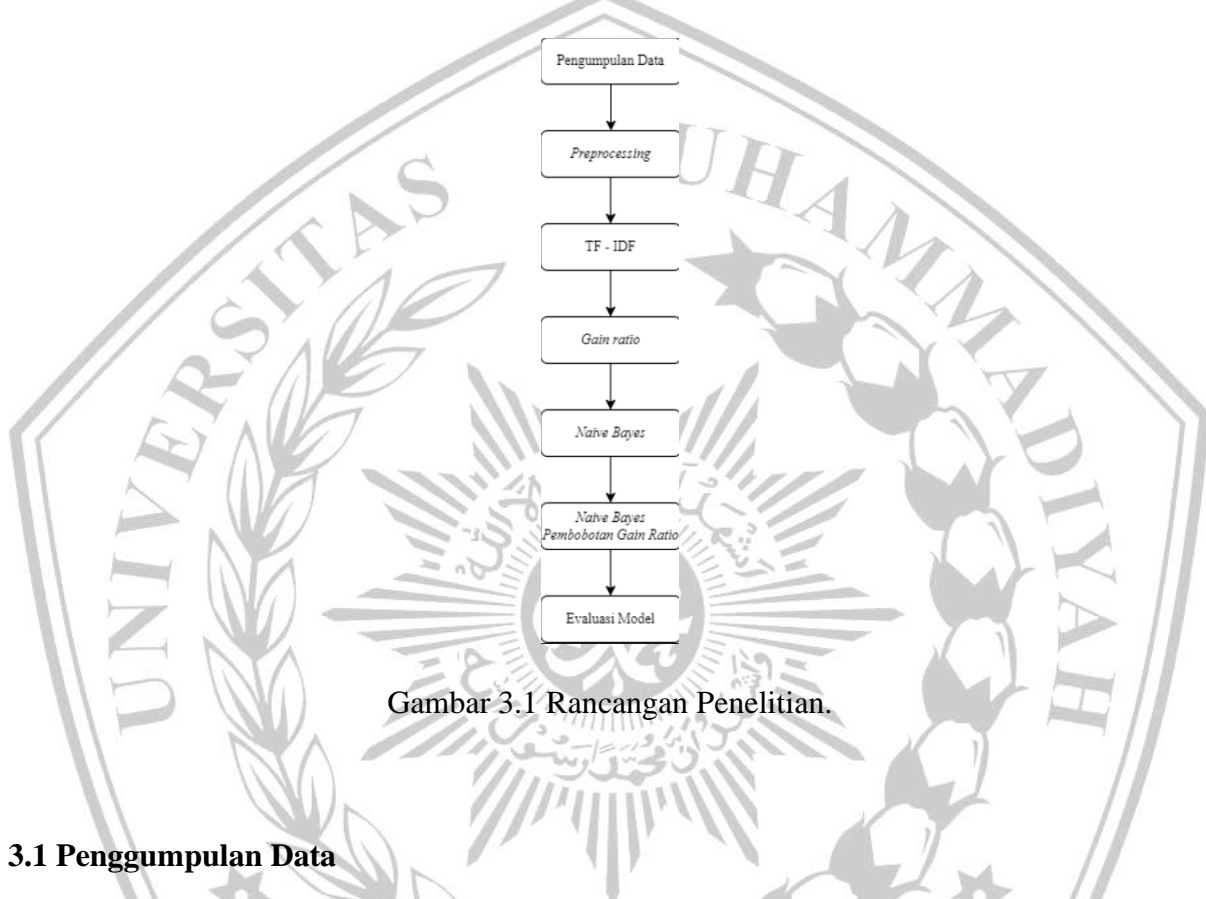


BAB 3 METODE PENELITIAN

Pada tahap metode penelitian ini menggunakan rancangan untuk pengambilan keputusan informasi yang bertujuan untuk mendapatkan hasil yang diharapkan. Tahapan rancangan metode penelitian ini terdapat pada Gambar 1 yang meliputi pengumpulan data, *preprocessing*, TF-IDF, *Gain Rasio*, Naïve Bayes, dan evaluasi model.



Gambar 3.1 Rancangan Penelitian.

3.1 Pengumpulan Data

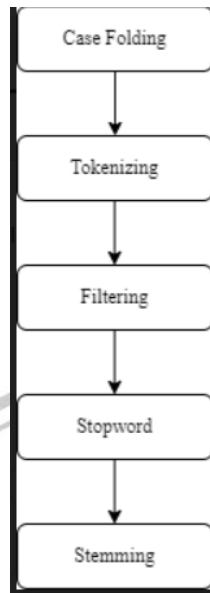
Pada proses pengumpulan data dilakukan secara manual dengan kurung waktu selama 1 minggu dengan cara menonton siaran langsung streamer youtube game online lalu memilah – milah beberapa komentar bullying dan non bullying pada kolom komentar. Labelling data dilakukan secara maual. Data yang diperoleh berjumlah 1000 dengan komentar negtive sebanyak 500 dan komentar positive sebanyak 500. Sumber data yang didapatkan beberapa dari *channel youtube Mobile Legends : Bang Bang Official*, Pascol Kintil, XINNN dan beberapa situs game online yang ada.

Tabel 2. contoh dataset

NO	TANGGAL	NAMA CHANNEL	JUDUL STREAMER	KOMENTAR	NAMA KOMENTAR	Label
1	21 - Februari - 2022	XINNN	HAYABUSA LANCELOT CLAUDE !!!	HALOOOOO BANGSIN..... DUHHH KANGEN BANGETTTTTTTTT	Tiara Viola	<i>Positive</i>
2	21 - Februari - 2022	XINNN	HAYABUSA LANCELOT CLAUDE !!!	GG gameplay core lu bg tpi mau aja ngisi gold #pindahaja	Tiara Viola	<i>Positive</i>
3	21 - Februari - 2022	XINNN	HAYABUSA LANCELOT CLAUDE !!!	bapak kinn ganteng	ayam	<i>Positive</i>
4	21 - Februari - 2022	Vin	Main Santai Spam Mage 1000 Point !	BANYAK NGOMONG SKIL NGGA PUNYA WKWK	VIN TAI YONG	<i>Negative</i>
5	21 - Februari - 2022	Vin	Main Santai Spam Mage 1000 Point !	youtuber cacad gaberani solo	agung renaldy	<i>Negative</i>

3.2 Preprocessing

Preprocessing merupakan tahap awal yang mempersiapkan data sebelum digunakan dengan bertujuan mendapatkan proses data yang terstruktur dan jelas sehingga menghasilkan informasi dengan kualitas yang baik dan siap digunakan.



Gambar 3.2 Proses *Preprocessing*.

a. Case Folding

Pada tahap ini, melakukan perubahan kata-kata huruf besar disetarakan menjadi huruf kecil. Contoh perubahan *case folding* dibawah ini.

Tabel 3. Perbandingan data tahapan Case Folding

Tahap	Komentar
Sebelum	SIAPA AJA BOLEH MVP FRANCONTOL PUN BOLEH ASAL BUKAN GRANGER!!!
Sesudah	siapa aja boleh mvp francontol pun boleh asal bukan granger!!!

b. Tokenizing

Pada tahap ini, melakukan penghapusan tanda baca dan simbol agar bisa dilakukannya pembobotan. Contoh perubahan *tokenizing* dibawah ini.

Tabel 4. Perbandingan data tahapan Tokenizing

Tahap	Komentar
Sebelum	siapa aja boleh mvp francontol pun boleh asal bukan granger!!!

Sesudah	siapa aja boleh.mvp francontrol pun boleh asal bukan granger
---------	--

c. Stopword Removal

Pada tahap ini, *stopword removal* bertugas untuk menghilangkan kata yang tidak penting pada dokumen, seperti kata sambung, kata ganti dan lain-lainya. Contoh perubahan *stopword removal* dibawah ini.

Tabel 5. Perbandingan data tahapan Stopword Removal

Tahap	Komentar
Sebelum	siapa aja boleh.mvp francontrol pun boleh asal bukan granger
Sesudah	siapa boleh.mvp francontrol asal bukan granger

d. Steaming

Pada tahap ini, dilakukan proses mengubah kata imbuhan menjadi kata dasar. Contoh perubahan *steaming* dibawah ini.

Tabel 6. Perbandingan data tahapan Steaming

Tahap	Komentar
Sebelum	siapa boleh.mvp francontrol asal bukan granger
Sesudah	Siapa boleh.mvp francontrol asal bukan granger

3.3 TF-IDF

Setelah melakukan tahap *preprocessing* maka dapat dilanjutkan melakukan tahapan TF-IDF untuk mendapatkan nilai bobotnya pada setiap dokumen. Berikut ini alur penjelasan tahapan TF-IDF :

1. Menyiapkan hasil data yang sudah diolah pada tahap *preprocessing*.

2. Dilakukan proses TF (*Term Frequency*) pada setiap kata yang terdapat pada dokumen.
3. Melakukan tahapan IDF (*Invers Document Frequency*) untuk menghitung ketersediaan *term* didalam dokumen tersebut.
4. Selanjutnya melakukan perhitungan dari nilai TF-IDF.
5. Output data hasil dari perhitungan TF-IDF.

Hasil dari proses tahapan TF-IDF mendapatkan data yang sudah memiliki nilai bobot disetiap kata yang terdapat pada dokumen. Setelah melakukan tahapan TF-IDF maka dilakukan proses selanjutnya *gain rasio*.

3.4 Gain Ratio

Setelah dilakukan tahap preprocessing maka selanjutnya di lakukan seleksi fitur *Gain Ratio*. *Gain Ratio* merupakan teknik pembobotan atribut yang pada umumnya digunakan pada metode *Decision Trees* [7]. Seleksi fitur ini berpengaruh besar pada proses prediksi dan fitur ini juga menyeleksi berdasarkan rangking nilai *Gain Ratio*. Maka semakin besar nilai yang dimiliki *gain rasio* dapat menjadi pengaruh besar terhadap proses prediksi. Berikut ini alur pembobotan *Gain Rasio* :

1. Menghitung nilai *Entropy* pada setiap fitur yang terdapat pada setiap dokumen.
2. Menghitung nilai *Information Gain* pada setiap fitur yang terdapat pada setiap dokumen.
3. Menghitung nilai *Split Information* setiap fitur yang terdapat pada setiap dokumen.
4. Menghitung nilai *Gain Ratio* setiap fitur yang terdapat pada setiap dokumen.
5. Mengurutkan posisi nilai yang terdapat pada *Gain Ratio* dan menggumpulkan fitur yang memiliki pengaruh terhadap hasil prediksi.

3.5 Naïve Bayes

Setelah melakukan tahapan *preprocessing* dengan melakukan pembobotan *Gain Ratio*. Maka selanjutnya dilakukan klasifikasi dengan menggunakan metode Naïve Bayes. Berikut ini alur klasifikasi Naïve Bayes :

1. Setelah dilakukan proses tahapan preprocessing dengan menggunakan tahapan pembobotan *Gain Ratio*.
2. Menghitung nilai probabilitas terdapat pada data *training*.
3. Setelah melakukan perhitungan probabilitas Naïve Bayes tersebut disimpan untuk dilakukannya pengujian pada data uji.
4. Selanjutnya menghitung nilai probabilitas data sentimen yang akan diujikan.

3.6 Evaluasi Pengujian

Dalam penelitian ini dilakukan pengujian yang dilakukan sebanyak 3 kali pengujian. Proses pengujian pertama yaitu klasifikasi menggunakan metode Naïve Bayes. Pengujian kedua dilakukan dengan metode Naïve Bayes dengan pembobotan *Gain Ratio* dengan membagi data *train* 90% dan data *test* 10%. Pengujian ketiga dilakukan pembagian data *train* 80% dan data *test* 20%. Pengujian keempat dilakukan pembagian data *train* 70% dan data *test* 30%. Pengujian dilakukan dengan cara menghitung nilai akurasi, *recall*, *precision*, dan *f-measure* disertai dengan tabel *confusion matrix* yang terdiri dari *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). Pada pengujian tersebut dilakukan untuk mengetahui pengujian mana yang baik untuk mendapatkan hasil akurasi terbaik.

