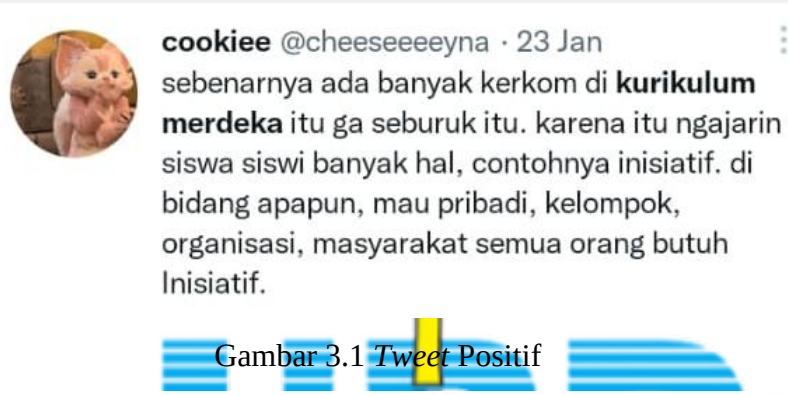
BAB III
METODE PENELITIAN

3.1 Objek Penelitian

Objek penelitian yang digunakan pada penelitian ini adalah *Tweet* dari para pengguna *Twitter* yang berkaitan dengan kata "Kurikulum Merdeka".



Gambar 3.1 *Tweet* Positif

Gambar 3.2 *Tweet* Negatif

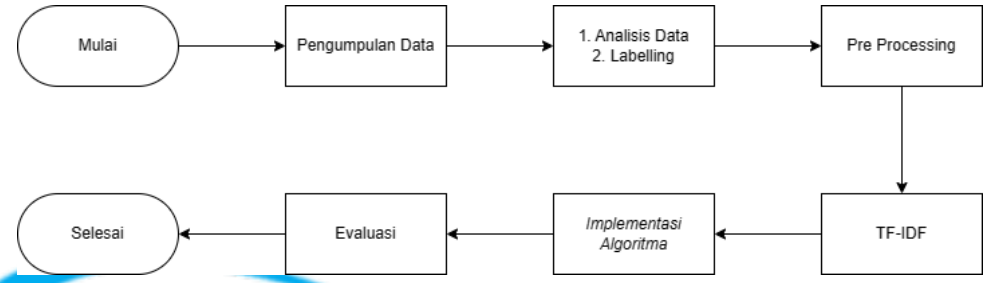
3.2 Peralatan Penelitian

Penelitian ini akan dilakukan dengan menggunakan peralatan yang terdiri dari perangkat keras dan perangkat lunak.

1. Perangkat keras yang digunakan:
   a. Laptop Acer Aspire E1-431 Intel(R) Core (TM) i7-3632QM CPU @ 2.20GHz
   b. RAM 8GB
   c. HDD 320GB
   d. SSD 256GB
2. Perangkat lunak yang digunakan:
   a. 64-*bit windows* 10 *pro operating system*
   b. *Jupyter Notebook*
   c. *Twitter*
   d. *Microsoft Office*

KARAWANG

3.3 Prosedur Penelitian



Gambar 3.3 *Flowchart* Prosedur Penelitian

Pada penelitian analisis diatas metode pengumpulan data menggunakan *library Snscrape* pada *Jupyter Notebook* menggunakan bahasa *python*, data yang sudah diambil kemudian dianalisis terlebih dahulu dan diberi label secara manual, kemudian melalui proses *pre processing* untuk dilakukan pembersihan data. Setelah data bersih kemudian dilakukan *feature extraction* yaitu mengubah data teks menjadi angka. Selanjutnya data akan di-*modelling* menggunakan algoritma, kemudian ke tahap evaluasi yaitu pengujian model dengan menggunakan *Confusion Matrix*.

3.3.1 Pengumpulan Data

Proses pengumpulan data terbagi menjadi dua tahap, studi literatur dan pengambilan data. (*Data Crawling*).

1. Studi literatur

   Studi literatur dilakukan untuk mencari dasar teori dan mengumpulkan referensi dari berbagai sumber seperti jurnal, buku, dan internet. Langkah ini mencakup pencarian definisi istilah, tinjauan pustaka, serta penelitian sebelumnya dan metode yang telah digunakan.

2. Crawling data

   Pengambilan data menggunakan *library snscrape* pada *jupyter notebook* menggunakan bahasa *python* dengan kata kunci "Kurikulum Merdeka".

	datetime	username	tweet
0	Fri Jun 02 10:09:08 +0000 2023	kyeomiecute_	@pupxkitten Heheh kbtulan dh jdi anak ipa kak ...
1	Fri Jun 02 10:16:54 +0000 2023	jawapos	Kurikulum Merdeka Kembangkan Keterampilan Sisw...
2	Fri Jun 02 06:14:38 +0000 2023	_1808jpg	tolong rekomendasi inn akuuubuku latsoll yang ...
3	Fri Jun 02 07:03:17 +0000 2023	kumparan	Dalam Kurikulum Merdeka, Projek Penguatan Prof...
4	Tue May 30 07:29:30 +0000 2023	numukuulyu	Dulu kupikir jurusan yg sama di semua univ tuh...
...	...	...	...
2060	Tue Apr 11 22:10:01 +0000 2023	kiivchrwt	@tanyarlfes masi kurikulum 2013 kak \nmungkin ...
2061	Wed Apr 12 01:47:36 +0000 2023	daniellsinaga_	Ada informasi tambahan kalau pendaftaran sekol...
2062	Wed Apr 12 02:34:54 +0000 2023	jwmnlatte	kurikulum merdeka presentasi mulu ah capek
2063	Wed Apr 12 00:02:04 +0000 2023	okeysja	@amyckerman @jodohchae @convomf intinya mereka...
2064	Wed Apr 12 01:12:02 +0000 2023	Daron202	Buku Informatika Kelas 9 SMP/MTS Kurikulum Mer...

Gambar 3.4 *Crawling* Data

3.3.2 Analisis Data

Analisis data dilakukan untuk mengetahui permasalahan pada media sosial *Twitter*, yaitu tanggapan masyarakat tentang kurikulum merdeka. Analisis ini diperlukan untuk mendapatkan sentimen positif dan negatif. Proses pengecekan data secara manual untuk menghilangkan data duplikat dan netral seperti iklan agar data yang diperoleh hanya sentimen positif dan negatif, menurut ahli bahasa yaitu Pionika Linda Sari, kalimat diaggap netral jika tidak membuat nilai positif dan negatif serta tidak memihak salah satu kebijakan. berikut data netral yang sudah di pisah dari hasil *crawling* sebelumnya.

	username	tweet
0	litbangdikbud	Yuk, kunjungi laman https://t.co/jwpeIrG0XH un...
1	tribunkaltim	Kunci Jawaban Sosiologi Kelas 11 Halaman 32 33...
2	LPMP_KepBabel	#KurikulumMerdeka #PembelajaranBerkualitasBagi...
3	litbangdikbud	Halo #Sahabatdikbud, tahukah kalian ada yang b...
4	macam_cara	Kumpulan Soal PTS Bahasa Indonesia Kelas 7 Sem...
...	...	...
1432	tribunkaltim	Kunci Jawaban USP IPS Kelas 6, 30 Soal Pilihan...
1433	tribunkaltim	INILAH KUNCI JAWABAN IPS KELAS 10 HALAMAN 80 T...
1434	tribunkaltim	Inilah kunci jawaban IPS kelas 10 halaman 80 t...
1435	LPMP_KepBabel	#KurikulumMerdeka #PembelajaranBerkualitasBagi...
1436	LPMP_KepBabel	Untuk mendaftar tonton dulu tutorial cara mela...

Gambar 3.5 Data Netral

3.3.3 *Labelling*

Setelah analsis data tahap selanjutnya adalah pelabelan, di mana label diberikan pada setiap data dalam dua kategori kelas yang berbeda, kelas positif dan kelas negatif:

1. Positif: Berdasarkan definisi yang ditemukan dalam Kamus Besar Bahasa Indonesia (KBBI), sentimen positif mengacu pada respons atau sikap yang meningkatkan nilai individu atau entitas tertentu. Berdasarkan penyusunan frasa dalam riset ini, kalimat dengan nuansa positif dapat diilustrasikan sebagai berikut: " Senang sekali dengan adanya beberapa taman dan ruang baca, semoga bisa membuat kota Pontianak lebih bersinar”
2. Negatif: Mengacu pada definisi dalam Kamus Besar Bahasa Indonesia (KBBI), sentimen negatif merujuk pada respon atau sikap yang merendahkan nilai individu atau entitas tertentu. Kalimat dengan nuansa negatif umumnya memiliki potensi untuk mengurangi persepsi positif terhadap hal tertentu, dan sering kali dapat mengarah pada penurunan tren. Biasanya, penggunaan kata-kata negasi menjadi ciri khas dari kalimat bersentimen negatif. Negasi adalah unsur yang ditemukan di semua bahasa, dan seringkali digunakan untuk mengubah polaritas suatu pernyataan. Berikut adalah contoh struktur kalimat bersentimen negatif: “Rencana pemerintah merenovasi Taman Budaya Kalimantan Barat tak pernah terealisasi. Tak pernah di perbaiki sejak di bangun tahun 1982”(Ardiani, Sujaini & Tursina 2020).

Tabel 3.1 Contoh *Labeling*

Kalimat	Label
Kurikulum ini bagus buat nyari kerja	Positif
Kebanyakan kerja kelompok jadi males	Negatif

3.3.4 Pre-processing

Tahap *Pre-processing* adalah langkah sebelum proses klasifikasi yang bertujuan mengubah kata-kata yang tidak terstruktur menjadi format yang lebih terstruktur, agar mempermudah tahap pemrosesan berikutnya. Tahap

ini melibatkan beberapa langkah, termasuk *Cleaning, Case Folding, Tokenizing, Normalisasi, Stopword,* dan *Stemming*.

1. *Cleaning* merupakan tahap di mana karakter-karakter tak diperlukan dalam teks, seperti tanda baca, angka, simbol, emotikon, dan tautan dari sumber lain, dihilangkan.

Tabel 3.2 Contoh *Cleaning*

Sebelum	Sesudah
Kurikulum Merdeka wktu istirahatnya jadi berkurang!	Kurikulum Merdeka wktu istirahatnya jadi berkurang

2. *Case Folding* adalah langkah mengubah seluruh huruf menjadi huruf kecil guna penyamaan.

Tabel 3.3 Contoh Case Folding

Sebelum	Sesudah
Kurikulum Merdeka wktu istirahatnya jadi berkurang	kurikulum merdeka wktu istirahatnya jadi berkurang

3. *Tokenizing* adalah pemecahan kalimat menjadi bagian yang disebut token, seperti kata.

Tabel 3.4 Contoh *Tokenize*

Sebelum	Sesudah
kurikulum merdeka wktu istirahatnya jadi berkurang	[kurikulum, merdeka, wktu, istirahatnya, jadi, berkurang]

4. *Normalize* mengubah koreksi kesalahan eja, mengubah kata singkatan ke bentuk umum.

Tabel 3.5 Contoh *Normalize*

Sebelum	Sesudah

[kurikulum, merdeka, wktu, istirahatnya, jadi, berkurang]	[kurikulum, merdeka, waktu, istirahatnya, jadi, berkurang]

5. *Stopword* proses menghilangkan kata-kata yang tidak memiliki pengaruh signifikan terhadap makna dan sentimen suatu kalimat

Tabel 3.6 Contoh *Stopword*

Sebelum	Sesudah
[kurikulum, merdeka, waktu, istirahatnya, jadi, berkurang]	[kurikulum, merdeka, waktu, istirahatnya, berkurang]

6. *Stemming* adalah proses mengubah kata-kata menjadi bentuk dasar. *Stemming* dapat mengurangi semua variasi dari suatu kata menjadi bentuk yang sama dan lebih sederhana.

Tabel 3.7 Contoh *Stemming*

Sebelum	Sesudah
[kurikulum, merdeka, waktu, istirahatnya, berkurang]	[kurikulum, merdeka, waktu, istirahat, kurang]

3.3.5 TF-IDF

Term Frequency-Inverse Document Frequency yaitu mengubah data *text* jadi data numerik agar data tersebut dapat diolah oleh komputer, salah satu teknik *feature extraction* adalah *Term Frequency-Inverse Document Frequency (TF-IDF)* merupakan metode yang digunakan untuk menghitung bobot kata dalam sebuah dokumen berdasarkan frekuensi kemunculan kata tersebut dalam dokumen itu sendiri serta jarangnya kata tersebut muncul dalam seluruh koleksi dokumen (Gifari et al., 2022).

TF-IDF dapat dihitung menggunakan rumus berikut:

$$Wt = tf \text{ x } idf$$

$$\text{IDF} = \log \frac{\text{N}}{\text{df}}$$

Keterangan:

TF = Jumlah kata dalam setiap dokumen

DF = Jumlah total kata dari seluruh dokumen

N = Jumlah total dataset

IDF = jumlah dokumen dalam kumpulan dokumen yang mengandung kosa kata.

3.3.6 Implementasi *Naïve bayes Multinomial* dan *Complement*

Setelah melakukan *pre-processing* selanjutnya akan di implementasikan dengan algoritma *multinomial naïve bayes* dan *complemet naïve bayes*. Klasifikasi suatu metode untuk mengelompokan dan mencari *itemset* sesuai dengan nilai bobot masing-masing. Pada penelitian ini terdapat 2 kelas yaitu positif dan negatif. Berikut perhitungan klasifikasi *multinomial naïve bayes* dan *complement naïve bayes.*

1. *Naïve Bayes Multinomial*

$$P(c) = \frac{Nc}{N}$$

- (Nc) = Menghitung berapa banyak dokumen atau data yang termasuk dalam kelas tertentu.
- (N) = menghitung total jumlah dokumen dalam dataset.
- P(c) = Nc/N untuk menghitung probabilitas prior dari kelas (c) dengan membagi jumlah dokumen dalam kelas tersebut (Nc) dengan total jumlah dokumen dalam dataset (N).

$$P(w,c) = \frac{count(w,c)+1}{\sum count(c)+|V|}$$

- P(w,c) = Probabilitas kondisional dari kata (w) terhadap kelas (c)
- Count(w,c) = Jumlah kemunculan kata (w) dalam dokumen yang termasuk dalam kelas (c)
- ∑count(c) = Jumlah total kemunculan semua kata dalam dokumen yang termasuk dalam kelas (c).
- |V| = Jumlah total kata.
- Penambahan angka 1 dalam pembilang digunakan untuk menghindari pembagian dengan nol (jika count(w,c) = 0)

$$P(w) = P(c) * P(w|c)*\ldots*P(w_n|c_n)$$

- P(w) = Peluang kategori a jika ada kemunculan kata b.

- P(c) = *Probabilitas* kemunculan nilai *prior*.
- P(w|c) = Peluang kata (w) yang muncul pada dokumen (c).
- P(wn|cn) = *Probabilitas* kondisional dari kata individu (wn) terhadap kelas yang sama (cn).

2. *Naïve Bayes Complement*

Metode *Complement Naïve Bayes* berfokus pada data yang tidak seimbang, dalam klasifikasi ini lebih berfokus pada data minoritas. Berikut adalah rumus dari algoritma *Complement*:

$$P(w) = P(\bar{c}) * P(w|c)*\ldots*P(w_n|c_n)$$

- P(w): Peluang kategori a jika ada kemunculan kata b.
- P($c\bar{}$): probabilitas komplementer atau probabilitas kebalikan dari kelas yang diprediksi.
- P(w|c): Peluang kata (w) yang muncul pada dokumen (c).

P(wn|cn): *Probabilitas* kondisional dari kata individu (wn) terhadap kelas yang sama (cn).

3.3.7 Evaluasi

Dalam evaluasi ini, *confusion matrix* digunakan sebagai metode untuk menilai kinerja dalam klasifikasi. Ini membantu menentukan nilai akurasi dari model yang telah dibuat. *Confusion matrix* terdiri dari 4 tabel: True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Dengan nilai-nilai yang diperoleh dari TP, FP, TN, dan FN, akurasi dapat dihitung.

Untuk nilai akurasi dapat dihitung sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+FN+TN}\ x\ 100\%$$