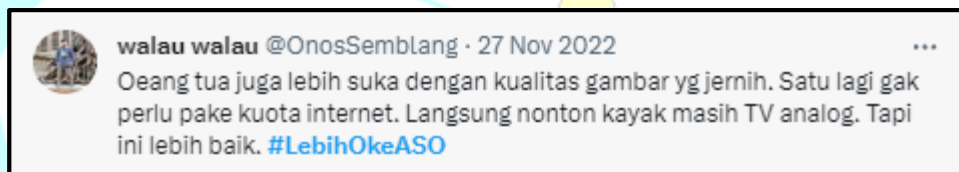


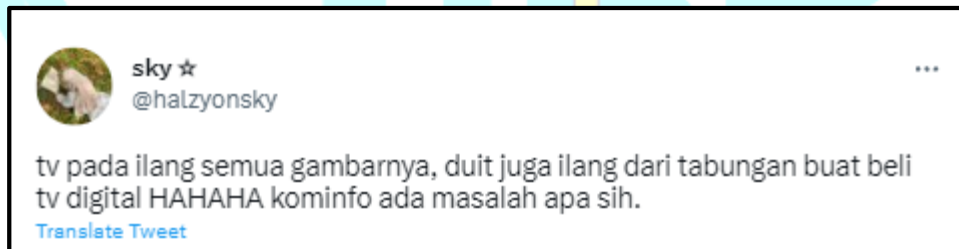
## BAB III METODE PENELITIAN

### 3.1 Objek Penelitian

Objek pada penelitian ini adalah *tweet* migrasi televisi analog ke televisi digital, dataset yang diperoleh dari media sosial *Twitter* dengan kata kunci “Migrasi Tv Digital”. Berdasarkan data tersebut akan dilakukan analisis sentimen menggunakan algoritma *Naive Bayes*.



Gambar 3. 1 *Tweet Positif*



Gambar 3. 2 *Tweet Negatif*

### 3.2 Peralatan Penelitian

Pada penelitian ini menggunakan perangkat keras dan perangkat lunak.

#### 3.2.1 Perangkat Keras

Perangkat keras yang digunakan pada penelitian ini antara lain:

1. Processor : AMD E2-9000e RADEON R2
2. Ram : 4 GB DDR 4
3. SSD : 128 GB

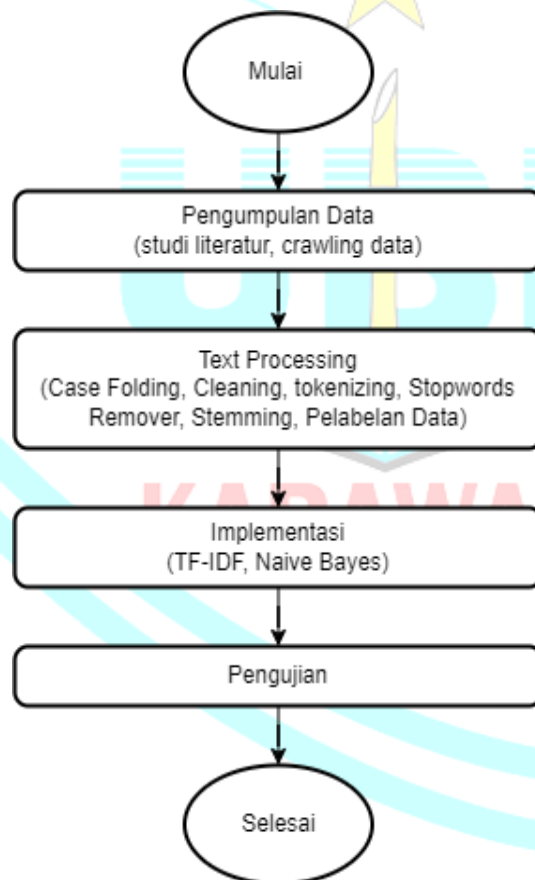
#### 3.2.2 Perangkat Lunak

Perangkat lunak yang digunakan pada penelitian ini antara lain:

1. *Operating system Windows 10*
2. *Microsoft office 2019*
3. *Google Chrome*
4. *Python*
5. *Google Colab*

### 3.3 Prosedur Penelitian

Untuk prosedur penelitian dapat digambarkan dengan bentuk *flowchart* seperti gambar dibawah ini :



Gambar 3. 3 Alur Penelitian

#### 3.3.1 Pengumpulan Data

Pada tahap pengumpulan data penelitian ini dibagi menjadi dua, antara lain sebagai berikut :

## 1. Studi Literatur

yaitu mencari landasan teori dan memperoleh data dari berbagai jurnal, buku, dan internet mendefinisikan dalam penggunaan *web browser*, contoh penelitian yang relevan, metode, data mining, analisis sentimen, dan data yang terkait dengan penelitian ini.

	Datetime	Tweet Id	Text	Username	Sentimen
0	2022-11-03 01:21:08+00:00	1.590000e+18	Jepang udah pindah ke tv digital dari 2011 dan...	misobach	-1
1	2022-11-03 00:02:34+00:00	1.590000e+18	@teguhjp14 @bagusajtm @DarahBI59166272 @idextr...	dwikyrestu	-1
2	2022-09-10 07:09:14+00:00	1.570000e+18	@Polytronindo kenapa video youtube di digital ...	ans_lbf	-1
3	2022-09-01 06:18:31+00:00	1.570000e+18	@Indosiar adakah kesalahan jaringan?soalnya di...	rosela_harum	-1
4	2022-11-02 23:21:44+00:00	1.590000e+18	@LUIPEMI kamu lupa ya?,kan migrasi besar besar...	KKholiq	1
...	...	...	...	...	...
1262	2022-08-26 13:29:34+00:00	1.560000e+18	keep fighting terus ini sampai titik darah pen...	Moektitto	1
1263	2022-08-26 13:27:08+00:00	1.560000e+18	Nek nonton nang vidio sakjane gak nurangi blas...	Moektitto	-1
1264	2022-08-15 00:49:38+00:00	1.560000e+18	@RRwfestealer @RawonStats @emosijiwakucom Kal...	GM_MAHARDIKA	-1
1265	2022-08-13 10:55:46+00:00	1.560000e+18	@YRadianito @kemkominfo @PlateJohnny Waduh agak...	UncleBACK165	-1
1266	2022-08-13 07:46:02+00:00	1.560000e+18	Masalah Umum TV Tidak Ada Sinyal dan Cara Meng...	MOTORSIA	1

1267 rows x 5 columns

## 2. Crawling Data

Data yang dikumpulkan diperoleh dari sebuah *tweet* pada media sosial *twitter*. Proses pengumpulan menggunakan *google colab* dengan cara *crawling* data dengan sebuah *query* “ Migrasi Tv Digital” dan terkumpul sebanyak 1267 *tweet*, sejak 29 November 2022 hingga 30 Desember 2022.

### 3.3.2 Text processing

Setelah melakukan proses pengumpulan data yang didapatkan merupakan data yang masih mentah dengan masih banyak kalimat-kalimat yang *noise* atau tidak terstruktur, oleh karena itu dibutuhkan *text processing* yang berguna untuk merubah data yang tidak struktur menjadi terstruktur. Tahapan pada *text processing* mempunyai tahapan sebagai berikut :

#### 1. Case Folding

Gambar 3. 4 Contoh data dengan kata kunci Migrasi Tv Digital

*Case Folding* adalah proses untuk menyamaratakan huruf kapital menjadi huruf kecil (*lowercase*). Tujuan dari proses *case folding* yaitu untuk

mempermudah pencarian dikarenakan semua dokumen teks tidak konsisten dengan hurufnya.

Tabel 3. 1 Contoh *Case Folding*

<i>Input</i>	<i>Output</i>
@indosiar adakah kesalahan jaringan?soalnya di tv digital saya indosiar tiba" gak ada padahal sebelum gak ada masalah sama sekali	adakah kesalahan jaringan soalnya di tv digital saya indosiar tiba gak ada padahal sebelum gak ada masalah sama sekali

## 2. *Cleaning*

*Cleaning* adalah proses dimana karakter-karakter tak diperlukan dalam teks, seperti tanda baca, *symbol* "@" yang digunakan sebagai nama pengguna, *hashtag* (#), *emoticon*, dan URL situs web dihilangkan dan juga pembersihan data yang tidak dibutuhkan dari hasil *crawling* data.

Tabel 3. 2 Contoh *Cleaning*

<i>Input</i>	<i>Output</i>
@indosiar adakah kesalahan jaringan?soalnya di tv digital saya indosiar tiba" gak ada padahal sebelum gak ada masalah sama sekali	adakah kesalahan jaringan soalnya di tv digital saya indosiar tiba gak ada padahal sebelum gak ada masalah sama sekali

## 3. *Tokenizing*

*Tokenizing* adalah metode untuk melakukan pemisahan kata dalam suatu kalimat dengan tujuan menjadikan kata menjadi lebih bermakna.

Tabel 3. 3 Contoh *Tokenizing*

<i>Input</i>	<i>Output</i>
adakah kesalahan jaringan soalnya di tv digital saya indosiar tiba gak ada padahal sebelum gak ada masalah sama sekali	['adakah', 'kesalahan', 'jaringan', 'soalnya', 'di', 'tv', 'digital', 'saya', 'indosiar', 'tiba', 'gak', 'ada', 'masalah sama sekali']

---

'padahal', 'sebelum', 'gak', 'ada',  
'masalah', 'sama', 'sekali']

---

#### 4. Normalization

*Normalization* adalah proses untuk mempermudah dalam analisis sentimen terhadap kata-kata yang tidak baku di dalam sebuah dokumen teks, contohnya singkatan kata.

Tabel 3. 4 Contoh *Normalization*

<i>Input</i>	<i>Output</i>
['adakah', 'kesalahan', 'jaringan', 'soalnya', 'di', 'tv', 'digital', 'saya', 'indosiar', 'tiba', 'gak', 'ada', 'padahal', 'sebelum', 'gak', 'ada', 'masalah', 'sama', 'sekali']	['adakah', 'kesalahan', 'jaringan', 'soal', 'di', 'televisi;', 'digital', 'saya', 'indosiar', 'tiba', 'tidak;', 'ada', 'padahal', 'sebelum', 'tidak;', 'ada', 'masalah', 'sama', 'sekali']

#### 5. Stopwords Removal

*Stopword Removal* adalah proses pembuangan kata-kata yang tidak memiliki arti atau tidak penting. Contoh *stopwords* Bahasa Indonesia antara lain ini, itu, yang, ke, di, dll. Tujuan dari penggunaan *stopwords* yaitu menghilangkan kata-kata informasi rendah dari teks, agar memudahkan kita fokus kepada kata-kata yang penting.

Tabel 3. 5 Contoh *Stopword Removal*

<i>Input</i>	<i>Output</i>
['adakah', 'kesalahan', 'jaringan', 'soal', 'di', 'televisi;', 'digital', 'saya', 'indosiar', 'tiba', 'tidak;', 'ada', 'padahal', 'sebelum', 'tidak;', 'ada', 'masalah', 'sama', 'sekali']	['adakah', 'kesalahan', 'jaringan', 'soal', 'televisi;', 'digital', 'indosiar', 'tidak;']

#### 6. Stemming

*Stemming* adalah menghilangkan imbuhan pada kata, dan membuat kata yang ada imbuhan menjadi kata dasar.

Tabel 3. 6 Contoh *Stemming*

Input	Output
['adakah', 'kesalahan', 'jaringan', 'soal', 'televisi;', 'digital', 'indosiar', 'tidak;', 'tidak;']	['ada', 'salah', 'jaringan', 'soal', 'televisi', 'digital', 'indosiar', 'tidak']

## 7. Pelabelan Data

Pelabelan data adalah proses penandaan pada data untuk menentukan kelas pada data tersebut. Contohnya sentimen negatif dan positif.

Teks	Label
jepang sudah pindah televisi digital indonesia tidak heran sumber daya saja gin geram banget tawa sepele darurat baca saja	-1
...	...
umum televisi sinyal atas pasti televisi digitalya	1

### 3.3.3 Implementasi

Setelah dilakukan tahapan *text processing* dan pelabelan data, kemudian data diimplementasikan dengan tahapan sebagai berikut :

#### 1. *TF-IDF*

*TF-IDF* (*Term Frequency – Inverse Document Frequency*) adalah metode algoritma yang menghitung nilai *Term Frequency* (*TF*) dan *Inverse Document Frequency* (*IDF*) pada setiap *token* (kata) di setiap dokumen dalam korpus (Hendra, 2021). Jadi dapat disimpulkan metode *TF-IDF*

digunakan untuk mengetahui kata yang sering muncul dalam sebuah dokumen.

*TF-IDF* dapat dihitung dengan rumus berikut

$$Wt = tf \times idf$$

$$IDF = \log \frac{N}{DF}$$

Keterangan :

TF : Jumlah kata dalam setiap dokumen

DF : jumlah total kata dari seluruh dokumen

N : jumlah total *dataset*

IDF : jumlah dokumen dalam Kumpulan yang mengandung kosa kata

## 2. Klasifikasi *Naive Bayes*

Setelah proses proses *TF-IDF* atau pembobotan kata, lalu data di klasifikasi menggunakan algoritma *naïve bayes*. Klasifikasi adalah metode untuk mengelompokkan dan menemukan *itemset* sesuai dengan nilai bobot yang ada dari setiap *itemset*. Dalam penelitian ini data yang dikelompokkan dibagi menjadi dua bagian yaitu *positif*, dan *negatif*. Algoritma *naïve bayes* dapat dihitung dengan rumus berikut :

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Dimana:

$P(X|H)$  : Probabilitas X berdasarkan kondisi pada hipotesis H

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu *class* spesifik

$P(H|X)$  : Probabilitas hipotesis H berdasarkan kondisi X

$P(H)$  : Probabilitas hipotesis H

$P(X)$  : Probabilitas X

### 3.3.4 Pengujian

Setelah melalui tahapan-tahapan sebelumnya pada pengujian ini menggunakan metode *Confusion Matrix*. *Confusion Matrix* merupakan salah satu metode yang dapat melakukan perbandingan tingkat akurasi dan dapat digunakan untuk menguji akurasi dalam model klasifikasi. Oleh karena itu pengujian ini bertujuan untuk mengetahui performa algoritma yang digunakan dan metode *Confusion Matrix* sangat membantu untuk menilai kinerja dari analisis klasifikasi.

Tabel 3. 7 *Confusion Matrix*

Kelas	Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Pengujian dilakukan dengan menghitung nilai *accuracy*, *precision* dan *recall* yang dapat ditampilkan dalam bentuk persentase.

#### 1. *Accuracy*

*Accuracy* merupakan tingkat kedekatan nilai prediksi dengan nilai aktual (sebenarnya). Untuk menghitung nilai *accuracy* digunakan persamaan dibawah ini :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

#### 2. *Precision*

*Precision* merupakan rasio prediksi benar *positif* dibandingkan dengan keseluruhan hasil yang diprediksi *positif*. Untuk menghitung nilai *precision* digunakan persamaan dibawah ini :

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

#### 3. *Recall*

*Recall* merupakan rasio prediksi benar *positif* dibandingkan dengan keseluruhan data yang benar *positif*. Untuk menghitung nilai *recall* digunakan persamaan dibawah ini :

$$Recall = \frac{TN}{FN + TN} * 100\%$$

