

**IMPLEMENTASI *WEB SCRAPING* UNTUK ANALISIS  
ULASAN FILM KKN DI DESA PENARI MENGGUNAKAN  
*NAÏVE BAYES CLASSIFIER***

**VEGA FOSWANTO**

**NIM. H1091191003**

**SKRIPSI**



**PROGRAM STUDI STATISTIKA  
JURUSAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS TANJUNGPURA  
PONTIANAK**

**2023**

**IMPLEMENTASI *WEB SCRAPING* UNTUK ANALISIS  
ULASAN FILM KKN DI DESA PENARI MENGGUNAKAN  
*NAÏVE BAYES CLASSIFIER***

**VEGA FOSWANTO**

**NIM. H1091191003**

**SKRIPSI**

Sebagai salah satu syarat untuk memperoleh gelar  
Sarjana Statistika pada Program Studi Statistika



**PROGRAM STUDI STATISTIKA  
JURUSAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS TANJUNGPURA  
PONTIANAK  
2023**

**IMPLEMENTASI *WEB SCRAPING* UNTUK ANALISIS ULASAN FILM KKN DI  
DESA PENARI MENGGUNAKAN *NAÏVE BAYES CLASSIFIER***

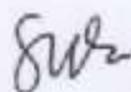
Tanggung Jawab Yuridis Material Pada



Vega Foswanto  
H1091191003

Disetujui Oleh

Pembimbing 1



Dr. Evy Sulistianingsih, M.Sc.  
NIP. 198502172008122006

Pembimbing 2



Hendra Perdana, S.Si., M.Sc.  
NIP. 198810102019031020

Disahkan  
Oleh

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Tanjungpura Pontianak



Dr. Gusrizal, S.Si., M.Si.  
NIP. 197108022000031001

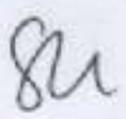
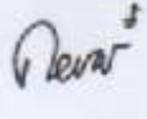
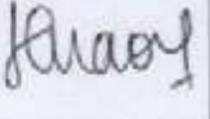
**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET DAN TEKNOLOGI  
UNIVERSITAS TANJUNGPURA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
PONTIANAK**

---

**TIM PENGUJI SKRIPSI**

**IMPLEMENTASI *WEB SCRAPING* UNTUK ANALISIS ULASAN FILM  
KKN DI DESA PENARI MENGGUNAKAN *NAÏVE BAYES CLASSIFIER***

**VEGA FOSWANTO  
H1091191003**

<b>NAMA/NIP</b>	<b>TIM PENGUJI</b>	<b>GOLONGAN/ JABATAN</b>	<b>TANDA TANGAN</b>
Dr. Evy Sulistianingsih, M.Sc. NIP. 198502172008122006	Pemimpin sidang merangkap anggota penguji	III/b Asisten Ahli	
Hendra Perdana, S.Si., M.Sc. NIP. 198810102019031020	Sekretaris sidang merangkap anggota penguji	III/b Tenaga Pengajar	
Neva Satyahadewi, S.Si., M.Sc., CRA., CRP., CRMP. NIP. 198212042005012001	Ketua penguji	III/d Lektor	
Naomi Nessyana Debataraja, M.Si. NIP. 198811232012122004	Anggota penguji	III/c Lektor	

Berdasarkan Surat Keputusan

Dekan Fakultas Matematika dan Ilmu Pengatahuan Alam

Universitas Tanjungpura Pontianak

Nomor : 1752/UN22.8/TD.06/2023

Tanggal : 19 Mei 2023

Tanggal Lulus : 23 Mei 2023

## HALAMAN PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi. Sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Pontianak, Mei 2023



Vega Foswanto

# IMPLEMENTASI *WEB SCRAPING* UNTUK ANALISIS ULASAN FILM KKN DI DESA PENARI MENGGUNAKAN *NAÏVE BAYES CLASSIFIER*

## INTISARI

Pada tahun 2022, film Indonesia dengan penonton paling banyak adalah film KKN di Desa Penari. Banyaknya penonton film tersebut menyebabkan banyaknya ulasan masyarakat terhadap film tersebut. Ulasan film KKN di Desa Penari dapat dilihat pada berbagai *website* dan pada penelitian ini sumber *website* yang digunakan adalah *twitter*. Metode yang digunakan di dalam penelitian ini menggunakan dua tahap, yaitu pengumpulan data menggunakan *Web Scraping* kemudian dilanjutkan analisis sentimen menggunakan *Naïve Bayes Classifier*. Data yang dikumpulkan menggunakan teknik *Web Scraping* adalah data ulasan film KKN di Desa Penari dari *twitter* pada hari pertama penayangan film tersebut di bioskop Indonesia yaitu pada tanggal 30 April 2022, kemudian penerapan *Naïve Bayes Classifier* untuk melakukan analisis sentimen dan melakukan klasifikasi terhadap ulasan film tersebut. Dari proses *Web Scraping* diperoleh jumlah *tweet* sebanyak 866 *tweets*. Data *tweet* yang diperoleh tidak seluruhnya adalah *tweet* yang membahas tentang ulasan film KKN di Desa Penari sehingga dilakukan seleksi data dan diperoleh data ulasan sebanyak 159 ulasan. Selanjutnya, data tersebut diolah dan dibersihkan untuk persiapan analisis sentimen menggunakan teknik dalam *data mining* yaitu *preprocessing* data. Dari proses *preprocessing* data didapat hasil sebanyak 127 kata yang berbeda untuk diproses menggunakan metode *Naïve Bayes Classifier*. Setelah data diolah dan dibersihkan, diperoleh hasil klasifikasi menggunakan metode *Naïve Bayes Classifier* diperoleh nilai akurasi sebesar 87,5% sehingga hasil klasifikasinya sangat baik dengan jumlah ulasan yang masuk ke dalam sentimen positif sebesar 69,18% dan sentimen negatif sebesar 30,82%.

**Kata kunci:** *Twitter*, Sentimen, *Preprocessing* Data

# **WEB SCRAPING IMPLEMENTATION FOR ANALYSIS OF KKN DI DESA PENARI MOVIE REVIEWS USING NAÏVE BAYES CLASSIFIER**

## **ABSTRACT**

In 2022, the Indonesian movie with the most viewers was KKN di Desa Penari. The large number of viewers of the movie caused many public reviews of the movie. Reviews of the KKN di Desa Penari movie can be seen on various websites and in this research the website source used is twitter. The method used in this research uses two stages, which are data collection using Web Scraping and then sentiment analysis using Naïve Bayes Classifier. The data collected using the Web Scraping technique is the review data of the film KKN in Desa Penari from twitter on the first day of the film's screening in Indonesian theaters on April 30, 2022, then the application of Naïve Bayes Classifier to conduct sentiment analysis and classify the film reviews. From the Web Scraping process, a total of 866 tweets were obtained. The tweet data obtained are not all tweets that discuss the KKN di Desa Penari movie reviews so that data selection is carried out and 159 reviews are obtained. Furthermore, the data is processed and cleaned in preparation for sentiment analysis using techniques in data mining, namely data preprocessing. From the data preprocessing process, 127 different words were obtained to be processed using the Naïve Bayes Classifier method. After the data is processed and cleaned, the classification results using the Naïve Bayes Classifier method obtained an accuracy value of 87.5% so that the classification results are very good with the number of reviews that fall into positive sentiment amounting to 69.18% and negative sentiment amounting to 30.82%.

**Keywords:** Twitter, Sentiment, Data Preprocessing

## KATA PENGANTAR

Puji dan syukur penulis ucapkan kehadiran Allah SWT karena berkat rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan penulisan skripsi yang berjudul “Implementasi *Web Scraping* Untuk Analisis Ulasan Film KKN di Desa Penari Menggunakan *Naïve Bayes Classifier*”. Tujuan penulisan skripsi ini adalah untuk memenuhi syarat dalam memperoleh gelar Sarjana Statistika pada Program Studi Statistika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura.

Dalam penulisan skripsi ini, tentu saja tidak lepas dari bimbingan, arahan, dan motivasi dari berbagai pihak. Maka dari itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua penulis, yaitu Ibu Tafiyah dan almarhum bapak saya yaitu Pak Samidi yang telah memberikan banyak dukungan dalam berbagai hal kepada penulis dan menjadi motivasi terbesar dalam kehidupan penulis.
2. Seluruh saudara penulis yaitu Budi Wijayanti, Krisuma Handayani, dan Tuwi Indah Pratiwi yang juga selalu memberikan banyak dukungan kepada penulis dalam berbagai hal.
3. Ibu Setyo Wira Rizki, S.Si., M.Sc., selaku dosen Pembimbing Akademik yang pertama yang telah memberikan bimbingan, dukungan, dan motivasi untuk penulis selama perkuliahan.
4. Ibu Wirda Andani, S.Si., M.Si., selaku dosen Pembimbing Akademik yang kedua yang telah memberikan arahan dan motivasi untuk penulis selama perkuliahan.
5. Ibu Dr. Evy Sulistianingsih, S.Si., M.Sc., selaku Dosen Pembimbing I yang telah memberikan arahan dan dukungan untuk menyelesaikan penulisan skripsi ini.
6. Bapak Hendra Perdana, S.Si., M.Sc., selaku Dosen Pembimbing II yang telah memberikan arahan, dukungan, dan motivasi untuk menyelesaikan penulisan skripsi ini.

7. Ibu Neva Satyahadewi, M.Sc., CRA., CRP., CRMP., selaku Dosen Penguji I yang telah memberikan saran dan masukan untuk perbaikan skripsi ini sehingga menjadi lebih baik.
8. Ibu Naomi Nessyana Debataraaja, M.Si., selaku Dosen Penguji II yang telah memberikan saran dan masukan untuk perbaikan skripsi ini sehingga menjadi lebih baik.
9. Teman-teman Statistika Angkatan 2019 yang telah kebersamai dan menjadikan kehidupan perkuliahan penulis menjadi penuh warna.
10. Seluruh rekan kerja Himasta FMIPA Untan 2021/2022 yang telah membantu penulis dalam menyelesaikan kehidupan organisasi kampus.
11. Saudara Wahyu Diyan Ramadana yang telah memberikan banyak motivasi agar penulis dapat menyelesaikan penulisan skripsi ini.
12. Rekan sesama *Web Scraping* yaitu saudara Frans Xavier dan Rahadi Ramlan yang telah saling membantu menemukan solusi ketika mengalami kesulitan dalam penulisan skripsi ini.
13. Serta semua pihak yang tidak dapat penulis sebutkan satu persatu.

Semoga Allah SWT memberikan banyak balasan atas bantuan yang telah diberikan dalam menyelesaikan skripsi ini dan semoga skripsi ini memberikan banyak manfaat bagi berbagai pihak yang membutuhkannya.

Pontianak, Januari 2023

Vega Foswanto

## DAFTAR ISI

<b>INTISARI</b> .....	i
<b>ABSTRACT</b> .....	ii
<b>KATA PENGANTAR</b> .....	iii
<b>DAFTAR ISI</b> .....	v
<b>DAFTAR TABEL</b> .....	vii
<b>DAFTAR GAMBAR</b> .....	viii
<b>DAFTAR LAMPIRAN</b> .....	ix
<b>DAFTAR SIMBOL</b> .....	x
<b>BAB I PENDAHULUAN</b> .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	3
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah.....	3
1.5 Tinjauan Pustaka .....	4
1.6 Metodologi Penelitian .....	5
<b>BAB II LANDASAN TEORI</b> .....	7
2.1 <i>Machine Learning</i> .....	7
2.2 <i>Data Mining</i> .....	7
2.3 <i>Text Mining</i> .....	8
2.4 <i>Natural Language Processing</i> .....	8
2.5 <i>Keyword Extraction</i> .....	9
2.6 <i>Opinion Mining atau Sentiment Analysis</i> .....	9
2.7 Probabilitas .....	10
<b>BAB III WEB SCRAPING DAN NAÏVE BAYES CLASSIFIER</b> .....	13
3.1 <i>Web Scraping</i> .....	13
3.2 <i>Text Preprocessing</i> .....	13

3.3 Pelabelan Kelas Sentimen .....	14
3.4 <i>Naïve Bayes Classifier</i> .....	15
3.5 Evaluasi Model Klasifikasi.....	17
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	20
4.1 Pengambilan Data menggunakan Teknik <i>Web Scraping</i> .....	20
4.2 <i>Preprocessing</i> Data .....	20
4.2.1 <i>Cleansing Data</i> .....	23
4.2.2 <i>Case Folding</i> .....	23
4.2.3 <i>Spelling Normalization</i> .....	24
4.2.4 <i>Filtering Data</i> .....	25
4.2.5 <i>Tokenizing</i> .....	25
4.3 Pelabelan Kelas Sentimen .....	25
4.4 Pembuatan Data Latih dan Data Uji.....	27
4.5 Klasifikasi Menggunakan Metode <i>Naïve Bayes Classifier</i> .....	27
<b>BAB V PENUTUP</b> .....	32
5.1 Kesimpulan.....	32
5.2 Saran.....	32
<b>DAFTAR PUSTAKA</b> .....	34
<b>LAMPIRAN</b> .....	37

## DAFTAR TABEL

<b>Tabel 3.1</b> <i>Confusion Matrix</i> .....	17
<b>Tabel 3.2</b> Tabel Golongan Klasifikasi.....	17
<b>Tabel 4.1</b> Ilustrasi Data <i>Tweet</i> Hasil <i>Web Scraping</i> dari <i>Twitter</i> .....	21
<b>Tabel 4.2</b> Ilustrasi Perbedaan <i>Tweet</i> Ulasan dan Bukan Ulasan .....	22
<b>Tabel 4.3</b> Ilustrasi Proses <i>Cleansing Data</i> .....	23
<b>Tabel 4.4</b> Ilustrasi Proses <i>Case Folding</i> .....	24
<b>Tabel 4.5</b> Ilustrasi Proses <i>Spelling Normalization</i> .....	24
<b>Tabel 4.6</b> Ilustrasi Proses <i>Filtering Data</i> .....	25
<b>Tabel 4.7</b> Ilustrasi Proses <i>Tokenizing</i> .....	25
<b>Tabel 4.8</b> Ilustrasi Perhitungan Skor Sentimen .....	26
<b>Tabel 4.9</b> Jumlah Ulasan Pada Kelas Sentimen .....	26
<b>Tabel 4.10</b> Pembagian Data Latih dan Data Uji.....	27
<b>Tabel 4.11</b> Frekuensi Kata Berdasarkan Kelas Sentimen .....	28
<b>Tabel 4.12</b> Probabilitas Kata Berdasarkan Kelas Sentimen .....	28
<b>Tabel 4.13</b> Ilustrasi Pengujian Probabilitas Data Latih.....	29
<b>Tabel 4.14</b> Hasil <i>Confusion Matrix</i> .....	29

## DAFTAR GAMBAR

<b>Gambar 1.1</b> <i>Flowchart</i> Penelitian.....	6
<b>Gambar 4.1</b> <i>Script Jupyter</i> Untuk Membuat Wadah <i>Data Crawling</i> .....	20
<b>Gambar 4.2</b> <i>Script Jupyter</i> Untuk Pengambilan Data .....	20
<b>Gambar 4.3</b> <i>Script Jupyter</i> Untuk Membuat <i>Dataframe</i> .....	21
<b>Gambar 4.4</b> <i>Script Jupyter</i> Untuk Menyimpan Hasil Pengambilan Data.....	21

## DAFTAR LAMPIRAN

<b>Lampiran 1</b> <i>Syntax</i> .....	37
<b>Lampiran 2</b> <i>Screenshot</i> Ulasan Film dari <i>Twitter</i> .....	38
<b>Lampiran 3</b> Hasil Pelabelan Kelas Sentimen .....	39
<b>Lampiran 4</b> Frekuensi Kata Data Latih Berdasarkan Kelas Sentimen.....	46
<b>Lampiran 5</b> Probabilitas Kata Data Latih Berdasarkan Kelas Sentimen.....	49

## DAFTAR SIMBOL

$P(A)$	: Probabilitas terjadinya kejadian $A$
$n(A)$	: Banyaknya anggota $A$
$n(S)$	: Banyaknya anggota $S$
$P(A \cup B)$	: Probabilitas kejadian $A$ dan $B$ yang terjadi saling lepas
$P(B)$	: Probabilitas terjadinya kejadian $B$
$P(A \cap B)$	: Probabilitas kejadian $A$ dan $B$ terjadi secara bersamaan
$P(A B)$	: Probabilitas kejadian $A$ dengan syarat kejadian $B$ telah terjadi
$P(A_i B)$	: Probabilitas kejadian $A_i$ dengan syarat kejadian $B$ telah terjadi
$P(V_j w_i)$	: Probabilitas kelas $j$ pada saat terdapat kemunculan kata $i$
$P(w_i V_j)$	: Probabilitas kata $i$ masuk ke dalam kelas $j$
$P(V_j)$	: Probabilitas kemunculan sebuah kelas $j$
$P(w_i)$	: Probabilitas kemunculan sebuah kata
$i$	: Indeks kata $i$ dimulai dari 1 hingga $k$
$j$	: Indeks kata $j$ dimulai dari 1 hingga $m$
$V_{MAP}$	: Probabilitas tertinggi dari semua kelas yang diujikan
$n_j$	: Jumlah ulasan pada kelas $j$
$N$	: Jumlah ulasan dari semua kelas
$w_{ij}$	: Jumlah kata $i$ pada kelas $V_j$
$w_j$	: Jumlah kata pada kelas $V_j$
$W$	: Jumlah seluruh kata pada semua kelas
$TP$	: <i>True Positive</i>
$TN$	: <i>True Negative</i>
$FP$	: <i>False Positive</i>
$FN$	: <i>False Negative</i>

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pada era *Society 5.0* sekarang, manusia disuguhkan dengan berbagai macam teknologi yang dapat mempermudah aktivitas manusia dalam kehidupan sehari-hari. Teknologi yang ada pada saat ini mengalami perkembangan yang semakin hari semakin cepat. Teknologi yang berkembang pesat saat ini adalah teknologi digital. Berkembangnya teknologi digital memudahkan manusia untuk mendapatkan berbagai macam hiburan, di antaranya adalah film.

Film adalah media hiburan yang banyak digemari oleh berbagai kalangan manusia karena menghadirkan berbagai genre pilihan ketika ingin mencari hiburan melalui film. Drama, komedi, tragedi, aksi, dan horor merupakan lima genre film yang saat ini diminati karena mengangkat latar belakang kehidupan masyarakat dalam dunia supranatural, takhayul, dan cerita-cerita mistis (Fitri, 2022). Berdasarkan informasi yang didapat dari [filmindonesia.or.id](http://filmindonesia.or.id), genre horor merupakan genre yang paling banyak diminati oleh masyarakat Indonesia pada tahun 2022 dengan penonton paling banyak adalah Film KKN di Desa Penari dengan jumlah penonton sebanyak 10.058.788. Film tersebut merupakan film Indonesia yang paling banyak ditonton sepanjang sejarah.

KKN di Desa Penari merupakan sebuah film adaptasi dari cerita horor di *twitter* yang sempat populer di tahun 2019. Menurut informasi yang didapat dari [kompas.com](http://kompas.com), terdapat beberapa alasan yang menyebabkan Film KKN di Desa Penari sangat diminati oleh masyarakat di antaranya yaitu ceritanya yang sangat populer di *twitter* dan cerita tersebut diklaim berdasarkan kisah nyata. Ceritanya yang sangat populer mengakibatkan banyaknya masyarakat yang penasaran dengan bentuk visual dari cerita tersebut dan cerita yang diklaim berdasarkan kisah nyata semakin membuat masyarakat penasaran karena masyarakat Indonesia sangat dekat dengan hal-hal mistis.

Banyaknya penonton Film KKN di Desa Penari menyebabkan banyaknya ulasan masyarakat terhadap film tersebut. Ulasan Film KKN di Desa Penari dapat dilihat pada berbagai *website* namun pada penelitian ini ulasan yang digunakan adalah ulasan yang berasal dari *twitter* karena *twitter* merupakan media sosial berbasis teks dan memiliki kemudahan untuk melihat topik yang sedang populer sehingga menjadi sumber data yang melimpah untuk melakukan analisis teks. Banyaknya ulasan tentang film tersebut yang terdapat di *twitter* menyebabkan kesulitan dalam mengambil data jika diambil secara manual. Metode yang digunakan di dalam penelitian ini menggunakan dua tahap, yaitu pengumpulan data menggunakan *Web Scraping* kemudian dilanjutkan analisis sentimen menggunakan *Naïve Bayes Classifier*. Data yang dikumpulkan menggunakan teknik *Web Scraping* adalah data ulasan film KKN di Desa Penari dari *Twitter* pada hari pertama penayangan film tersebut di bioskop Indonesia yaitu pada tanggal 30 April 2022, kemudian penerapan *Naïve Bayes Classifier* untuk melakukan analisis sentimen dan melakukan klafisikasi terhadap ulasan film tersebut.

*Web Scraping* adalah proses pengambilan sebuah data semi terstruktur dari internet, umumnya dapat berupa halaman-halaman *website* dalam bahasa *markup* seperti HTML atau XHTML, dan menganalisis data tersebut untuk diambil dari halaman tersebut untuk dimanfaatkan bagi kepentingan lain (Turland, 2010). Data yang telah diambil menggunakan teknik *Web Scraping* belum dapat digunakan secara langsung untuk dianalisis karena data yang diperoleh dari hasil *Web Scraping* memiliki bentuk yang tidak terstruktur sehingga perlu dilakukan *Text Preprocessing* agar data tersebut lebih terstruktur.

Setelah data tersebut sudah terstruktur kemudian ulasan-ulasan tersebut diklasifikasikan untuk mengetahui kelas sentimen apa yang lebih banyak diberikan oleh para penonton film KKN di Desa Penari. Proses klasifikasi tersebut dilakukan dengan menggunakan *Naïve Bayes Classifier*. *Naïve Bayes Classifier* merupakan algoritma yang dipakai untuk mencapai nilai probabilitas tertinggi dalam mengklasifikasikan data uji pada kategori yang paling tepat (Feldman and Sanger, 2007).

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Bagaimana implementasi *Web Scraping* untuk mengambil data ulasan film KKN di Desa Penari di media sosial *twitter*?
2. Bagaimana implementasi *Naïve Bayes Classifier* untuk mengelompokkan ulasan film KKN di Desa Penari ke dalam tipe ulasan positif dan tipe ulasan negatif?

## 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disampaikan, tujuan dalam penelitian ini adalah sebagai berikut.

1. Mengimplementasikan *Web Scraping* untuk mengambil data ulasan film KKN di Desa Penari di media sosial *twitter*.
2. Mengimplementasikan *Naïve Bayes Classifier* untuk mengelompokkan ulasan film KKN di Desa Penari ke dalam tipe ulasan positif dan tipe ulasan negatif.

## 1.4 Batasan Masalah

Batasan masalah dalam penelitian ini yaitu sebagai berikut.

1. Data yang digunakan adalah ulasan dari film KKN di Desa Penari yang diambil dari media sosial *twitter* dengan menggunakan kata kunci “KKN di Desa Penari”.
2. Pengambilan ulasan dari media sosial *twitter* menggunakan bantuan *software Jupyter Notebook* dan *Microsoft Excel*.
3. Data ulasan yang diambil adalah ulasan pada hari pertama penayangan film tersebut di bioskop Indonesia yaitu dari tanggal 30 April 2022.
4. Ulasan yang diambil adalah ulasan yang menggunakan Bahasa Indonesia dan ulasan yang digunakan adalah ulasan film yang berisi pendapat, opini, dan analisis mengenai unsur-unsur film seperti plot, karakter, akting, sinematografi, dan unsur-unsur film lainnya, sedangkan hal-hal yang tidak berhubungan

dengan film dianggap bukan ulasan. Pembagian ulasan dan bukan ulasan dilakukan berdasarkan subjektif peneliti.

5. Ulasan dibagi menjadi ulasan positif dan ulasan negatif yang pembagiannya dilakukan berdasarkan subjektif peneliti. Ulasan positif adalah sentimen yang didominasi oleh kata-kata yang berisi penghargaan, kepuasan, atau pujian terhadap unsur-unsur film, sedangkan ulasan negatif adalah sentimen yang didominasi oleh kata-kata yang berisi kritik, ketidakpuasan, dan kesan buruk terhadap unsur-unsur film.

### 1.5 Tinjauan Pustaka

Penelitian menggunakan *Naïve Bayes Classifier* yang dilakukan oleh Fanny dan Suroyo (2022) untuk menentukan klasifikasi sentimen tanggapan pengguna *Twitter* terhadap Omnibus Law berdasarkan *hashtag*. Pada analisis tersebut diperoleh sentimen netral sebesar 55%, sentimen positif sebesar 35% dan sentimen negatif sebesar 10%. Kemudian dengan menggunakan *Naïve Bayes Classifier* terhadap *hashtag* pro diperoleh nilai rata-rata akurasi sebesar 92,1%, nilai rata-rata presisi sebesar 94,8% dan nilai rata-rata *recall* sebesar 90,7%, sedangkan untuk *hashtag* kontra diperoleh nilai rata-rata akurasi sebesar 98,3%, nilai rata-rata presisi sebesar 97,6% dan nilai rata-rata *recall* sebesar 98,7%. Hasil dari analisis dari gabungan *hashtag* pro dan *hashtag* kontra tersebut didapat kata yang paling dominan adalah Omnibus Law yang artinya berdasarkan seluruh *hashtag* yang diperoleh dari hasil *scraping* memang benar topik utamanya adalah tentang Omnibus Law.

Penelitian menggunakan *Naïve Bayes Classifier* yang dilakukan oleh Putri, Sugiyarto, dan Salafudin (2021) untuk menentukan klasifikasi sentimen tentang Covid-19 berdasarkan data yang diperoleh dari hasil *scraping* di *Twitter*. Pada analisis tersebut diperoleh sentimen netral sebesar 28,3%, sentimen positif sebesar 36,7% dan sentimen negatif sebesar 35%. Kemudian dengan menggunakan *Naïve Bayes Classifier* dengan perbandingan *training* dan *testing data* sebesar 70%:30% diperoleh nilai akurasi sebesar 83,1%.

Penelitian menggunakan *Naïve Bayes* dengan *Ensemble Feature* dan *Pearson Correlation Coefficient* yang dilakukan oleh Rangkuti, Fauzi, Arum Sari, dan

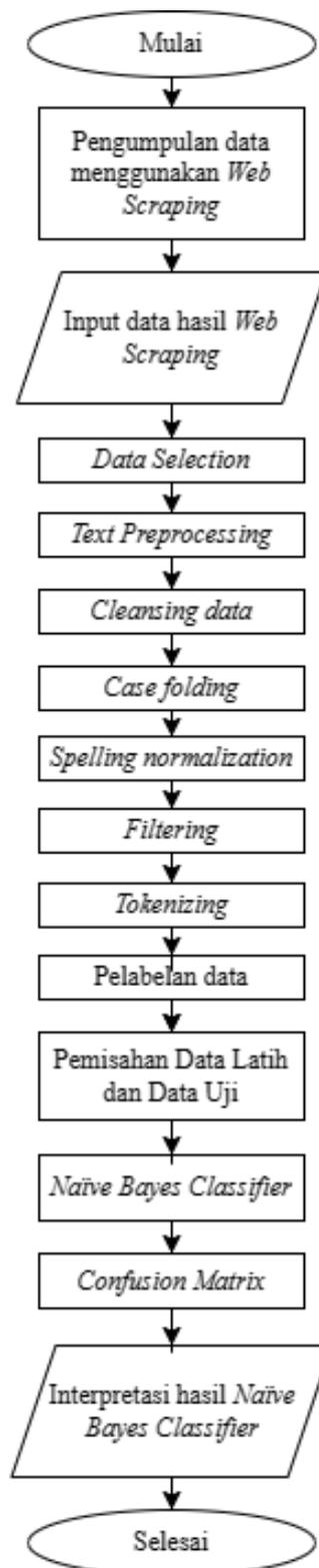
Lukmana Sari (2018) untuk mengklasifikasi sentimen opini film. Pada analisis tersebut dengan menggunakan kata yang tidak baku pada *tweet* didapat akurasi sebesar 82%, presisi sebesar 86%, *recall* sebesar 79,62% dan *f-measure* sebesar 82,69% pada seleksi fitur 20%. Setelah dilakukan pembakuan kata didapat akurasi sebesar 90%, presisi sebesar 92%, *recall* sebesar 88,46% dan *f-measure* sebesar 90,19% pada seleksi fitur 85%.

## 1.6 Metodologi Penelitian

Dalam penelitian ini metode yang digunakan diawali dengan studi literatur yaitu dengan mencari referensi yang sesuai dengan permasalahan yang diteliti. Data dalam penelitian ini adalah data sekunder yang didapat dari media sosial *twitter* menggunakan teknik *Web Scraping* menggunakan kata kunci “KKN di Desa Penari”. Data tersebut adalah ulasan Film KKN di Desa Penari di media sosial *twitter* pada hari pertama film tersebut tayang di bioskop yaitu pada tanggal 30 April 2022.

Adapun langkah-langkah dalam melakukan *Naïve Bayes Classifier* adalah sebagai berikut.

1. *Data selection*, yaitu proses melakukan seleksi terhadap data yang termasuk ke dalam ulasan dan bukan ulasan. Setelah melakukan seleksi, data yang bukan ulasan kemudian disingkirkan dan data yang masuk ke dalam ulasan kemudian digunakan untuk proses analisis berikutnya.
2. *Text Preprocessing*, yaitu proses memperbaiki data dari data yang tidak terstruktur ke dalam bentuk yang lebih terstruktur. Tahapan *preprocessing* terdiri dari *cleansing data*, *case folding*, *spelling normalization*, *filtering*, dan *tokenizing*.
3. Pelabelan dan pembobotan data, yaitu proses membagi data ulasan ke dalam bentuk ulasan positif dan ulasan negatif.
4. Langkah berikutnya adalah melakukan proses pengujian hasil klasifikasi menggunakan metode *Naïve Bayes Classifier* yang dilakukan untuk mendapatkan nilai akurasi. Alur dalam penelitian dapat dilihat pada *flowchart* berikut.



**Gambar 1.1** *Flowchart Penelitian*