:

1) Tokenize

Pada tahapan ini kalimat pada *tweet* akan diperiksa secara menyeluruh.

Setelah itu dilakukan tahap *tokenize*, *tokenize* adalah tahap pemisahan kata, simbol, frasa, dan entitas penting lainnya (disebut token) dari teks (Rahman, 2020). Contoh *tokenize*:

Tabel 3.1 Tokenize

<i>Input</i>	<i>Output</i>
Ada rasa senang yg sampe bikin ga bisa tidur? Ada. Ada rasa sedih yg bikin ga bisa tidur? Ada. Keduanya bisa terjadi karena dikasihi dan mengasihi. Dan keduanya bisa terjadi dalam waktu yg bersamaan? Bisa!	['Ada', 'rasa', 'senang', 'yg', 'sampe', 'bikin', 'ga', 'bisa', 'tidur?', 'Ada.', 'Ada', 'rasa', 'sedih', 'yg', 'bikin', 'ga', 'bisa', 'tidur?', 'Ada.', 'Keduanya', 'bisa', 'terjadi', 'karena', 'dikasihi', 'dan', 'mengasihi.', 'Dan', 'keduanya', 'bisa', 'terjadi', 'dalam', 'waktu', 'yg', 'bersamaan?', 'Bisa!']

2) Transform Cases

Transform Cases proses mengubah seluruh karakter pada data sesuai keinginan, seperti mengubah huruf besar menjadi huruf kecil dan

sebaliknya (Rahman, 2020). Contoh *transform cases* untuk *lowercase*(tanpa huruf kapital):

Tabel 3.2 Transform cases

<i>Input</i>	<i>Output</i>
Ada rasa senang yg sampe bikin ga bisa tidur? Ada. Ada rasa sedih yg bikin ga bisa tidur? Ada. Keduanya bisa terjadi karena dikasihi dan mengasihi. Dan keduanya bisa terjadi dalam waktu yg bersamaan? Bisa!	ada rasa senang yg sampe bikin ga bisa tidur? ada. ada rasa sedih yg bikin ga bisa tidur? ada. keduanya bisa terjadi karena dikasihi dan mengasihi. dan keduanya bisa terjadi dalam waktu yg bersamaan? bisa!

3) *Filter Stopword*

Setelah dilakukan tahap *transform cases* dan *tokenize* tahap selanjutnya adalah *filter stopword*, pada tahap ini, semua kata diperiksa, jika dalam dokumen terdapat kata-kata yang tidak memberikan kontribusi banyak seperti kata sambung, kata depan, atau kata ganti, maka kata-kata tersebut akan dihilangkan (Maulana et al., 2020). Contoh *filter stopword*:

Tabel 3.3 Filter stopword

<i>Input</i>	<i>Output</i>
Ada rasa senang yg sampe bikin ga bisa tidur? Ada. Ada rasa sedih yg bikin ga bisa tidur? Ada. Keduanya bisa terjadi karena dikasihi dan mengasihi. Dan keduanya bisa terjadi dalam waktu yg bersamaan? Bisa!	['rasa', 'senang', 'yg', 'sampe', 'bikin', 'ga', 'tidur', 'rasa', 'sedih', 'yg', 'bikin', 'ga', 'tidur', 'ada.', 'keduanya', 'terjadi', 'dikasihi', 'mengasihi.', 'keduanya', 'terjadi', 'waktu', 'yg', 'bersamaan']

3.4.1 Ekstrasi Fitur n-gram

Setelah melalui tahap praproses data(*tokenize, transform data, dan filter stopword*) akan dilakukan proses *n-gram* pada dokumen teks. Pada penelitian ini akan menggunakan *n-gram*, n=1 atau disebut dengan *unigram* dan n=2 atau biasa disebut *bigram*. Contoh *n-gram*:

Tabel 3.4 N-gram

<i>Unigram</i>	<i>Bigram</i>
Ada rasa senang yg sampe bikin ga bisa tidur? Ada. Ada rasa sedih yg bikin ga bisa tidur? Ada. Keduanya bisa terjadi karena dikasihi dan mengasihi. Dan keduanya bisa terjadi dalam waktu yg bersamaan? Bisa!	“Ada rasa”, “rasa senang”, “senang yg”, “yg bikin”, “bikin ga”, “ga bisa”, “bisa tidur”, “tidur ada”. “ada rasa”, “rasa sedih”, “sedih yg”, “yg bikin”, “bikin ga”, “ga bisa”, “bisa tidur”, “tidur ada”. “keduanya bisa”, “bisa terjadi”, “terjadi karena”, “karena dikasihi”, “dikasihi dan”, “dan mengasihi”. “dan keduanya”, “keduanya bisa”, “bisa terjadi”, “terjadi dalam”, “dalam waktu”,

	“waktu yg”, “yg bersamaan”, “bersamaan bisa”
--	---

3.5 Pembobotan Kata

Setelah dilakukan tahap praproses data akan dilanjutkan dengan tahap pembobota kata dengan TF-IDF, pada tahap ini penulis menggunakan aplikasi Microsoft Excel untuk melakukan penghitungan TF-IDF.

3.6 Pengujian Algoritma Naïve Bayes

Setelah dilakukan tahap pembobotan kata selanjutnya dilakukan eksperimen dengan menguji algoritma *Naïve Bayes*, *software* yang digunakan sebagai alat bantu adalah Rapidminer.

3.7 Evaluasi Dan Validasi

Untuk mengetahui performa klasifikasi, diperlukan pengukuran akurasi, presisi, dan *recall*. Pengukuran ini biasa disebut dengan *confusion matrix*. *Confusion Matrix* dengan *true positive* (*TP*), *true negative* (*TN*), *false positive* (*FP*) dan *false negative* (*FN*) sebagai indikator. Validasi yang digunakan dalam penelitian ini adalah *k-fold validation* menggunakan k=5 sampai k=10.