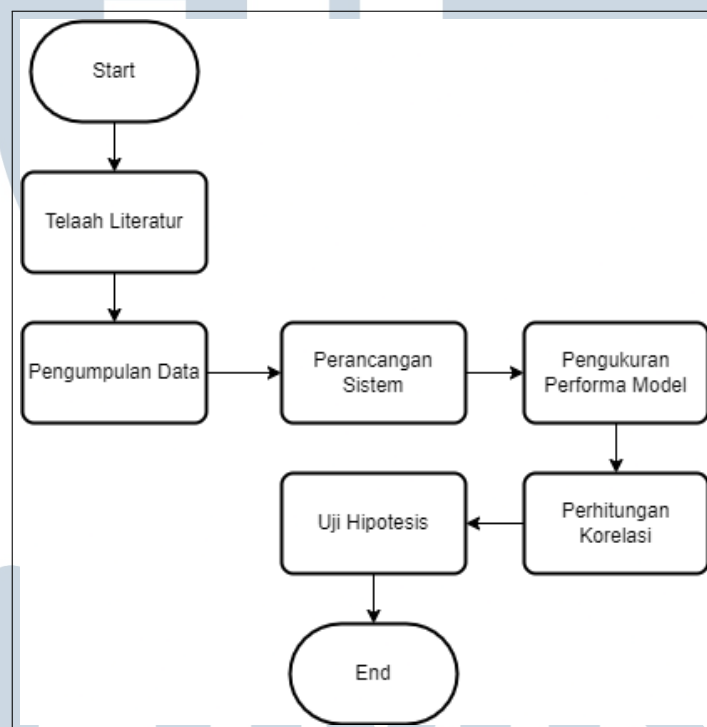


## BAB 3 METODOLOGI PENELITIAN

### 3.1 Metodologi Penelitian

Metodologi penelitian digambarkan melalui Gambar 3.1. Metodologi terdiri atas telaah literatur, proses pengumpulan data, perancangan sistem, evaluasi model, perhitungan korelasi, dan uji hipotesis.



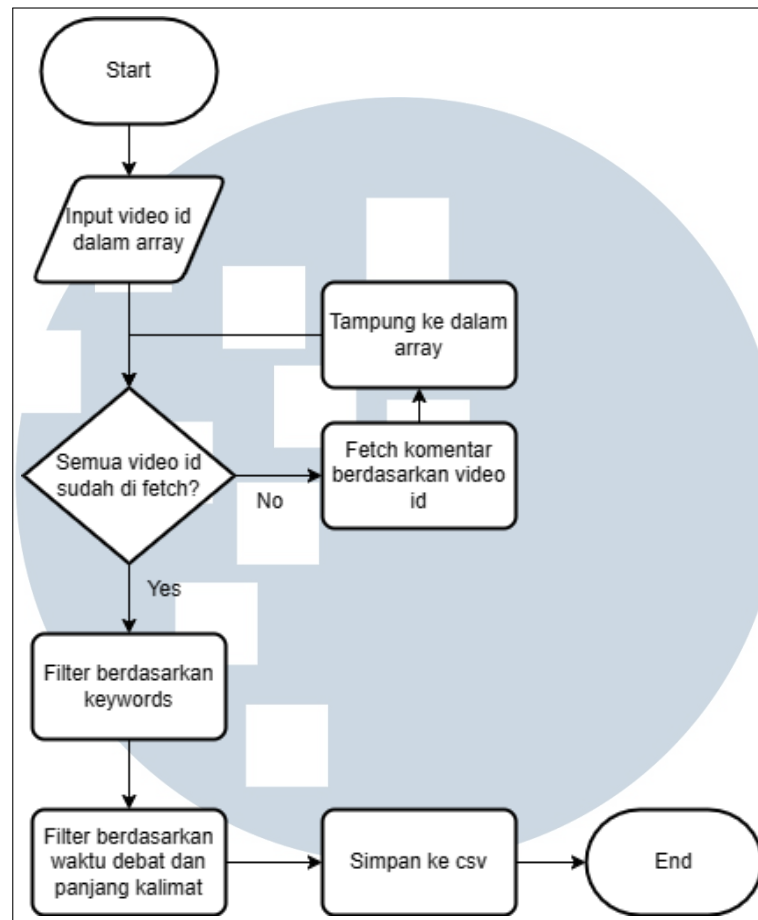
Gambar 3.1. Flowchart metodologi

#### 3.1.1 Telaah Literatur

Telaah literatur dilakukan dengan mengumpulkan teori-teori yang terkait dengan penelitian yang dilakukan.

#### 3.1.2 Teknik Pengumpulan Data

Alur pengumpulan data menggunakan Youtube Data API digambarkan melalui *flowchart* pada Gambar 3.2.



Gambar 3.2. Flowchart pengumpulan data

Gambar 3.2 menunjukkan *flowchart* pengumpulan data. Pengumpulan data dilakukan pada kolom komentar Youtube. Pengumpulan data dilakukan dengan Youtube Data API untuk mengambil komentar pada video debat pilpres berdasarkan video id. Data komentar di-filter berdasarkan *keywords* setiap paslon. Berikut merupakan *keywords* untuk setiap paslon.

1. paslon 1: anies, imin, amin
2. paslon 2: prabowo, gibran, pragib
3. paslon 3: ganjar, mahfud, gama

Selain *keywords*, data di-filter berdasarkan waktu survei elektabilitas dilakukan di mana hanya komentar yang diperbarui sebelum waktu survei berakhir yang diambil dan berdasarkan jumlah kata di bawah 50 kata. Filter berdasarkan jumlah kata dilakukan untuk memaksimalkan performa model dan menghapus

komentar yang tidak berisi opini terhadap hasil debat pilpres. Data komentar disimpan ke dalam csv. Data yang diambil total berjumlah 17872 komentar yang terbagi menjadi 15 *dataset*.

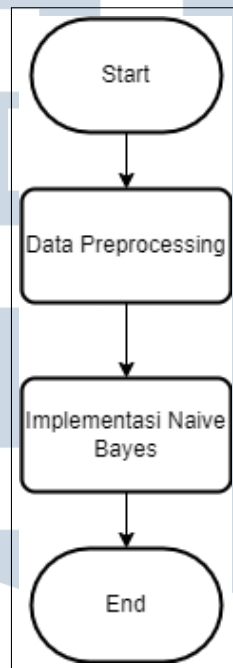
Pengumpulan data diambil dari kanal Youtube KPU dan Najwa Shihab pada 11 April 2024. Tabel 3.1 menunjukkan video yang digunakan dalam melakukan pengumpulan data dan jumlah data yang didapatkan.

Tabel 3.1: Daftar video

Debat	Kanal	Judul	Jumlah Data
Debat capres 1	KPU RI	Debat Pertama Calon Presiden Pemilu Tahun 2024	7141
	Najwa Shihab	[LIVE] Musyawarah Nobar Debat Pilpres 2024   Musyawarah	
Debat cawapres 1	KPU RI	Debat Kedua Calon Wakil Presiden Pemilu Tahun 2024	3825
	Najwa Shihab	[LIVE] Part 2 Nobar Debat Cawapres 2024   Musyawarah	
Debat capres 2	KPU RI	Debat Ketiga Calon Presiden Pemilu Tahun 2024	2986
	Najwa Shihab	[FULL] Debat Capres 2024, Nobar Debat Ronde Ketiga di Musyawarah   Musyawarah	
Debat cawapres 2	KPU RI	Debat Keempat Calon Wakil Presiden Pemilu Tahun 2024	1972
	Najwa Shihab	[FULL] Debat Cawapres 2024, Nobar Debat Ronde Keempat di Musyawarah	
Debat capres 3	KPU RI	Debat Kelima Calon Presiden Pemilu Tahun 2024	1948
	Najwa Shihab	[FULL] Layar Tancap Mata Najwa, Nobar Debat Capres Ronde Kelima   Mata Najwa	
Total Data			17872

### 3.1.3 Perancangan Sistem

Alur perancangan sistem digambarkan melalui *flowchart* pada Gambar 3.3. Perancangan terdiri atas data *preprocessing* dan implementasi model dengan algoritma Naive Bayes.

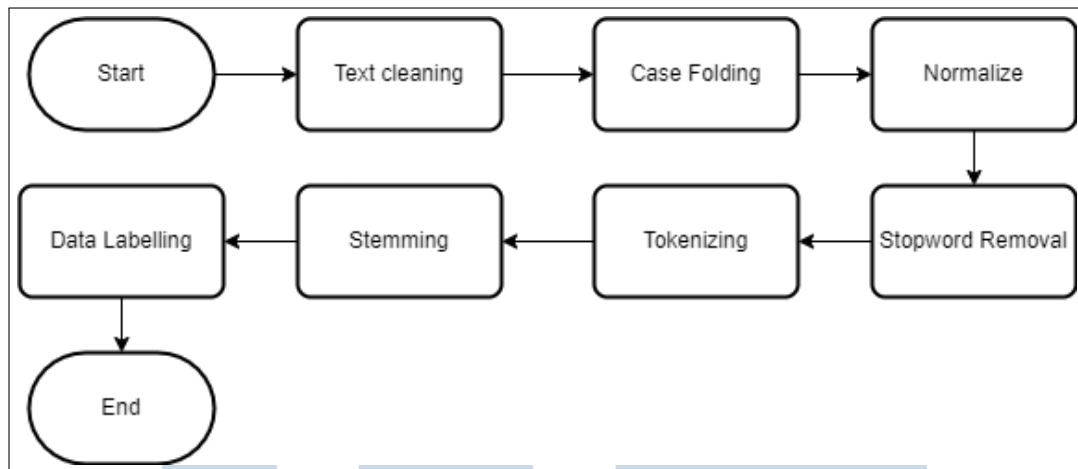


Gambar 3.3. Flowchart perancangan sistem

#### A Data Preprocessing

Alur data *preprocessing* ditunjukkan pada Gambar 3.4. Data *preprocessing* terdiri atas *text cleaning*, *case folding*, *normalize*, *stopword removal*, *tokenizing*, *stemming*, dan data *labelling*.

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA



Gambar 3.4. Flowchart data preprocessing

### A.1 Text Cleaning

Proses *text cleaning* dilakukan untuk membersihkan teks. Terdapat beberapa hal yang dilakukan dalam proses *text cleaning* yaitu:

1. menghapus *hashtag*, dan simbol *tag*
2. menghapus karakter non-alfanumerik
3. menghapus angka
4. menghapus tautan
5. menghapus kata yang hanya memiliki satu karakter
6. menghapus *whiteline* dan *newline*

### A.2 Case Folding

Proses untuk menyamakan *case* pada semua karakter dalam teks. *Case folding* mengubah semua karakter pada teks menjadi huruf kecil.

### A.3 Normalize

Proses untuk melakukan normalisasi pada teks. Proses normalisasi memperbaiki kata salah ketik dan *slangword* menjadi bentuk baku. Kata-kata salah ketik dan *slangword* dimasukkan ke dalam kamus berbentuk csv yang didapatkan dari *resource* Github [32] dan dengan melakukan pengecekan kata pada *dataset*.

#### A.4 Stopword Removal

Proses untuk menghilangkan kata yang tidak penting dan tidak relevan dalam kalimat. Kata-kata *stopword* didapatkan dari *library* Sastrawi dengan beberapa penyesuaian. Penyesuaian berupa penghilangan kata "tidak" pada *stopword* dan penambahan beberapa *stopword*. Penambahan *stopword* terbagi menjadi 2 macam yaitu penambahan *stopword* yang spesifik untuk masing-masing paslon dan penambahan *stopword* untuk semua paslon. Berikut merupakan *stopword* yang ditambahkan.

1. paslon 1: prabowo, gibran, ganjar, mahfud, ganjarmahfud, prabowogibran
2. paslon 2: anies, imin, ganjar, mahfud, aniesimin, ganjarmahfud
3. paslon 3: prabowo, gibran, anies, imin, aniesimin, prabowogibran
4. semua paslon: jakarta, jawa, timur, barat, tengah, indonesia, bapak, pak, mas, bpk, gus, cak, oppa, jokowi, opa, jokowidodo, abah, nya

#### A.5 Tokenizing

Proses untuk memenggal teks menjadi bagian-bagian yang disebut dengan *token*. Proses *tokenizing* menghasilkan sebuah array yang berisi setiap kata dalam suatu kalimat.

#### A.6 Stemming

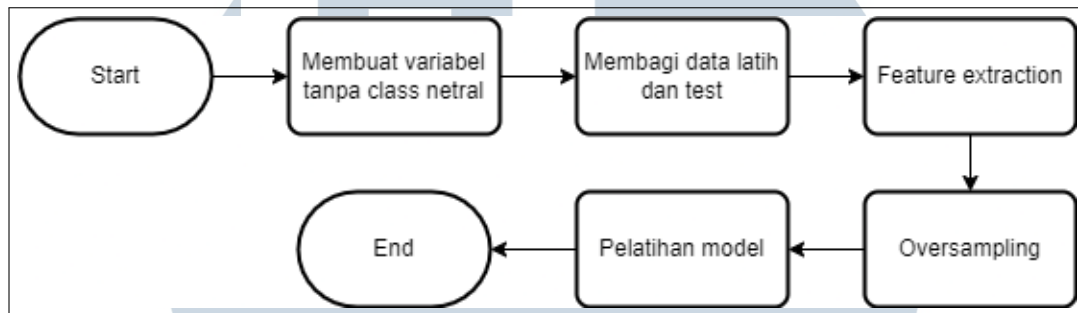
Proses *stemming* dilakukan untuk mengubah kata berimbuhan menjadi bentuk dasarnya. Proses *stemming* dilakukan menggunakan *library* Sastrawi.

#### A.7 Data Labelling

Proses data *labelling* dilakukan untuk memberi label sentimen pada data. Label sentimen terdiri dari label positif, netral, dan negatif. Proses data *labelling* dilakukan secara otomatis melalui VADER.

## B Implementasi Naive Bayes

Alur implementasi Naive Bayes ditunjukkan pada Gambar 3.5. Implementasi terdiri atas pembuatan variabel tanpa *class* netral, pembagian data latih dan tes, ekstraksi fitur, *oversampling*, dan pelatihan model.



Gambar 3.5. Flowchart implementasi Naive Bayes

### B.1 Pembuatan Variabel Tanpa Kelas Netral

Klasifikasi sentimen dilakukan pada 2 kondisi yaitu dengan *class* netral dan tanpa *class* netral. Untuk itu, membuat variabel baru yang menampung *dataset* tanpa kelas netral.

### B.2 Pembagian Data Latih dan Tes

Data dibagi menjadi data latih dan data uji. Data latih berjumlah 80% dan data uji berjumlah 20% dari total data pada *dataset*.

### B.3 Feature Extraction

Setelah data terbagi menjadi data latih dan data uji, dilakukan ekstraksi fitur menggunakan BOW untuk mengubah data teks menjadi sebuah vektor berisi frekuensi kata atau *token*.

### B.4 Oversampling

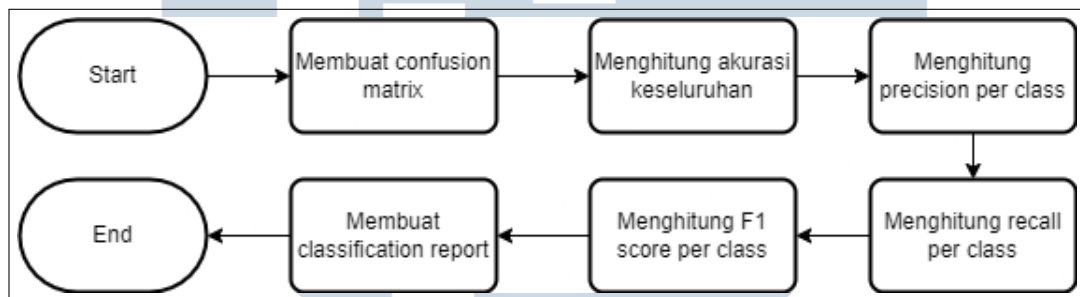
Setelah melakukan *feature extraction*, yang dilakukan selanjutnya yaitu melakukan *oversampling* pada *class* minoritas untuk mengatasi ketidakseimbangan data.

## B.5 Pelatihan Model

Setelah melakukan *overasampling*, dilakukan pelatihan model Naive Bayes. Naive Bayes melakukan klasifikasi sentimen pada data uji.

### 3.1.4 Evaluasi Model

Setelah melakukan perancangan sistem, dilakukan evaluasi model untuk mengukur performa model Naive Bayes dalam melakukan klasifikasi. Alur evaluasi model digambarkan melalui Gambar 3.6.

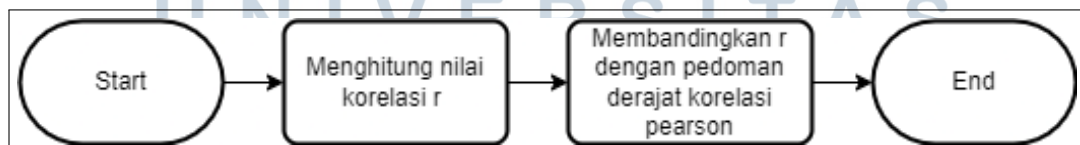


Gambar 3.6. Flowchart evaluasi

Gambar 3.6 menunjukkan alur evaluasi model. Evaluasi model dilakukan dengan membuat *confusion matrix*, menghitung nilai akurasi secara keseluruhan, *precision*, *recall*, *F1 score* untuk setiap *class*, dan membuat *classification report*.

### 3.1.5 Pehitungan Korelasi

Setelah melakukan evaluasi model, dilakukan perhitungan korelasi Pearson antara sentimen debat pilpres pada media sosial terhadap elektabilitas paslon. Alur perhitungan korelasi ditunjukkan melalui Gambar 3.7.



Gambar 3.7. Flowchart perhitungan korelasi

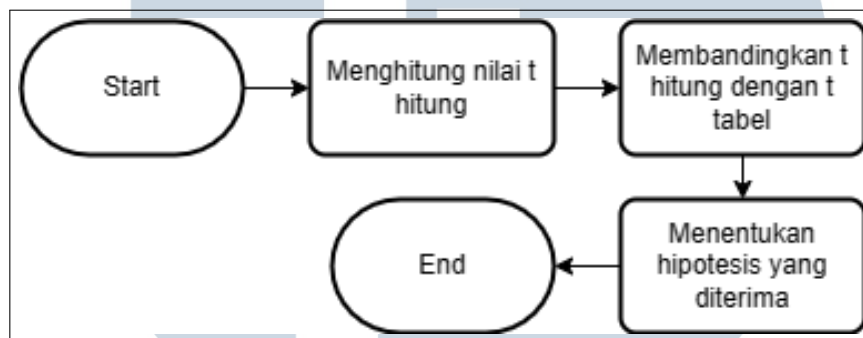
Gambar 3.7 menunjukkan alur perhitungan korelasi. Perhitungan korelasi dilakukan pada hasil persentase sentimen positif dan negatif pada *dataset* dengan *class* netral. Setelah didapatkan nilai korelasi *r*, dilakukan perbandingan nilai *r*



dengan pedoman derajat korelasi Pearson yang ditunjukkan pada Tabel 2.2 untuk mengetahui hubungan antara sentimen debat pilpres pada media sosial terhadap elektabilitas paslon.

### 3.1.6 Uji Hipotesis

Setelah mendapatkan derajat hubungan korelasi Pearson, dilakukan pengujian hipotesis. Alur pengujian hipotesis ditunjukkan melalui Gambar 3.8.



Gambar 3.8. Flowchart uji hipotesis

Gambar 3.8 menunjukkan alur uji korelasi. Uji korelasi dilakukan dengan menghitung nilai t hitung. Setelah didapatkan nilai t hitung, dilakukan perbandingan nilai t hitung dengan t tabel. Setelah itu, dilakukan penentuan hipotesis mana yang diterima. Hipotesis tersebut meliputi:

H<sub>0</sub>: tidak ada hubungan antara sentimen debat pilpres pada media sosial terhadap elektabilitas paslon.

H<sub>1</sub>: ada hubungan antara sentimen debat pilpres pada media sosial terhadap elektabilitas paslon.

Kriteria pengujian dengan taraf signifikansi 0,05 dan *degree of freedom* n-2 sebagai berikut.

1. Jika t hitung bernilai positif:

(a) Jika  $t_{hitung} > t_{tabel}$ , tolak H<sub>0</sub>

(b) Jika  $t_{hitung} < t_{tabel}$ , terima H<sub>0</sub>

2. Jika t hitung bernilai negatif:

(a) Jika  $-t_{hitung} < -t_{tabel}$ , tolak H<sub>0</sub>

(b) Jika  $-t_{hitung} > -t_{tabel}$ , terima H<sub>0</sub>