

BAB III

METODE PENELITIAN

3.1 Objek Penelitian

Objek Penelitian pada tugas akhir ini yaitu pembelajaran daring pada masa pandemi covid-19 Pengambilan bahan objek penelitian ini di ambil dari opini warga media sosial *witter*.

3.2 Alat dan Bahan Penelitian

Pada tahap ini menjelaskan tentang alat dan bahan apa saja yang digunakan saat penelitian

3.2.1 Alat Penelitian

Penelitian ini memakai perangkat keras dan perangkat lunak sebagai berikut :

1. Perangkat Keras (*Hardware*)
 - a. Laptop (Processor Intel Core i3-3200, 6GB RAM, 210 GB SSD, 512 HDD, Sistem Operasi Windows 10)
 - b. Mouse
 - c. Printer
2. Perangkat Lunak (*Software*)
 - a. *Twitter* (Sebagai media untuk mencari data)
 - b. *Rapidminer* (Untuk menarik data dari *Twitter*)
 - c. *Google Chorme* (Untuk mencari referensi jurnal maupun artikel)
 - d. *Microsoft Office* (Untuk membuat laporan dan proposal)
 - e. *Anaconda* (Sebagai *text editor* dalam penelitian ini)

3.2.1 Bahan Penelitian

Data yang digunakan dalam penelitian ini yaitu *tweet*/opini masyarakat terhadap konsep pembelajaran daring di masa pandemi Covid-19 dengan total data sebanyak 2587 *tweet* yang berbahasa Indonesia, hasil pengumpulan data ini disimpan ke dalam format *excel* (.xlsx).

3.3 Lokasi dan Waktu Penelitian

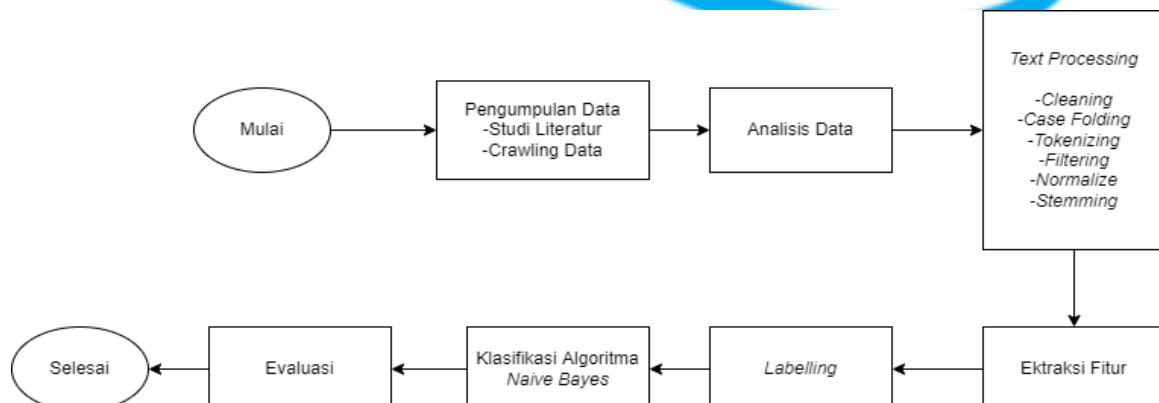
Penelitian dilakukan pada laboratorium riset dan teknologi Gedung A lantai II Universitas Buana Perjuangan Karawang. Kemudian pelaksanaan penelitian berlangsung sejak bulan September 2021 sampai bulan Mei 2022.

Tabel 3. 1 Waktu Penelitian

Kegiatan	1	2	3	4	5	6	7	8
Pengumpulan Data								
Analisis Data								
Preprocessing								
Labelling								
Klasifikasi								
Evaluasi								

3.4 Prosedur Penelitian

Prosedur penelitian yang digambarkan dengan menggunakan bentuk *Flowchart* dalam Gambar 3.1. Penelitian ini dimulai dari Analisis data lalu pengumpulan data yaitu pengumpulan studi literatur untuk mencari referensi dan *Crawling data* menggunakan API Twitter di *Rapidminer Studio*. Kemudian data yang sudah terkumpul akan di proses pada bagian *Preposcesing* yang terdiri dari dari proses *Case Folding*, *Tokenizing*, *Filtering*, *Stemming* untuk membersihkan data, setelah data bersih yaitu lalu ke tahap Ekstraksi Fitur dan di lanjut dengan pemberian label data dalam kategori kelas yang berbeda yaitu kelas positif, negatif dan netral, kemudian data yang sudah di klasifikasi akan implementasikan kepada algoritma *Naive Bayes* kemudian dilanjutkan ke tahap evaluasi yaitu pengujian model klasifikasi dengan menggunakan *Confusion Matrix*.



Gambar 3. 1 Prosedur Penelitian

3.5 Pengumpulan Data

Pengumpulan data dari penelitian ini dibagi menjadi dua bagian antara lain :

3.5.1 Studi Literatur

Studi literatur dilakukan sebagai pencarian landasan teori dan untuk mendapatkan referensi dari beraneka macam jurnal, buku, dan juga internet seperti mencari pengertian pengertian terhadap untuk tinjauan pustaka, penelitian terdahulu, metode yang di pakai

3.5.2 Crawling Data

Proses mengumpulkan data dengan mengambil data *tweet* yang dikumpulkan menggunakan *software Rapidminer Studio*, pencarian dengan kata kunci “Pembelajaran Online”, “Sekolah Online”, “Kuliah Online”, “Sekolah Daring” dan “Kuliah Daring”. Data diambil dengan total 2587 *tweet*, sejak awal mula penerapan Pembelajaran daring yaitu pada 16 Maret 2020 hingga 15 Januari 2022. Data kemudian dibagi dua data latih sejumlah 70% dan data uji sejumlah 30% yang di klasifikasi dengan Algoritma *Naive Bayes*.

Row No.	Id	Text ↓
618	1483791008...	hari ini daring , tadi aku cuma nyeleseini tugas agama sama mtkw aja soalnya tadi ekskul luring di sekolah , dan ini baru sampe rum...
61	1486550773...	hah ini sekolah daring lagi?
608	1483809778...	gw sih terserah mau sekolah daring atau ptm asal JANGAN ADA TUGAS KELOMPOK AJG
416	1484534338...	guej dong sekolah tapii besok jatah gue daring si
606	1483815939...	gue lagi ada di masa 'yaudah lah masuk terus ga perlu ada daring daring lagi nanggung' kek percuma juga kalo daring lagi sekolah t
83	1486495600...	good mwonee, yang hari ini sekolah/daring semangat yaa, yang kerja juga semangat, yang ga ngapa ngapain juga semangat
191	1485973538...	gapapaaa deh sekolah bakal daring lagi, yang penting sekolah gua wajib ada perpisahan anj gamau tau gua sial
76	1486505977...	ga kebayang kalo sekolah gue dadakan juga (walau emg suka dadakan) gue lagi di jalan, otomatis ga buka hape, nyampe sekolah k
615	1483795058...	dulu males daring

Gambar 3. 2 Contoh data dengan kata kunci sekolah daring

12	1486677631...	kok ya hybrid... tapi cuma satu matkul doang sih. gua mau ngisi kuliah full daring tapi takut ngga punya temen kalo gak mas...
67	1486174697...	kok malah deket2 mau masuk gini malah kesibukan baru dilimpahkan semua yaaa... aku benci itu ♥👀
185	1484239834...	kesel banget kalo ditanyain temen beda kampus tentang jadinya kuliah luring/daring/hybrid, kenapa sih unesa blm pasti ...
79	1485962905...	kerasa banget ya kalo udah luring gini, dikit banget ilmu yang nyangkut di otak selama kuliah daring 🤔
23	1486572435...	kehidupan maba era pandemi pertama kali ngerasain kuliah offline & ketemu temen kelas di RL setelah sekian bulan kulia...
53	1486293681...	kalo kuliah hybrid seribet itu mending aku pilih kuliah daring aja sii kayaknya
110	1485608178...	kalo ada hal yang sia sia tapi tetep dijaninin mungkin salah satunya ya kuliah daring
189	1484183351...	jujur marah bgt karena angkatan 2019 daring lagi kaya ya Allahhhhhhhhhhh BISA GAK SIH LU SEMUA DIEM DIRUMAH BIAR...
108	1485618076...	ini malah tulisannva hvbrid doano. beda kalo va online va emano darina ot. Ni hvbrid mksdnva apaan dah tolonga diberikan s...

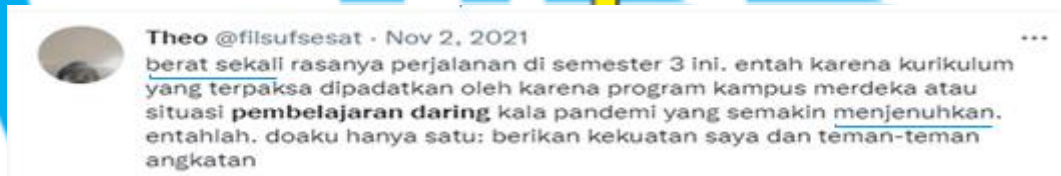
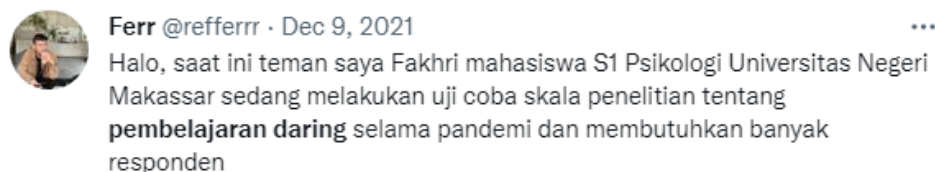
Gambar 3. 3 Contoh data dengan kata kunci kuliah online

1484134886...	ukt gratis 2 semester, pembelajaran onien (sisa matkui yg ga begitu krusial), dapet kuota internet selama 1 tahun, proses dim...
1483796891...	tambahan poin
1484885483...	semoga pihak kampus mikirin mateng2 deh keputusan online/offline buat pembelajaran kuliah semester ini, berhubung kasu...
1484530764...	semester kemarin temen aku ambil MK pilihan, nah aku baru mau ambil MK pilihan itu di semester ini dan semester ini kan b...
1486177785...	semangatin aku dongg, pembelajaran dari semester 1 online, tp uas semester 3 ini offline🙄🙄
1486161704...	sejauh ini menurut kalian enak pembelajaran online atau offline?

Gambar 3. 4 Contoh data dengan kata kunci Pembelajaran Online

3.6 Analisis Data

Analisis data dilakukan untuk mempelajari permasalahan yang terdapat pada media sosial *Twitter* yaitu opini masyarakat tentang pembelajaran daring. Analisis data dilakukan untuk mengetahui sentimen positif, negatif, dan netral. Contoh *tweet* sentimen positif, negatif, netral.

Gambar 3. 5 Contoh *tweet* negatifGambar 3. 6 Contoh *tweet* positifGambar 3. 7 Contoh *tweet* netral

3.7 Text Processing

Text processing merupakan tahap sebelum klasifikasi yaitu untuk mengganti kata yang tidak terstruktur menjadi kata yang terstruktur sehingga mempermudah pemrosesan dataset kata tersebut. Tahap ini melakukan *Cleansing*, *Case Folding*, *Tokenizing*, *Filtering*, *Normalize* dan *Stemming*

3.7.1 Cleaning

Cleaning yaitu tahap untuk menghapus karakter non-alfabetis dan menghilangkan tanda baca, simbol-simbol seperti tanda '@' untuk *username*, *hashtag* (#), *emoticon* dan url dari situs web.

Tabel 3. 2 *Cleansing*

Sebelum	Sesudah
Pengalaman saat menerima pembelajaran mata kuliah awal rasanya biasa aja sih, lama kelamaan seru juga ditambah pembelajaran daring yg membuat semakin seru #kultwit #literateproject40 #prodiperpustakaan #vokasiunair	Pengalaman saat menerima pembelajaran mata kuliah awal rasanya biasa aja sih, lama kelamaan seru juga ditambah pembelajaran daring yg membuat semakin seru

3.7.2 Case Folding

Case Folding yaitu mengubah huruf menjadi *lowercase* atau huruf kecil.

Tabel 3. 3 *Case Folding*

Sebelum	Sesudah
Pengalaman saat menerima pembelajaran mata kuliah awal rasanya biasa aja sih, lama kelamaan seru juga ditambah pembelajaran daring yg membuat semakin seru	pengalaman saat menerima pembelajaran mata kuliah awal rasanya biasa aja sih, lama kelamaan seru juga ditambah pembelajaran daring yg membuat semakin seru

3.7.3 Tokenizing

Tokenizing merupakan tahap memisahkan string atau kata terhadap suatu teks berdasarkan tiap kata pada kalimat kalimat yang terdapat dari *tweet*

Tabel 3. 4 *Tokenizing*

Sebelum	Sesudah
Pengalaman saat menerima pembelajaran mata kuliah awal rasanya biasa aja sih, lama kelamaan seru juga ditambah pembelajaran daring yg membuat semakin seru	'pengalaman' 'saat' 'menerima' 'pembelajaran' 'mata' 'kuliah' 'awal' 'rasanya' 'biasa' 'aja' 'sih' 'lama' 'kelamaan' 'seru' 'juga' 'ditambah' 'pembelajaran' 'daring' 'yg' 'membuat' 'semakin' 'seru'

3.7.4 *Stopwords*

Menghilangkan kata yang kurang relevan atau tidak penting/tidak bermakna pada analisis sentimen tahap ini juga di kenal sebagai *stopword removal*

Tabel 3. 5 *Filtering*

Sebelum	Sesudah
pengalaman saat menerima pembelajaran mata kuliah awal rasanya biasa aja sih lama kelamaan seru juga ditambah pembelajaran daring yg membuat semakin seru	pengalaman menerima pembelajaran mata kuliah awal rasanya biasa kelamaan seru ditambah pembelajaran daring semakin seru

3.7.5 *Normalize*

Proses *normalize* yaitu mengubah kata, kesalahan dalam pengetikan (*typo*) menjadi kata yang mudah di pahami atau sesuai KBBI

Tabel 3. 6 *Normalize*

Sebelum	Sesudah
pengalaman menerima pembelajaran mata kuliah awal rasanya biasa kelamaan seru ditambah pembelajaran daring semakin seru	pengalaman menerima pembelajaran mata kuliah awal rasanya biasa kelamaan seru ditambah pembelajaran daring semakin seru

3.7.6 *Stemming*

Stemming adalah tahap menghapus imbuhan kata atau mengubah kata menjadi kata dasar pada kata dari hasil *normalize*

Tabel 3. 7 *Stemming*

SEBELUM	SESUDAH
pengalaman menerima pembelajaran	pengalaman terima belajar mata
mata kuliah awal rasanya biasa kelamaan	kuliah awal rasa biasa lama seru
seru ditambah pembelajaran daring	tambah belajar daring makin seru
semakin seru	

3.8 Ekstraksi Fitur

Pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan tahap mengubah data teks menjadi data numerik. TFIDF ini yaitu sebuah ukuran statistik yang digunakan untuk mengetahui banyaknya sebuah kata di dalam dataset.

Adapun rumus dari pembobotan kata TD-IDF adalah:

$$W_{t,d} = t_{ft,d} \times id_{ft} = t_{ft,d} \times \log \frac{N}{d_{ft}} \quad (6)$$

Keterangan :

$W_{t,d}$ = Bobot TF-IDF

$t_{ft,d}$ = Jumlah frekuensi kata

id_{ft} = Jumlah inverse frekuensi dokumen tiap kata

d_{ft} = Jumlah frekuensi dokumen tiap kata

N = Jumlah total dataset

Pembobotan kata dengan TF-IDF ini adalah perkalian dari nilai TF dan IDF yang akan mendapatkan bobot lebih kecil jika kata tersebut selalu muncul pada setiap dataset, sebaliknya bobot TF-IDF akan lebih besar jika kata tersebut tidak sering muncul pada setiap *dataset*.

Tweet 1 : "bingung" "tugas" "video" "sekolah"											
Tweet 2: "bosan" "rasa" "sekolah"											
Tweet 3: "sumpah" "jadwal" "bingung"											
Token	TF			TF Normalisasi			DF	IDF	TF-IDF		
	D1	D2	D3	D1	D2	D3			D1	D2	D3
bingung	1	0	1	0,25	0	0	2	0,176	0,0440	0	0,0587
tugas	1	0	0	0,25	0	0	1	0,477	0,1193	0	0
video	1	0	0	0,25	0	0	1	0,477	0,1193	0	0
sekolah	1	1	0	0,25	0,33	0	2	0,176	0,0440	0,0587	0
sumpah	0	1	0	0	0,33	0	1	0,477	0	0,1590	0
jadwal	0	0	1	0	0	0,33	1	0,477	0	0	0,1590
bosan	0	1	0	0	0,33	0	1	0,477	0	0,1590	0
rasa	0	1	0	0	0,33	0	1	0,477	0	0,1590	0
Panjang Dokumen	4	3	3								

Gambar 3. 8 Contoh Perhitungan TF-IDF

3.9 Labelling Data

Setelah *text processing* dan pembobotan TF-IDF tahap selanjutnya *labelling* atau Pelabelan pemberian label pada setiap data dalam kategori kelas yang berbeda yaitu kelas positif, netral dan negatif.

3.10 Klasifikasi Naive Bayes

Klasifikasi adalah suatu pengkategorian data dimana data yang digunakan tersebut mempunyai kelas atau label. *Naive Bayes* adalah pengklasifikasi probabilistic yang tidak rumit berguna untuk menghitung sekumpulan probabilitas dengan menambahkan frekuensi dan kombinasi nilai dalam kumpulan data yang diberikan, keunggulan *Naive Bayes* ini hanya memerlukan sedikit data latih untuk menentukan estimasi parameter yang diperlukan dalam proses klasifikasi. Persamaan dari teorema Bayes adalah :

$$P(H|X) = \frac{P(X|K).P(K)}{P(X)} \quad (7)$$

Keterangan :

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu *class* spesifik

P(H|X) : Probabilitas hipotesis H berdasar kondisi X

P(H) : Probabilitas hipotesis H

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

P(X) : Probabilitas X

3.11 Evaluasi

Pada proses terakhir yaitu evaluasi, penelitian ini menggunakan *Confusion Matrix* untuk melakukan tahapan pengujian atau evaluasi. *Confusion Matrix* merupakan salah satu metode yang dapat dipakai untuk mengukur kinerja suatu metode klasifikasi. Pengujian dilakukan dengan menghitung *accuracy*, *recall*, *precision*, *f1-score* dan ditampilkan dalam bentuk persentase, terdapat 6 istilah yaitu *True Positif* (TP), *True Negatif* (TN), *True Netral* (TNt), *False Positif* (FP), *False Negatif* (FN), dan *False Netral* (FNt). *Confusion Matrix* dapat disajikan seperti pada Tabel 2.2

Tabel 3.7 *Confusion matrix*

Kelas	Positif	Netral	Negatif
Positif	True Positif (TP)	False Netral (FNt)	False Negatif (FN)
Netral	False Positif (FP)	True Netral (TNt)	False Negatif (FN)
Negatif	False Positif (FP)	False Netral (FNt)	True Negatif (TN)

Keterangan untuk tabel diatas yaitu:

TP (*True Positif*) = Data positif terklasifikasi dengan benar oleh sistem.

FNt (*False Netral*) = Data netral namun terklasifikasi salah oleh sistem.

FN (*False Negatif*) = Data negatif namun terklasifikasi salah oleh sistem.

FP (*False Positif*) = Data positif namun terklasifikasi salah oleh sistem

TNt (*True Netral*) = Data netral namun terklasifikasi benar oleh sistem.

TN (*True Negatif*) = Data negatif yang terklasifikasi dengan benar oleh sistem.

Pengujian dilakukan dengan menghitung nilai *accuracy*, *precision*, *recall* dan *f1-score* lalu ditampilkan dalam bentuk persentase

1. Accuracy

Accuracy merupakan mengetahui jumlah data yang di klasifikasikan secara benar. Berikut rumus perhitungan *Accuracy*

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FN+TN} \times 100 \quad (8)$$

2. Precision

Precision merupakan perbandingan jumlah data relevan yang ditemukan terhadap jumlah data yang ditemukan. Berikut rumus perhitungan *Precision*

$$\text{Precision} = \frac{TP}{TP+FP} \quad (9)$$

3. *Recall*

Jumlah data yang skala pada jumlah teks yang relevan dikalangan dokumen teks pada koleksi. Berikut rumus perhitungan *Recall*

$$Recall = \frac{TP}{TP+FP} \quad (10)$$

4. *F1-Score*

F1-Score merupakan parameter tunggal ukuran keberhasilan retrieval yang menggabungkan *recall* dan precision

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (11)$$



