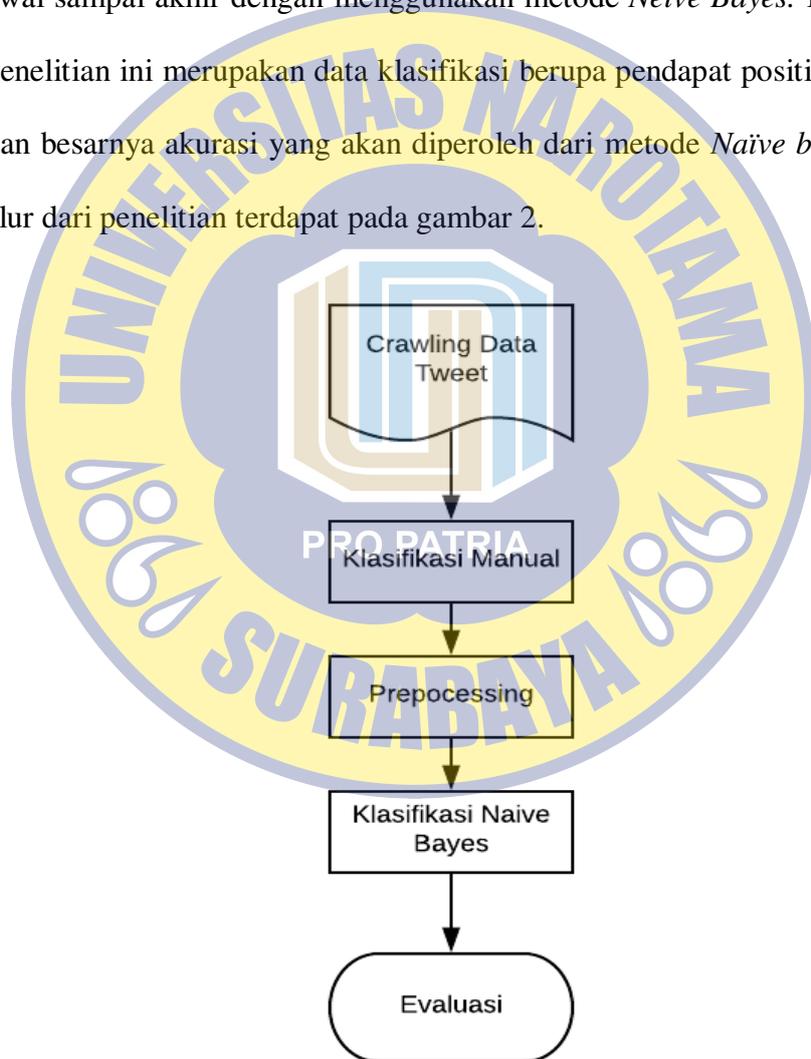


BAB III

METODOLOGI PENELITIAN

3.1 Deskripsi Penelitian

Deskripsi ini adalah tahap-tahap bagaimana alur dari penelitian ini. Deskripsi ini bertujuan supaya bisa menggambarkan proses klasifikasi dari awal sampai akhir dengan menggunakan metode *Neive Bayes*. Keluaran dari penelitian ini merupakan data klasifikasi berupa pendapat positif atau negatif dan besarnya akurasi yang akan diperoleh dari metode *Naïve bayes*. Berikut alur dari penelitian terdapat pada gambar 2.



Gambar 2. Gambaran Deskripsi Penelitian

Penelitian diawali dari proses memasukkan (*input*) data *tweet* hasil *crawling*. Dokumen hasil *crawling* akan diklasifikasi manual dengan memberikan identitas (label) positif atau negatif pada *dataset*.

3.2 *Input*

Input disini adalah dokumen dalam *dataset* hasil dari proses *crawling* berupa *tweet* yang telah dilabeli positif atau negatif, seperti yang terdapat pada tabel 2.

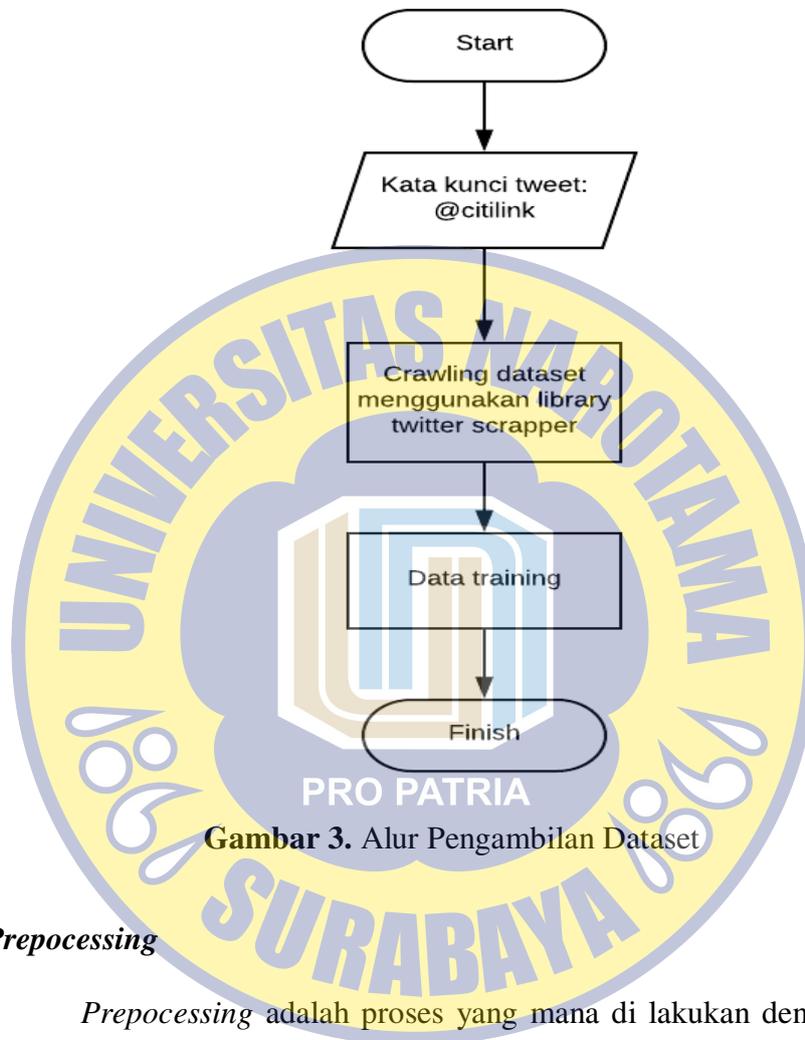
Table 2. Contoh Dokumen Input

Opini	Tweet
P	@Citilink #Terima kasih Citilink atas pelayanan prima merubah last min tiket mudik kt krn wafatnya ayah kami td mlm.Semoga makin sukses
N	@Citilink pramugari tidak ramah, hanya berdiri saja tidak membantu penumpang

3.3 *Dataset*

Dataset berupa teks yang diambil untuk penelitian menggunakan kata kunci(*query*) '@citilink'. kata kunci tersebut merupakan akun resmi dari Citilink Indonesia yang ada pada media sosial Twitter. Pengambilan *dataset* dengan cara *crawling* menggunakan *library twitter scrapper* yang bisa jalankan langsung pada komputer masing-masing. *Dataset* hasil *crawling* akan dikalifikasi secara manual, dengan cara memberikan identitas (label)

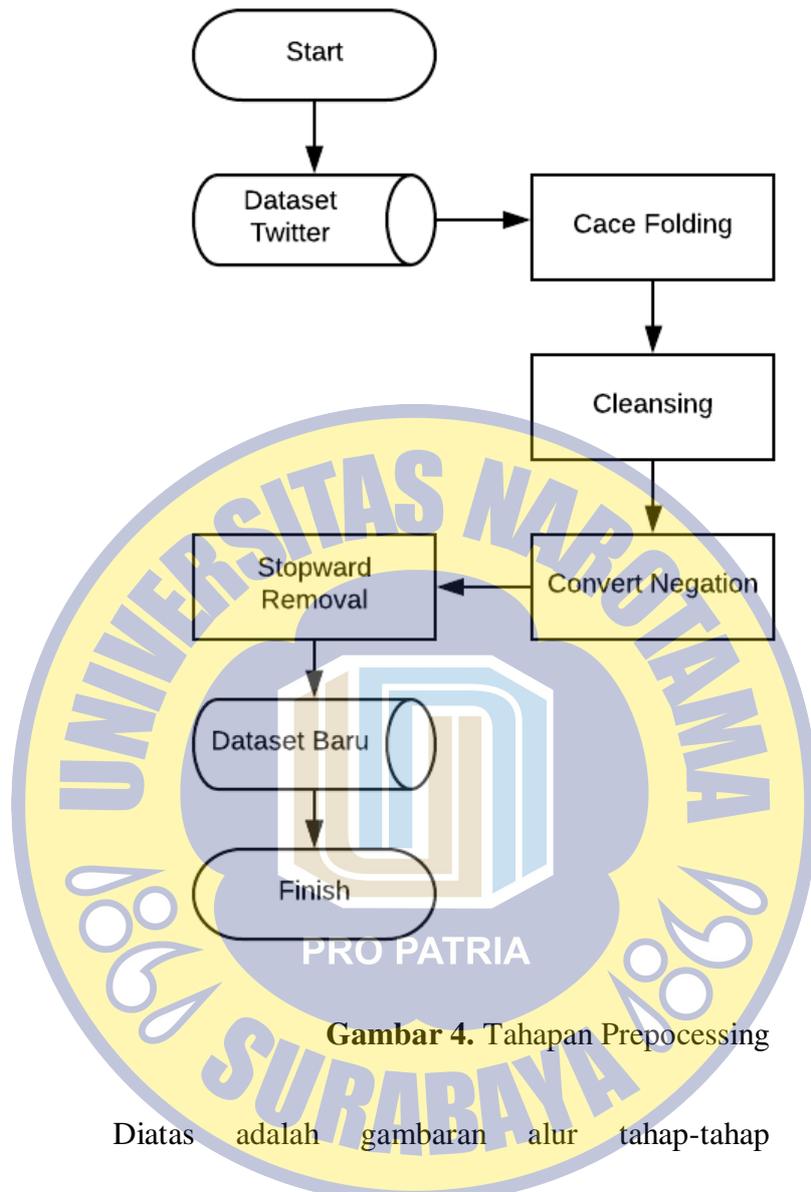
positif atau negatif. Berikut alur proses pengambilan *dataset* terdapat pada gambar 3.



Gambar 3. Alur Pengambilan Dataset

3.4 Preprocessing

Preprocessing adalah proses yang mana dilakukan dengan beberapa tahap sebelum masuk pada tahap klasifikasi dengan metode *Naive Bayes*. Proses ini untuk mengolah dokumen *tweet* dari *dataset*, membaca setiap kata pada dokumen yang dimasukkan, dan menghasilkan dokumen *tweet* baru yang akan mempengaruhi hasil dari proses klasifikasi *naive bayes*. Dalam proses *preprocessing* tahap yang hendak dilaksanakan. Berikut tahapan-tahapan *preprocessing* seperti pada gambar 4.

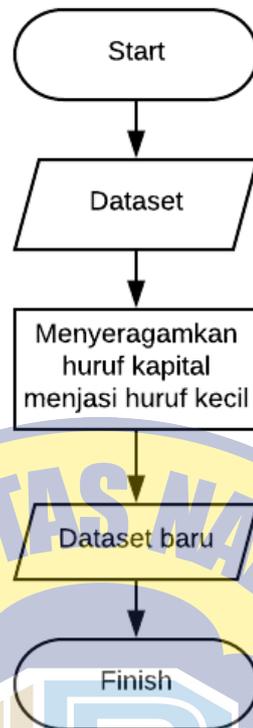


Gambar 4. Tahapan Preprocessing

Diatas adalah gambaran alur tahap-tahap dari proses *preprocessing*. Berikut penjelasan pada setiap tahap-tahap diatas :

3.4.1 Case Folding

Tahap *case folding* merupakan proses penyamaan huruf pada dokumen *tweet* dengan mengubah huruf kapital jadi huruf kecil, supaya merenggangkan kata berlebih-lebihan dikerenakan perbedadaan huruf, dan akan menghasilkan dokumen *tweet* baru. Berikut gambaran pada tahap *case folding* terlihat pada gambar 5.



Gambar 5. Tahap Case Folding

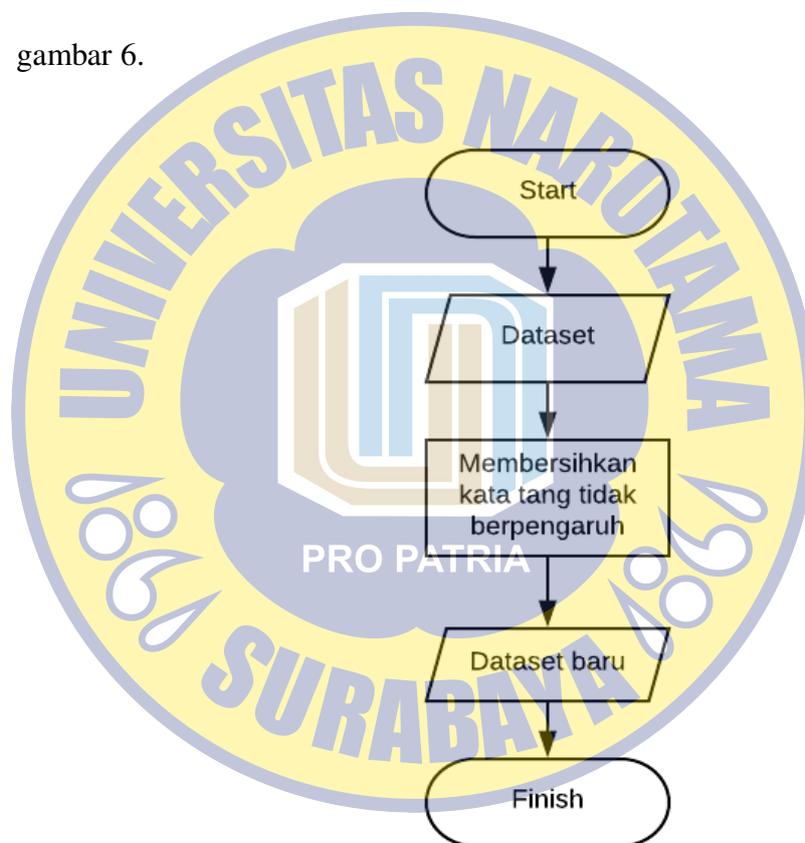
Dari proses diatas akan menghasilkan *tweet* baru. Berikut contoh *tweet* yang dihasilkan dari tahap *case folding* bisa dilihat pada tabel 3

Table 3. Hasil Proses Case Folding

Input	Output
@Citilink #Terima kasih Citilink atas pelayanan prima merubah last min tiket mudik kt krn wafatnya ayah kami td mlm.Semoga makin sukses	@citilink #terima kasih citilink atas pelayanan prima merubah last min tiket mudik kt krn wafatnya ayah kami td mlm.Semoga makin sukses

3.4.2 *Cleansing*

Tahap *cleansing* adalah proses yang bertujuan untuk menghilangkan *tweet* dari kata yang kurang dibutuhkan seperti simbol HTML, kata kunci, *hashtags*(#), *username*, email dan simbol (!@#%<>[]=;:”,’&^\$*+_-/?), yang akan menghasilkan dokumen *tweet* baru. Berikut gambaran pada tahap *cleansing* seperti yang ditampilkan gambar 6.



Gambar 6. Tahapan Cleansing

Dari proses diatas akan menghasilkan *tweet* baru. Berikut contoh dokumen *tweet* yang menjadi keluaran dari tahap *cleansing* bisa dilihat pada tabel 4.

Table 4. Contoh tweet hasil cleansing

Input	output
@Citilink #Terima kasih Citilink atas pelayanan prima merubah last min tiket mudik kt krn wafatnya ayah kami td mlm.Semoga makin sukses	terima kasih Citilink atas pelayanan prima merubah last min tiket mudik kt krn wafatnya ayah kami td mlm.Semoga makin sukses

3.4.3 *Convert Negation*

Proses *convert negation* bertujuan untuk transformasi kata negasi yang tergabung dalam *tweet*. Kata tersebut akan mengganti arti suatu sentimen pada *tweet* tersebut, kata negasi akan tergabung dengan kata setelahnya. Contoh kata negasi 'tidak' banyak lagi kata negasi yang akan menghasilkan dokumen *tweet* baru, dan akan memberikan pengaruh terhadap proses klasifikasi dengan metode *naïve bayes* pada penelitian ini. Berikut gambaran tahap *convert negation* bisa dilihat seperti gambar 7.



Gambar 7. Tahap Convert Negation

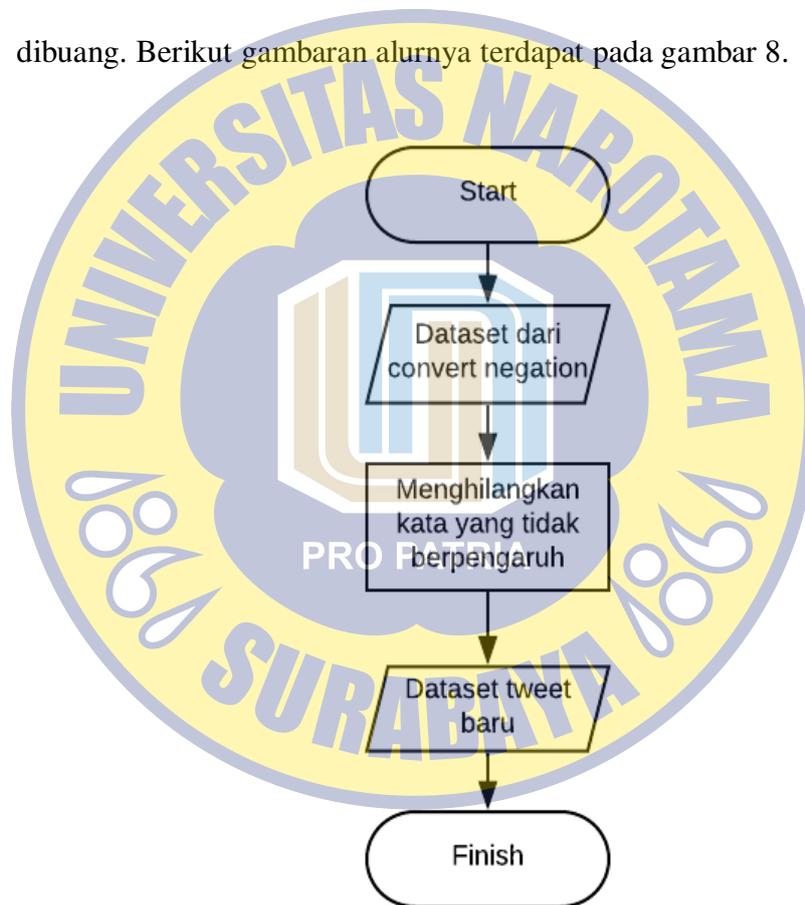
Dari proses diatas akan menghasilkan *tweet* baru. Berikut contoh *tweet* yang dihasilkan dari tahap *convert negation* terdapat pada tabel 5.

Table 5. Contoh Tweet hasil Convert Negation

Inpu	Output
citilink pramugari tidak ramah, hanya berdiri saja tidak membantu penumpang	citilink pramugari tidakramah, hanya berdiri saja tidak membantu penumpang

3.4.4 Stopword Removal

Stopword Removal, yakni teknik yang bertujuan untuk menghapus kata didalam dokumen *dataset*, kata tersebut seperti “di”, “ke”, “dari”, “yang”, “sedang”, “ini”, karena dianggap tidak memberikan deskripsi sesuatu dalam bahasa Indonesia. Tapi pada teks klasifikasi kata semacam “tidak”, “bukan”, “tanpa” kurang penting sehingga kata tersebut tidak ikut dibuang. Berikut gambaran alurnya terdapat pada gambar 8.



Gambar 8. Tahap Stopword Removal

Dari proses diatas akan menghasilkan *tweet* baru. Berikut contoh hasil tahap *stopword removal* terdapat pada tabel 6.

Table 6. Contoh Hasil Stopword Removal

<i>Input</i>	<i>Output</i>
citilink terima kasih Citilink atas	Citilink terima kasih pelayanan
pelayanan prima merubah last min	prima merubah last min tiket
tiket mudik kt krn wafatnya ayah	mudik kt wafatnya ayah kami mlm
kami td mlm.Semoga makin sukses	Semoga sukses

3.5 Klasifikasi Naïve Bayes

Naive Bayes (Naive Bayes Classifier) adalah algoritma yang sangat efektif dalam permasalahan klasifikasi atau penggolongan. Algoritma ini bekerja berdasarkan probabilitas yang sudah ada untuk menentukan probabilitas yang akan datang. Meskipun *dataset* yang digunakan ada jutaan jumlahnya, *Naive Bayes* adalah pendekatan yang dianjurkan untuk digunakan. Untuk memahami lebih jauh tentang *Naive Bayes* maka perlu terlebih dahulu memahami dalil *Bayes*. Dalil *Bayes* dinamakan sesuai dengan nama penemunya Rev. Thomas Bayes.[8] Algoritma ini bekerja berdasarkan probabilitas bersyarat, yaitu probabilitas yang menjelaskan tentang sesuatu akan terjadi berdasarkan peristiwa yang telah terjadi sebelumnya.[10] Secara sederhana persamaan probabilitas dapat dirumuskan sebagai berikut.

$$P(A/B) = \frac{P(A/B)P(A)}{P(B)}$$

3.6 Evaluasi

Evaluasi bertujuan untuk mengecek kebenaran hasil klasifikasi dengan menghitung berapa angka yang diperoleh. Acuan yang dipakai untuk tahap evaluasi yakni perhitungan tabel matriks klasifikasi (*confussion matrix*). Berikut tabel perhitungan *confussion matrix* bisa dilihat pada tabel 7.

Tabel 7. *Confussion Matrix*

		True Class	
		Positive	Negative
Kelas Terprediksi	Positif	TP(True Positif)	FP(False Positive)
	Negatif	FN(False Negative)	TN(True Negative)

Pada matriks diatas mempunyai empat penilaian yang digunakan sebagai parameter dalam perhitungan, yaitu :

- 1) *True Positive*, pada saat kelas diprediksi positif dan kenyataannya positif.
- 2) *True Negative*, pada saat kelas diprediksi negatif dan kenyataannya negatif.
- 3) *False Positive*, pada saat kelas diprediksi positif dan kenyataannya negatif.
- 4) *False Negative*, pada saat kelas diprediksi negatif dan kenyataannya positif.

Akurasi sebagai opsi kebenaran hasil dari klasifikasi dengan rumus sebagai berikut :

$$\text{Akurasi}(A) = \frac{TP + TN}{(TP + TN + FP + FN)}$$

