

***PARTICLE SWARM OPTIMIZATION (PSO) SEBAGAI OPTIMASI
ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN TWITTER
PADA DRAMA QUEEN OF TEARS***

SKRIPSI



Oleh:

IKFINA AZIZAH

2020502049

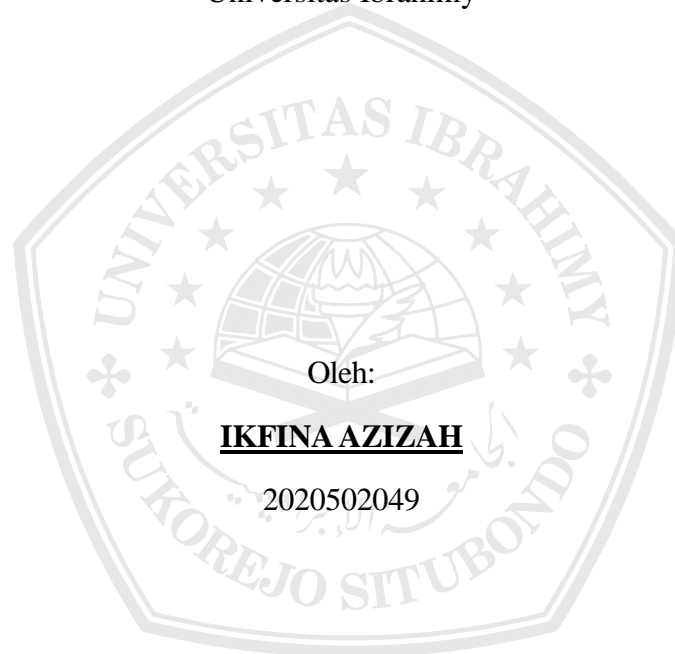
**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS IBRAHIMY
SITUBONDO**

2024

***PARTICLE SWARM OPTIMIZATION (PSO) SEBAGAI OPTIMASI
ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN TWITTER
PADA DRAMA QUEEN OF TEARS***

SKRIPSI

Diajukan untuk Memenuhi Salah Satu Persyaratan dalam Menyelesaikan Program Sarjana (S-1) pada Program Studi Sistem Informasi Fakultas Sains dan Teknologi Universitas Ibrahimy



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS IBRAHIMY
SITUBONDO
2024**

PERNYATAAN KEASLIAN TULISAN

Saya yang bertandatangan dibawah ini:

Nama : **Ikfina Azizah**
NPM : 2020502049
Program Studi : S-1 Sistem Informasi
Fakultas : Fakultas Sains dan Teknologi

Menyatakan dengan sebenarnya, bahwa tugas akhir/skripsi ini secara keseluruhan adalah hasil penelitian atau karya saya sendiri, kecuali pada bagian-bagian yang dirujuk sebagai sumber referensi dan disebutkan dalam daftar pustaka. Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa tugas akhir/skripsi ini hasil plagiasi, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Situbondo, 30 Agustus 2024

Saya yang menyatakan,

Ikfina Azizah

PERSETUJUAN PEMBIMBING

Skripsi ditulis oleh :

Nama : Ikfina Azizah

NPM : 2020502049

Judul : Particle Swarm Optimization (PSO) sebagai Optimasi
Algoritma Naive Bayes untuk Analisis Sentimen Twitter
Pada Drama Queen Of Tears

Telah ditelaah dan disetujui oleh pembimbing untuk diuji pada sidang/munaqosah.

Telah disetujui oleh :

Pembimbing 1

Pembimbing 2

Nur Azise, M.Kom

NIDN:0730108802

Muhammad Ali Ridla, M.Kom

NIDN:0728089102

PENGESAHAN**SKRIPSI****PARTICLE SWARM OPTIMIZATION (PSO) SEBAGAI OPTIMASI
ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN TWITTER
PADA DRAMA QUEEN OF TEARS****IKFINA AZIZAH**
2020502049

Telah dipertanggung jawabkan di depan dewan penguji Sidang/Munaqosah Skripsi pada hari Sabtu, Tanggal 31 Agustus 2024 sebagai salah satu syarat memperoleh gelar Sarjana (S.Kom) pada Fakultas Sains dan Teknologi Universitas Ibrahimi.

Tim Penguji,

Ketua Sidang,

Sekertaris Sidang,

Dr. Ach. Khumaidi, M.P
NIDN: 0722049001**Uslan Hidayat, S.Kom**
NIDN

Penguji I,

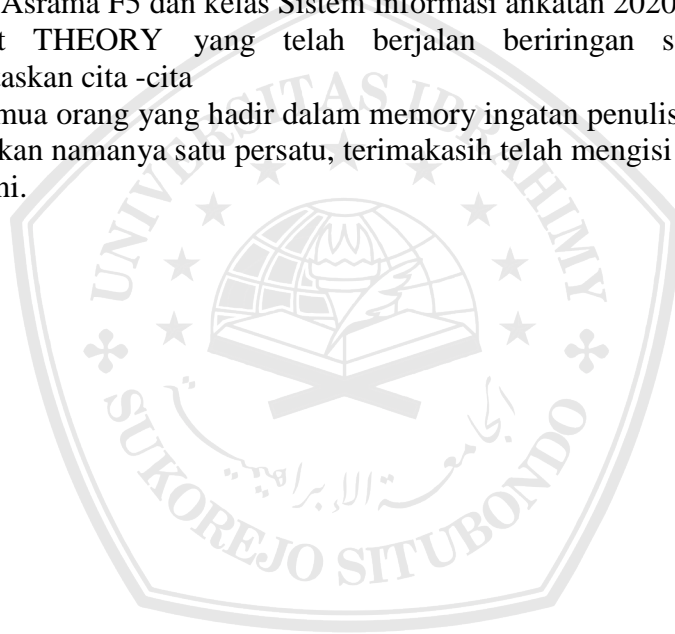
Penguji II,

Firman Santoso, M.Kom
NIDN: 0722129201**Ahmad Homaidi, M.Kom**
NIDN: 0705078901Mengetahui
Dekan,**Abd. Ghofur, M.Kom**
NIDN: 0711088303

PERSEMBAHAN

Skripsi ini dipersembahkan teruntuk orang-orang yang telah berkontribusi besar dalam kehidupan peneliti serta memberi kontribusi dalam proses pembuatan skripsi yang berjudul Particle Swarm Optimization ini :

1. Kedua orang tua yang telah berjuang baik dari segi spiritual dan juga finansial
2. Ke dua Kakak yang selalu mensupport langkah demi langkah perjuangan dalam meniti karir
3. Teman – teman seperjuangan yang selalu menyuntikkan semangat berjuang baik di Asrama F5 dan kelas Sistem Informasi angkatan 2020
4. Sahabat THEORY yang telah berjalan beriringan sama berjuang menuntaskan cita -cita
5. Dan semua orang yang hadir dalam memory ingatan penulis, yang tak bisa disebutkan namanya satu persatu, terimakasih telah mengisi hal baik dalam hidup ini.



KATA PENGANTAR

Segala puji bagi Allah SWT yang telah melimpahkan rahmat dan hidayahnya bagi sekalian alam. Sholawat serta salam tak henti-hentinya tetap tercurah limpahkan kepada Baginda Nabi Muhammada SAW yang telah menebarkan cahaya keilmuan kepada ummatnya. Berkat rahmat dan hidayah-NYA lah skripsi ini dapat dirampungkan oleh penulis dengan sebaik-baiknya oleh karena itu penulis juga menyajikan banyak terimakasih kepada:

1. Guru kami KHR. Azaim Ibrahimy, S.Sy. M.HI. pengasuh Pondok Pesantren Salafiyah Syafi'iyah Sukorejo.
2. Kedua orang tua dan kedua kakak perempuan Umamah Aini dan Kinanah Aini yang selalu memberikan dukungan baik dari segi spiritual maupun material.
3. KH. Ach. Fadlail, SH, M.H Selaku Rektor Universitas Ibrahimy
4. Abd. Ghofur, M.Kom Selaku Dekan Fakultas Sains dan Teknologi.
5. Dr. Ach. Khumaidi, M.P Selaku Wakil Dekan 1 Fakultas Sains Dan Teknologi.
6. Abd. Wafi, M.P Selaku Wakil Dekan 2 Fakultas Sains dan Teknologi.
7. Akhmad Luthfi, M.Kom Selaku Wakil Dekan 3 Fakultas Sains dan Teknologi.
8. Ahmad Baujiri, M.Kom Selaku Ketua Prodi Sistem Informasi.
9. Ibu Nur Azizse, M.Kom dan bapak Muhammad Ali Ridla, M.Kom selaku dosen pembimbing pada penelitian ini, yang telah memberikan arahan dan bimbingan.
10. Seluruh Dosen Fakultas Sains dan Teknologi yang telah memberikan bekal ilmu hingga rampunglah skripsi ini.

Situbondo, 30 Agustus 2024

Peneliti

ABSTRAK

Ikfina Azizah, 2024. ***PARTICLE SWARM OPTIMIZATION (PSO) SEBAGAI OPTIMASI ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN TWITTER PADA DRAMA QUEEN OF TEARS.*** Skripsi Program Studi Sistem Informasi, Universitas Ibrahimy. Pembimbing: (I) Nur Azise, M.Kom., (II) Muhammad Ali Ridla, M.Kom.

Analisis sentimen atau yang bisa disebut juga sebagai penggalian opini adalah proses menganalisis teks dalam jumlah besar untuk menentukan apakah teks tersebut mengekspresikan sentimen positif, sentimen negatif, atau sentimen netral. Banyaknya penelitian mengenai analisis sentimen membuktikan bahwa analisis sentimen memberikan banyak manfaat yang cukup besar, seperti mengetahui respon masyarakat terhadap suatu objek tertentu. Analisis sentimen menggunakan algoritma *Naive Bayes* sudah banyak dilakukan akan tetapi algoritma *Naive Bayes* juga memiliki kekurangan di mana sebuah probabilitas tidak bisa mengukur seberapa besar tingkat akurasi sebuah prediksi, selain itu *Naive Bayes* juga memiliki kelemahan pada seleksi atribut sehingga mempengaruhi nilai akurasi, oleh karena itu algoritma *Naive Bayes* perlu dioptimasi dengan cara mengoptimasi menggunakan algoritma *Particle Swarm Optimization (PSO)* untuk meningkatkan akurasi pada algoritma *Naive Bayes*, algoritma PSO memiliki konsep yang sederhana dan mudah diimplementasikan serta efisien dalam perhitungan jika dibandingkan dengan algoritma matematika dan teknik optimasi lainnya. Hasil yang didapatkan dalam penelitian ini menunjukkan bahwa optimasi algoritma NBC menggunakan PSO menghasilkan nilai Accuracy sebesar 68%.

kunci: Analisis Sentimen, PSO, NBC

ABSTRACT

Ikfina Azizah, 2024. **Particle Swarm Optimization As a Naive Bayes Algorithm Optimization For Twitter Sentiment Analysis in the Drama Queen Of Tears.** Programming Languages Thesis, Information Systems Studi Program, Ibrahimi University.

Supervisor: (I) Nur Azise, M.Kom., (II) Muhammad Ali Ridla, M.Kom.

Sentiment analysis or what can also be called opinion mining is the process of analyzing large amounts of text to determine whether the text expresses positive sentiment, negative sentiment, or neutral sentiment. Many studies on sentiment analysis prove that sentiment analysis provides many significant benefits, such as knowing the public's response to a particular object. Sentiment analysis using the Naive Bayes algorithm has been widely carried out, but the Naive Bayes algorithm also has shortcomings where a probability cannot measure the level of accuracy of a prediction, besides that Naive Bayes also has weaknesses in attribute selection which affects the accuracy value, therefore the Naive Bayes algorithm needs to be optimized by optimizing using the Particle Swarm Optimization (PSO) algorithm to improve accuracy in the Naive Bayes algorithm, the PSO algorithm has a simple concept and is easy to implement and is efficient in calculation when compared to mathematical algorithms and other optimization techniques. The results obtained in this study indicate that the optimization of the NBC algorithm using PSO produces an Accuracy value of 68%.

Keywords: Analisis Sentimen, NBC, PSO

DAFTAR ISI

Cover	
PERNYATAAN KEASLIAN TULISAN	i
PERSETUJUAN PEMBIMBING	ii
PENGESAHAN	iii
SKRIPSI	iii
PERSEMBAHAN	iv
KATA PENGANTAR	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR SEGMENT PROGRAM	xiii
BAB I	1
PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah	4
1.3. Rumusan Masalah	4
1.4. Batasan Masalah	5
1.5. Tujuan Penelitian	5
1.6. Manfaat Penelitian	6
1.7. Metode Penelitian	6
1.8. Jenis Penelitian	6
1.9. Metode Pengumpulan Data	7
1.10. Sistematika Pembahasan	8
BAB II	10
TINJAUAN PUSTAKA	10
2.1 Penelitian Terdahulu	10
2.2 Landasan Teori	13

2.2.1	Analisis Sentimen	13
2.2.2	Algoritma Optimasi.....	13
2.2.3	Twitter	14
2.2.4	Drama Korea <i>Queen Of Teras</i>	14
2.2.5	Algoritma <i>Naive Bayes</i>	15
2.2.6	<i>Particle Swarm Optimization (PSO)</i>	17
2.2.7	<i>Confusion Matrix</i>	20
2.2.8	<i>Python</i>	22
2.3	Pemodelan.....	23
2.3.1	<i>Use Case Diagram</i>	23
2.3.2	<i>Flowchart</i>	23
Tabel 2.1	Simbol <i>Flowchart</i>	23
2.4	Perangkat Lunak Yang Digunakan	25
2.4.1	Google Colab.....	25
2.4.2	Microsoft Excel.....	25
BAB III	26
ANALISIS DAN PERANCANGAN	26
3.1	Gambaran Umum Objek Penelitian.....	26
3.1.1	<i>Crawling Data Twitter</i>	26
3.2	Alur Proses.....	43
3.2.1	Identifikasi dan Analisis proses Bisnis.....	43
3.2.2	Analisis Proses Bisnis	43
3.2.3	Identifikasi dan Analisis Kebutuhan	47
3.2.4	Identifikasi dan Analisis Alternatif Solusi.....	49
Tabel 3.1	Identifikasi Alternatif Solusi	49
Tabel 3.2	Analisis Kelayakan.....	51
3.2.5	Desain Proses	52
IMPLEMENTASI SISTEM	57
4.1	Konstruksi Sistem.....	57
4.1.1	Kebutuhan Sistem	57
4.2	Pengumpulan Data	57

4.3	<i>Text Preprocessing</i>	44
Tabel 4.1	Proses Cleaning Tweet	45
Tabel 4.2	Proses Case Folding	48
Tabel 4.3	Proses Normalisasi	49
Tabel 4.4	Proses Tokenizing	52
Tabel 4.5	Proses Stopword Removal	53
Tabel 4.6	Proses Steaming	55
4.4	Wordcloud	56
4.5	Pelabelan Otomatis dengan lexicon Based	57
4.6	Split Data Training dan Data Testing	59
4.7	Metode Naive Bayes Clasifier	60
Tabel 4.7	Confusion Matrix NB	62
Tabel 4.8	Hasil Evaluasi NB	63
4.8	Particle Swarm Optimization	64
Tabel 4.18	Hasil confusion matrix	68
Tabel 4.18	Hasil dari metode PSO	69
BAB V	69
PENUTUP	69
5.1	Kesimpulan	69
5.2	Saran	69
DAFTAR PUSTAKA	71
LAMPIRAN	73

DAFTAR GAMBAR

Gambar 3.1 Data Hasil Crawling Twitter.....	27
Gambar 3.2 Tahapan <i>Preprocessing</i>	29
Gambar 3.3 Tahapan Analisis Sentimen.....	32
Gambar 3.4 Preprocessing data.....	38
Gambar 3.5 Alur Labeling.....	39
Gambar 3.6 Alur metode NBC.....	39
Gambar 3.7 Alur metode PSO.....	40
Gambar 3.8 Arsitektur Sistem.....	41
Gambar 4.1 Hasil Preprocessing Data.....	56
Gambar 4.2 Wordcloud sebelum preprocessing.....	56
Gambar 4.3 Wordcloud setelah preprocessing.....	57
Gambar 4.4 Hasil Labeling Data.....	58
Gambar 4.5 Diagram hasil labeling.....	59
Gambar 4.6 Split data.....	59
Gambar 4.7 Input hasil Labeling data.....	61

DAFTAR TABEL

Tabel 2.1 Simbol <i>Flowchart</i>	24
Tabel 3.1 Identifikasi Alternatif Solusi.....	34
Tabel 3.1 lanjutan	35
Tabel 3.2 Analisis Kelayakan	36
Tabel 3.2 lanjutan	37
Tabel 4.1 Proses Cleaning Tweet	45
Tabel 4.2 Proses Case Folding.....	48
Tabel 4.3 Proses Normalisasi.....	49
Tabel 4.3 lanjutan	50
Tabel 4.4 proses Tokenizing	52
Tabel 4.5 Proses Stopword Removal.....	53
Tabel 4.5 lanjutan	54
Tabel 4.6 Proses Steaming.....	55
Tabel 4.7 Confusion Matrix NB.....	62
Tabel 4.8 Hasil Evaluasi NB	63
Tabel 4.9 Hasil Confusion Matrix	68
Tabel 4.10 Hasil dari Metode PSO.....	69

DAFTAR SEGMENT PROGRAM

Segment Program 4.1 <i>Twitter Auth Token</i>	43
Segment Program 4.2 <i>Crawling Data Twitter</i>	44
Segment Program 4.3 Proses Cleaning.....	46
Lanjutan Segment Program 4.3	47
Segment Program 4.4 Proses Case Folding.....	49
Segment Program 4.5 Proses Normalisasi.....	49
Lanjutan segment 4.5.....	49
Segment Program 4.6 Proses Tokenizing.....	53
Segment Program 4.7 Proses Stopword Removal.....	54
Segment Program 4.8 Proses Steaming.....	55
Segment Program 4.9 Naive Bayes	63
Segment Program 4.10 PSO.....	64
Segment Program 4.11 Inisialisasi.....	65
Segment lanjutan 4.11	65

BAB I

PENDAHULUAN

1.1. Latar Belakang

Analisis sentimen atau yang bisa disebut juga sebagai penggalian opini adalah proses menganalisis teks dalam jumlah besar untuk menentukan apakah teks tersebut mengekspresikan sentimen positif, sentimen negatif, atau sentimen netral. Banyaknya penelitian mengenai analisis sentimen membuktikan bahwa analisis sentimen memberikan banyak manfaat yang cukup besar, seperti mengetahui respon masyarakat terhadap suatu objek tertentu [1]. Termasuk analisis sentimen mengenai percakapan di Twitter yang banyak mengandung respon masyarakat terkait suatu hal tertentu. Twitter atau yang saat ini dikenal dengan aplikasi X adalah layanan jejaring sosial atau bisa juga disebut mikroblog daring yang memungkinkan pengguna untuk mengirim, membaca, dan membalas pesan teks hingga 280 karakter (dikenal dengan sebutan kicauan (*Tweet*))[2].

Menurut laporan dari *We Are Social* pengguna Twitter diseluruh dunia mencapai angka 564,1 juta per Juli 2023, jumlah pengguna Twitter naik sebesar 16,1% dibandingkan tahun sebelumnya. Indonesai tercatat sebagai peringkat ke-4 dengan pengguna Twitter terbanyak didunia pada juli 2023 yang mengalami kenaikan sebesar 71,2% dengan jumlah pengguna mencapai 25,5 juta dari laporan sebelumnya

Indonesia berada diperingkat keenam terbanyak dunia pada Mei 2023, sementara di urutan pertama Amerika masih menjadi negara dengan pengguna Twitter terbanyak di Dunia dengan total pengguna sebanyak 98,5 juta [3].

Dalam beberapa tahun terakhir, popularitas K-Drama di Indonesia terus meningkat yang didorong oleh beberapa faktor seperti naiknya platform *streaming* dan daya tarik budaya Korea, serta meningkatnya percakapan di Twitter tentang konten hiburan dari Korea Selatan, Twitter sendiri menjadi rumah bagi banyaknya pelaku industri hiburan Korea Selatan termasuk drama Korea hal ini dibuktikan dengan adanya percakapan di Twitter mengenai hiburan Korea Selatan yang memecahkan rekor *tweet* secara global mencapai 7,8 miliar pada tahun 2021 dan pada tahun sebelumnya tercatat ada 6,7 miliar *tweet* seputar hiburan Korea Selatan. Dengan demikian secara global perbincangan seputar hiburan Korea di *Twitter* meningkat sebesar 16 % [4].

Berdasarkan data internal Twitter yang dihimpun mulai 1 Januari sampai 3 Desember 2021, Indonesia tercatat sebagai negara yang paling banyak membicarakan hiburan Korea di Twitter termasuk yang baru-baru ini tayang diplatform *streaming* Netflix yakni drama *Queen of Tears* yang dibintangi oleh Kim So Hyun dan Kim Jiwon sebagai pemeran utama dalam drama ini, drama bergenre romantis komedi ini berjumlah 16 episode [5]. Dari survei global Twitter mengenai drama ini ada sebanyak 1530 ribu percakapan di Twitter yang menggunakan tagar *#Queenoftears* dengan

berbagai macam opini penonton mengenai drama ini, tentu saja hal ini memberikan efektifitas kepada pelaku industri hiburan Korea drama *Queen of Tears* untuk mengetahui tanggapan para penggemar drama ini. Dari pemaparan diatas perlu dilakukan penelitian untuk mengetahui tanggapan penonton berdasarkan sentimen dimedia sosial *Twitter* terhadap alur cerita drama Korea *Queen of tears* dengan mengambil sample percakapan *Twitter* yang menggunakan tagar *#Queenoftears*.

Analisis sentimen menggunakan algoritma *Naive Bayes* sudah banyak dilakukan diantaranya adalah peneliaian opini *Twitter* terhadap tokoh publik dengan algoritma *Naive Bayes* dan *Support Vector Machine*, penelitian ini dilakukan oleh Tanthy Tawaqalia Widawat, penelitian ini menggunakan data dengan dua metode klasifikasi yang berbeda yakni *Naive Bayes* dan *SVM*, hasil akurasi dari *Naive Bayes* adalah sebesar 91,48% sedangkan *SVM* sebesar 85,47%, dari hasil akurasi tersebut menunjukkan bahwa metode klasifikasi *Naive Bayes* memiliki nilai akurasi yang lebih tinggi dibandingkan nilai akurasi dari *SVM* [6]. Akan tetapi algortima *Naive Bayes* juga memiliki kekurangan dimana sebuah probabilitas tidak bisa mngukur seberapa besar tingkat akurasi sebuah prediksi, selain itu *Naive Bayes* juga memiliki kelemahan pada seleksi atribut sehingga mempengaruhi nilai akurasi, oleh karena itu algoritma *Naive bayes* perlu dioptimasi dengan cara memberikan bobot pada atribut agar *Naive Bayes* bisa bekerja lebih efektif [7].

Berdasarkan uraian permasalahan tersebut, algoritma *Particle Swarm Optimization* (PSO) dapat digunakan dalam melakukan pembobotan atribut untuk meningkatkan akurasi pada algoritma *Naive Bayes*, algoritma PSO memiliki konsep yang sederhana dan mudah diimplementasikan serta efisien dalam perhitungan jika dibandingkan dengan algoritma matematika dan teknik optimasi lainnya [8]. Maka dari itu diusulkan penelitian menggunakan *Particle Swarm Optimization* (PSO) dalam menganalisis sentimen opini penonton drama *Queen Of Tears*. Optimasi ini bertujuan untuk mempengaruhi tingkat akurasi klasifikasi dari algoritma *Naive Bayes*.

1.2. Identifikasi Masalah

Berdasarkan latar belakang penelitian ini terdapat beberapa identifikasi permasalahan sebagai berikut :

- a. Banyaknya percakapan di sosial media Twitter dengan tagar *#Queenoftears* yang menunjukkan sentimen penonton terhadap kepuasan dan ketidakpuasan penonton terhadap drama ini.
- b. Dibutuhkan metode untuk menganalisis sentimen opini penonton drama ini dengan menggunakan *tweet* dengan tagar *#Queenoftears*.
- c. Dibutuhkan optimasi pada algoritma *Naive bayes* dalam menganalisis sentimen opini Twitter.

1.3. Rumusan Masalah

Rumusan masalah yang dapat diambil dari penelitian ini adalah :

- a. Bagaimana penerapan metode *Naive Bayes* dalam menganalisis opini penonton drama *Queen Of tears* melalui percakapan di Twitter.
- b. Bagaimana hasil pengujian akurasi metode *Particle Swarm optimization* (PSO) dalam mengoptimasi algoritma *Naive bayes*.

1.4. Batasan Masalah

Batasan untuk penelitian ini adalah sebagai berikut :

- a. Data yang diambil adalah *tweet* dengan tagar *#Queenoftears* yang bersumber dari media sosial Twitter yang diunggah sejak tanggal 30 April sampai 08 Mei 2024.
- b. Data yang digunakan adalah dokumen *tweet* yang bersumber dari media sosial Twitter yang berbahasa Indonesia.
- c. Algoritma yang digunakan adalah algoritma *Naive Bayes* dengan optimasi menggunakan metode *Particle Swarm Optimization* (PSO).

1.5. Tujuan Penelitian

Dari rumusan masalah di atas, dapat diambil kesimpulan tujuan dari penelitian ini sebagai berikut :

- a. Mengetahui presentase ketepatan klasifikasi sentimen penonton dengan metode *Naive Bayes* dan optimasi menggunakan *Particle Swarm optimization* (PSO) pada drama *Queen of Tears*.
- b. Mendapatkan visualisasi pemetaan positif atau negatif pada masing-masing *tweet* sentimen dari penonton drama *Queen of tears*.

1.6. Manfaat Penelitian

Adapun manfaat yang dihasilkan dari penelitian ini adalah sebagai berikut:

- a. Menerapkan dan mengembangkan ilmu yang telah diperoleh selama masa perkuliahan.
- b. Menambah pengetahuan dalam bidang data & machine learning terkait analisis sentimen sehingga mendapatkan hasil yang bermanfaat.
- c. Sebagai referensi dan tolak ukur terkait kepuasan dan tidak puasnya penonton suatu drama yang menjadi *trending topic* dikalangan penikmat drama Korea.

1.7. Metode Penelitian

Metode penelitian merupakan sesuatu yang dilakukan untuk mendapatkan pengetahuan baru atau memecahkan masalah tertentu melalui pengumpulan, analisis dan interpretasi data. Metode yang digunakan dalam penelitian ini adalah sebagai berikut :

1.8. Jenis Penelitian

a. Pendekatan Kuantitatif

Menurut Emzir (2009:28), pendekatan kuantitatif adalah satu pendekatan yang secara primer menggunakan paradigma *postpositivist* dalam mengembangkan ilmu pengetahuan seperti pemikiran sebab akibat, reduksi kepada variabel, hipotesis dan pernyataan spesifik menggunakan pengukuran dan observasi pengujian teori, dalam pendekatan kuantitatif menggunakan strategi penelitian seperti eksperimen dan survei yang memerlukan data statistik, sehingga sesuai dengan namanya yakni

banyak menggunakan angka, mulai dari data penafsiran setra hasilnya [8].

b. Eksperimen

Eksperimen yang digunakan dalam penelitian ini adalah menggunakan bahasa pemrograman *Python* untuk menganalisis penggunaan algoritma *Naive Bayes* serta mengoptimasi algoritma *Naive Bayes* dengan *Particle Swarm Optimization* (PSO).

c. Library Research

Penelitian kepustakaan ini adalah jenis penelitian yang pada umumnya dilakukan dengan mencari sumber datanya, riset ini dilakukan berdasarkan karya-karya tulis, termasuk hasil penelitian baik yang sudah maupun belum dipublikasi [10].

1.9. Metode Pengumpulan Data

a. Studi Literatur

Studi Pustaka merupakan teknik pengumpulan data untuk memperoleh data dari sebuah tulisan yang berkaitan dalam pembangunan sistem informasi seperti menghimpun informasi dari skripsi, jurnal maupun tugas akhir dari para peneliti sebelumnya [11].

b. Crawling Data Twitter

Crawling data merupakan teknik mengumpulkan data pada sebuah website dengan memasukkan *Uniform Resource*

Locator (URL), untuk penerapannya pada data Twitter *crawling* menggunakan *Application Programming Interface* (API) sebagai jalur komunikasi dalam mendapatkan data [12].

1.10. Sistematika Pembahasan

Penulisan ini terdiri dari beberapa bab keseluruhan, bab ini berisi uraian tentang bahasan pemecahan masalah secara berurutan. Berikut adalah sistematika pembahasan yang digunakan :

Bab I : PENDAHULUAN

Pada bab ini menjelaskan tentang latar belakang, identifikasi masalah, runusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian, metode pengumpulan data dan sistematika pembahasan.

Bab II : TINJAUAN PUSTAKA

Pada bab ini menjelaskan tentang tinjauan pustaka dari adanya penelitian ini, tinjauan pustaka berisi referensi seperti jurnal, skripsi atau lainnya. Pada bab ini terdapat beberapa sub bab yakni penelitian terdahulu, landasan teori, pemodelan dan perangkat lunak yang digunakan.

Bab III : ANALISIS DAN PERANCANGAN

Pada bab ini menjelaskan tentang gambaran alur proses dan desain dari objek penelitian yang diteliti, terdapat beberapa sub bab pada bagian ini yakni identifikasi dan analisis proses bisnis,

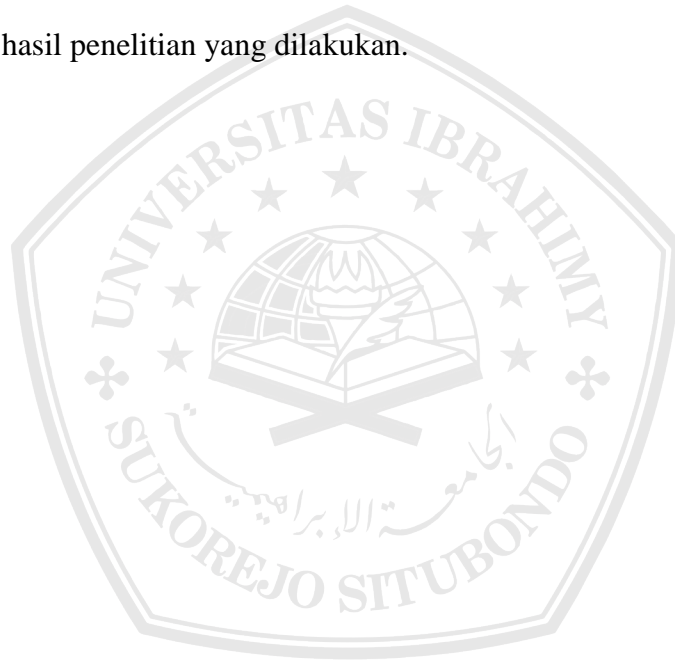
identifikasi dan analisis kebutuhan, identifikasi dan alternatif solusi, sedangkan desain sistem membahas tentang perancangan alur sistem.

Bab IV : HASIL DAN PEMBAHASAN

Pada bagian ini menjelaskan tentang alur dari perhitungan algoritma, pengujian algoritma.

BAB V : PENUTUP

Pada bab ini membahas tentang kesimpulan dan sari dari hasil penelitian yang dilakukan.



BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

a. Comparison of Naive Bayes and PSO-Based Naive Bayes Algorithm for prediction of Covid-19 Patient Recovery Data in Indonesia

Penyakit baru yang dikenal sebagai COVID 19 diidentifikasi pada tahun 2019 namun belum menginfeksi manusia (Organisasi Kesehatan Dunia, 2019). Kelompok virus ini dapat menginfeksi mamalia termasuk manusia dan juga burung dan menyebabkan penyakit. Orang biasanya tertular virus corona melalui flu dan penyakit pernapasan ringan lainnya, namun mereka juga dapat menyebarkan penyakit serius seperti SARS, MERS, dan COVID-19 yang mematikan. Agar tidak ada lagi korban jiwa, jumlah ini harus dikurangi. Penting untuk memahami variabel-variabel yang benar-benar dapat mengurangi risiko kematian dan mengukur kecenderungan kesembuhan pasien Covid-19. Beberapa teknik dalam penambangan data dapat digunakan untuk memperkirakan tingkat kesembuhan pasien tergantung pada berbagai karakteristik. Kriteria penelitian ini meliputi jenis kelamin, usia, provinsi, dan status. Algoritma *Naive Bayes* (NB) dan *Naive Bayes* berbasis PSO dibandingkan dalam penelitian ini menggunakan kumpulan data

pasien untuk menentukan apakah strategi lebih akurat. Temuan penelitian ini mengungkapkan bahwa penggunaan metode NB memiliki tingkat akurasi sebesar 94,07%, nilai presisi sebesar 14%, nilai recall sebesar 1%, dan nilai AUC sebesar 0,613, sesuai dengan data penelitian. Tingkat akurasi *Naive Bayes* berbasis PSO sebesar 95,56%, presisi sebesar 25%, recall sebesar 1%, dan AUC sebesar 0,540 [13].

b. Analisis Sentimen Pemilu 2024 dengan *Naive bayes* Berbasis *Particle Swarm Optimization* (PSO)

Penelitian yang dilakuakn oleh tommy Dwi Putra, Ema Utami dan Mei P Kurniawan yang membahas tentang pemilu 2024 dengan objek data yang diambil dari Twitter, menggunakan metode *Naive Bayes* (NB) berbasis *Particle Swarm Optimization* (PSO) dalam menganalisa label sentimen positif, negatif dan netral pada postingan para pengguna media sosial Twitter yang berkaitan dengan pemilu 2024. Jumlah dataset setelah pemrosesan menjadi 1.000 data. Tahap evaluasi menggunakan *Confusion Matrix* diperoleh nilai akurasi algoritma yaitu untuk *Naive Bayes* tanpa *Particle Swarm Optimization* (PSO) nilai akurasi sebesar 73,67%. Sedangkan untuk algoritma *Naive Bayes* berbasis *Particle Swarm Optimization* (PSO) nilai akurasi sebesar 78,33%. Dapat diketahui bahwa tingkat akurasi dari algoritma *Naive Bayes* berbasis *Particle*

Swarm Optimization (PSO) lebih unggul dibandingkan algoritma *Naive Bayes* tanpa *Particle Swarm Optimization* (PSO) [14].

c. Perbandingan Algoritma *Naive Bayes* dan *Naive Bayes* Berbasis PSO untuk Analisis Kredit Pada PT. BPR Syariah Paokmotong

Menurut UU No.21 tahun 2008, dinyatakan bahwa bank syariah adalah badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkan kepada masyarakat dalam bentuk kredit dan/atau bentuk lainnya dalam rangka meningkatkan taraf hidup rakyat. Kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan pinjam meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam melunasi hutangnya setelah jangka waktu tertentu dengan pemberian bunga (UU Perbankan No.10 Tahun 1998). Dalam melakukan analisa sebuah kredit terkadang seorang analis melakukan analisa kurang akurat, sehingga ada beberapa nasabah yang kurang mampu dalam melakukan pembayaran angsuran kredit, sehingga mengakibatkan kredit kurang lancar bahkan macet. Dari permasalahan di atas, peneliti melakukan analisis kredit menggunakan teknik komputerisasi dengan memanfaatkan perangkat lunak RapidMiner dalam melakukan pengolahan data. Teknik pengolahan data yang tepat untuk digunakan adalah klasifikasi. Salah satu metode klasifikasi data mining adalah algoritma *Naive Bayes*. Peneliti menggunakan pembobotan dengan

menerapkan *Particle Swarm Optimization* (PSO) untuk seleksi atribut guna meningkatkan akurasi *Naive Bayes*. Setelah melakukan pengujian dengan model algoritma *Naive Bayes* sebesar 71.00%, sedangkan dengan model algoritma *Naive Bayes* berbasis *Particle Swarm Optimization* menghasilkan nilai akurasi yang lebih tinggi yaitu 88.51% dibandingkan dengan model algoritma *Naive Bayes*. Dari hasil tersebut didapatkan selisih antara kedua model yaitu 17.51% [15].

2.2 Landasan Teori

2.2.1 Analisis Sentimen

Analisis sentimen atau penggalian opini adalah suatu proses menganalisis teks dalam jumlah besar untuk menentukan apakah teks tersebut mengekspresikan sentimen positif, sentimen negatif atau sentimen netral. Analisis sentimen penting untuk dilakukan agar memberikan hasil yang lebih objektif untuk beberapa subjek tertentu sebagai penggalian ulasan dari beberapa target, atau bisa juga untuk mencapai skalabilitas yang lebih besar dari suatu program yang dibuat oleh pihak tertentu sebagai pengembangan bisnis stake holder.[16]

2.2.2 Algoritma Optimasi

Algoritma Optimasi adalah sebuah proses menemukan nilai x dengan beberapa proses perhitungan sehingga

menemukan nilai $f(x)$ yang bernilai sekecil atau sebesar mungkin suatu fungsi f yang diberikan, yang mungkin disertai dengan beberapa batasan pada x . Di mana x bisa berupa skalar atau vektor dari nilai-nilai kontinu maupun diskrit [7].

2.2.3 Twitter

Twitter atau yang saat ini berganti nama menjadi X sejak juli 2023 adalah media sosial daring dan layanan jejaring sosial yang dioperasikan oleh perusahaan Amerika Serikat X Corp, pengguna Twitter dapat memposting teks, gambar, dan video serta pengguna juga dapat memposting (*Tweet*), menyukai (*Like*) atau juga bisa memposting ulang (*retweet*), memberi komentar dan mengutip postingan hingga mengirim pesan langsung kepada pengguna lainnya.[17]

Tagar atau *Hashtag* (#) adalah sebuah penanda yang mengikat suatu topik. Tagar memiliki beberapa fungsi disosial media mulai dari pengelompokan konten dan memudahkan pencarian, tagar yang digunakan bisa juga ditambahkan foto, teks atau bahkan video. Tagar juga digunakan untuk kebutuhan bisnis atau sebuah acara.[18]

2.2.4 Drama Korea *Queen Of Tears*

Drama *Queen Of Tears* adalah drama televisi Korea yang disiarkan di Tvn mulai tanggal 9 maret hingga 28 April

2024, drama ini disutradarai oleh Jang Young-woo dan Kim Hee- Won dengan naskah yang ditulis oleh penulis ternama Park Ji-Eun. Drama ini mengisahkan tentang kisah cinta pasangan suami istri Baek Hyun-Woo yang diperankan oleh Kim soohyun dan Hong Hea-In yang diperankan oleh Kim Ji-Won. Drama ini turut ditayangkan melalui platform Netflix, memiliki jumlah episode 16 episode dengan tambahan 2 episode spesial yang berisi tentang *behind the scene* dari drama ini.[5]

2.2.5 Algoritma *Naive Bayes*

Naive Bayes atau yang terkadang disebut juga *Naive Bayes Classifier* adalah algoritma *machine learning probabilistik* yang digunakan dalam berbagai macam tugas klasifikasi. Berikut adalah rumus umum dari *Teorema Bayes* yang menjadi dasar *Naive Bayes*

$$P(H | X) = \frac{P(X | H) \cdot P(H)}{P(X)} \quad (2.1)$$

Penjelasan :

H : Hipotesis data X yang merupakan suatu class

X : Data dengan class yang belum diketahui

P(X) : Probabilitas prior bukti X

P(H) : adalah probabilitas prior hipotesis H

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

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

Algoritma *Naive Bayes* sebagian besar digunakan dalam penyaringan spam, analisis sentimen, hingga sistem rekomendasi. Alasan penggunaan algoritma *Naive Bayes* ini adalah karena implementasinya yang cepat dan mudah, *Naive Bayes* berhubungan erat dengan klasifikasi dan *machine learnig*. Terdapat dua tahapan dalam klasifikasi tweet, tahap pertama adalah pelatihan terhadap tweet yang telah diketahui kategorinya, sedangkan taha kedua adalah proses klasifikasi *tweet* yang belum diketahui kategorinya. Dalam algoritma *Naive Bayes* setiap dokumen direpresentasikan dengan atribut “ a_1, a_2, \dots, a_n ” dimana a_1 adalah kata pertama, a_2 adalah kata kedua dan seterusnya, sedangkan V adalah himpunan kategori *tweet*. Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi ($\arg \max$) dari semua kategori dokumen yang diujikan (VMAP).[19]

Untuk mencari VMAP maka dibutuhkan persamaan sebagai berikut :

$$V_{MAP} = \mathbf{arg\ max} \ P(v_j) \prod_i P(a_i | v_j) \quad (2.2)$$

Nilai $P(V_j)$ dapat dihitung pada saat training data, melalui persamaan sebagai berikut :

$$P(v_j) = \frac{|doc\ j|}{|training|} \quad (2.3)$$

dimana $|doc\ j|$ merupakan jumlah tweet yang memiliki kategori j dalam training. Sedangkan $|training|$ merupakan jumlah tweet dalam contoh yang digunakan untuk training. Untuk setiap probabilitas kata x_1 untuk setiap kategori $P(a_1|v_j)$, dihitung pada saat training dengan persamaan berikut:

$$P(a_i | v_j) = \frac{n_i + 1}{n + |kosakata|} \quad (2.4)$$

Dimana n_i adalah jumlah kemunculan kata a_i dalam tweet yang berkategori v_j , sedangkan n adalah banyaknya seluruh kata dalam tweet dengan kategori v_j dan $|kosakata|$ adalah banyaknya kata dalam data *training*.

2.2.6 Particle Swarm Optimization (PSO)

Particle Swarm Optimization adalah metode yang terinspirasi dari beberapa sifat alami sekelompok binatang seperti burung yang memiliki kebiasaan berkelompok dalam kesehariannya. Metode ini juga disebut sebagai algoritma *behaviorally inspired*, pada PSO terdapat tiga komponen penting antara lain adalah partikel, komponen kognitif dan

komponen sosial, serta kecepatan partikel. Pada setiap partikel tersebut mewakili soluis dari suatu permasalahan yang dihadapi, PSO dimulai dengan secara acak menginisialisasi segerombolan partikel diruang pencarian, setiap partikel diberi posisi dan vektor kecepatan, posisi mewakili solusi potensial untuk maslaah tersebut sedangkan vektor kecepatan menentukan arah dan kecepatan pergerakan partikel. Dalam PSO, populasi disebut dengan *Swarm* dan individu-individunya disebut partikel, *Swarm* didefinisikan sebagai himupna :

$$S = \{C_1, C_2, \dots, C_m\} \quad (2.5)$$

S adalah Populasi Partikel, C lokasi Partikel sedagkan m adalah banyaknya partikel (calon solusi), yang didefinisikan sebagai berikut :

$$C_i = \{C_{i1}, C_{i2}, \dots, C_{im}\}^T \in A, i = 1, 2, \dots, m \quad (2.6)$$

Partikel bergerak didalam ruangan pencarian dengan menyesuaikan dengan posisi masing-masing dari mereka dengan kecepatan (velocity) yang didefinisikan dengan persamaan berikut :

$$V_i = \{V_{i1}, V_{i2}, \dots, V_{im}\}^T \in A, i = 1, 2, \dots, m \quad (2.7)$$

Kecepatan disesuaikan dengan cara bertahap sehingga partikel mampu mengunjungi setiap wilayah A.

Jika t menunjukkan jumlah iterasi, maka posisi yang sedang berlangsung dari partikel ke- i dan kecepatannya dinotasikan sebagai $C_i(t)$ dan $V_i(t)$. Kecepatan diperbarui berdasarkan informasi yang diperoleh dari langkah algoritma sebelumnya, hal ini diimplementasikan dalam sebuah memori, dimana masing-masing partikel dapat menyimpan memori posisi terbaik yang pernah dikunjungi selama pencarian, sehingga S berisi posisi partikel saat ini. Memori yang disimpan oleh PSO adalah :

$$P = \{P_{1best}, P_{2best}, \dots, P_{mbest}\} \quad (2.8)$$

Yang berisi posisi terbaik

$$P_{ibest} = \{P_{i1}, P_{i2}, \dots, P_{in}\}^T \in A, i = 1, 2, \dots, m \quad (2.9)$$

dan pernah dikunjungi oleh masing-masing partikel.

Sehingga posisi ini didefinisikan sebagai berikut :

$$P_{ibest}(t) = \arg \min f_i(t) \quad (2.10)$$

dengan $\arg \min f_i(t)$ adalah partikel dengan $f_i(t)$ yang minimal. PSO didasarkan pada model simulasi perilaku sosial sehingga mekanisme pertukaran informasi harus ada untuk memungkinkan partikel saling berkomunikasi. Algoritma akan mendekati minimum global dengan posisi terbaik yang pernah dikunjungi oleh semua partikel.

Diberikan g sebagai posisi terbaik dengan nilai fungsi terendah dalam P pada t iterasi yang diberikan yaitu :

$$g_{ibest}(t) = \arg \min f (P_{ibest}(t)) \quad (2.11)$$

dengan $\arg \min f (P_{ibest}(t))$ adalah g_{ibest} dengan $f (g_{ibest}(t))$ yang minimal. Masing-masing partikel memperbarui kecepatan dan posisinya dengan persamaan sebagai berikut :

$$V_{in} = wV_{in} + c_1r_1(P_{ibest} - C_{in}) + c_2r_2(g_{ibest} - C_{in}) \quad (2.12)$$

$$C_{in} = C_{in} + V_{in} \quad (2.13)$$

Dengan w adalah koefisien inersia pada interval $[0,1]$, c_1c_2 adalah koefisien akselerasi yang berupa konstanta positif, r_1r_2 adalah bilangan acak pada interval $[0,1]$.

2.2.7 Confusion Matrix

Pada tahapan ini evaluasi dilakukan untuk mengetahui kinerja klasifikasi yang digunakan, *Confusion Matrix* adalah *matrix* yang menyajikan tabel hasil prediksi dari satu model klasifikasi, terdapat 4 nilai yang disajikan dalam confusion matrix [20]:

1. TP (*True Positive*) adalah jumlah data positif yang terklasifikasi benar sebagai positif.
2. TN (*True Negatif*) adalah jumlah data dari kelas negatif yang diprediksi benar masuk kelas negatif.
3. FP (*False Positive*) adalah jumlah data dari kelas negatif yang salah prediksi sebagai positive
4. FN (*False Negative*) adalah jumlah data positif yang salah sebagai negatif.

Berdasarkan nilai-nilai yang merupakan representasi dari kinerja klasifikasi tersebut dapat dihitung [21]:

1. Akurasi, perbandingan antara data yang telah terklasifikasi benar dengan keseluruhan data.

$$\text{Akurasi} = \frac{TP+TN}{TP+FN+FP+TN} \quad (3.1)$$

2. *Recall (Sensitivity)*, perbandingan antara data positif yang terklasifikasi benar dengan keseluruhan data positif.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3.2)$$

3. *Precision*, perbandingan antara data positif yang terklasifikasi benar dengan keseluruhan data yang terklasifikasi positif

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3.3)$$

4. *F-measure*, digunakan untuk pengukuran *precision* dan *recall* secara bersamaan

$$\text{F-measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (3.4)$$

2.2.8

Python

Python adalah bahasa pemrograman interpretatif yang dianggap mudah dipelajari serta fokus pada code yang mudah dibaca oleh sistem. Bahasa pemrograman *python* dikembangkan oleh Guido Van Rossum pada tahun 1990 di Amsterdam, Belanda sebagai kelanjutan dari bahasa pemrograman ABC. Keunggulan bahasa pemrograman ini adalah memiliki koleksi perpustakaan yang banyak, struktur bahasa yang sederhana dan mudah dipelajari, memiliki sistem pengolahan memori otomatis dan mudah dikembangkan dengan bahasa C / C++ [22].

2.3 Pemodelan

2.3.1 *Use Case Diagram*

Use Case Diagram adalah diagram yang menggambarkan fungsi, ruang lingkup dan interaksi dari pengguna dengan sistem yang dibuat. Fungsi dari *Use case diagram* antara lain [23]:

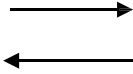
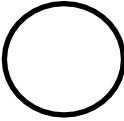




1. Sebagai bagan perinci kebutuhan pengguna
2. Pemodelan interaksi pengguna dan sistem
3. Menampilkan fungsionalitas sistem
4. Memfasilitasi komunikasi antara sistem dan pengguna
5. Memvalidasi pengkajian sistem

2.3.2 *Flowchart*

Flowchart (Bagan Alir) adalah bagan (chart) yang menunjukkan alir program atau prosedur sistem secara logika. Berikut adalah simbol – simbol dari *flowchart* serta fungsinya. *Flowchart* berfungsi untuk memudahkan dalam menganalisis proses dari sistem.

Tabel 2.1 Simbol *Flowchart*

Nama	Keterangan
Simbol	

	<p>Flow</p> <p>Simbol yang digunakan untuk menggabungkan antara simbol yang satu dengan simbol yang lain.</p>
	<p>On-Page Reference</p> <p>Simbol untuk keluar-masuk atau penyambungan proses dalam lembar kerja yang sama</p>
	<p>Off-Page Reference</p> <p>Simbol untuk keluar-masuk atau penyambungan proses dalam lembar kerja yang berbeda.</p>
	<p>Terminator</p> <p>Simbol yang menyatakan awal atau akhir</p>
	<p>Process</p> <p>Simbol yang menyatakan suatu proses yang dilakukan komputer</p>
	<p>Input/output</p> <p>Simbol yang menyatakan proses input atau output</p>

2.4 Perangkat Lunak Yang Digunakan

2.4.1 Google Colab

Google Colaboratory atau yang lebih umum disebut Googel Colab adalah compiler dan editor bahasa pemrograman python secara online, keunggulan menggunakan Google colabs ada banyak diantaranya tidak perlu instalasi, bisa berkolaborasi dengan sesam penggunanya, penggunaan GPU dan TPU gratis dan Pengolahan data cepat [24].

2.4.2 Microsoft Excel

Microsoft excel adalah sebuah program atau aplikasi yang dibuat oleh perusahaan Microsoft satu paket dengan *Microsoft Office*, dibuat dengan fungsi sebagai aplikasi pengolahan angka menggunakan spreadsheet yang terdiri dari baris dan kolom untuk mengeksekusi perintah [25]

BAB III

ANALISIS DAN PERANCANGAN

3.1 Gambaran Umum Objek Penelitian

Drama Queen of Tears adalah drama korea yang dibintangi oleh Kim Jiwoon dan Kim Soo Hyun, drama ini tayang pada bulan maret sampai april tahun 2024, sebanyak 16 episode. Drama ini mejandai trending 1 di aplikasi X atau yang dulu dikenal dengan Twitter, karena memiliki alur cerita yang bagus dan pemain yang profesional. Berbagai opini penonton dituangkan dalam bentuk tulisan di Twitter untuk menunjukkan berbagai tanggapan mengenai drama ini, dari survei internal Twitter mengenai drama ini da sebanyak 1530 ribu percakapan di twitter yang menggunakan tagar #queenoftears dengan berbagai macam opini penonton. Banyaknya opini penonton drama ini yang menunjukkan berbagai sentimen sehingga bisa jika dilakukan pengolahan data tweet yang benar maka akan memperoleh informasi yang sangat menguntungkan bagi pelaku industri drama baik diranah nasioanal maupun internasional.

3.1.1 *Crawling Data Twitter*

Data *Tweet* merupakan data yang diperoleh dari aplikasi twitter dengan cara crawling data Twitter. *Crawling* data Twitter pada penelitian ini menggunakan fungsi *Tweet Harvest*. *Tweet harvest* adalah produk yang dibuat oleh *Node.js* dimana pengguna harus memasukkan

Twitter Auth Token untuk bisa menuju halaman twitter yang digunakan untuk *crawling* data Twitter. Selanjutnya adalah proses mencari data yang dibutuhkan dengan memasukkan kata kunci yang sesuai dengan data yang dicari, dimana hal ini untk mendapatkan data yang relevan dengan penelitian ini memasukkan tagar dengan kata kunci *#queenoftears*, sehingga program akan mencari data yang sesuai dengan kata kunci yang sudah dimasukkan. Data yang digunakan dalam penelitian ini sebanyak 408 data yang hasil outputnya berupa data excel. Berikut adalah dataset yang berhasil dilakukan dengan cara *crawling* data Twitter menggunakan *Tweet harvest*. Berikut ini adalah tahapan dari proses *Crawling* data Twitter:

1. Masukkan *twitter Auth Token*
2. Masukkan keyword *#queenoftears* untuk mencari data tweet yang berhubungan dengan drama *queen of tears*
3. Menyimpan hasil *crawling* twitter dalam bentuk csv

conversation_id_str	created_at	favorit_count	full_text	id_str	image_url1	in_reply_to_screen_name	lang	location	quote_count	reply_count	retweet_count	tweet_url	user_id_str	username	tanggal	waktu
0	171915071128457757 Tue Jun 04 15:25:58 +000 2024	0	@QueenofTears @chryme_martina queen of tears gash	1758014338112814847		NaN	rawantha	in	NaN	0	1	0	https://x.com/faiyocapri/status/1758014338172...	1200163095174702362	faiyocapri	Tue Jun 04 15:25:58
1	1758013807615729505 Tue Jun 04 15:27:51 +000 2024	2	Lu tau nih ada yg ngidat begini pis wajah Pda...	1758013807615729505	https://ps.twimg.com/media/GPPI8A4sAAH0D.jpg		NaN	in	Indonesia	1	0	0	https://x.com/tutu_sgh/status/1758013807615729505	1310811045386133762	tutu_sgh	Tue Jun 04 15:27:51
2	1738008763867426825 Tue Jun 04 15:27:26 +000 2024	1	@omkassalus He kabin sudah pake make	1758013702769397973		NaN	omkassalus	in	NaN	0	1	0	https://x.com/dorminny/status/17580137027...	10272885070663745	dorminny	Tue Jun 04 15:27:26
3	1758011235787916936 Tue Jun 04 15:17:38 +000 2024	0	queen of tears each night pastoral mission org.	1758011235787916936		NaN	NaN	in	us23	0	0	0	https://x.com/taagony/status/1758011235787...	129532189322048334	taagony	Tue Jun 04 15:17:38
4	1758004792328439950 Tue Jun 04 15:17:29 +000 2024	0	@Charismad_ringsari anjali dulu	1758011963070784419		NaN	lennaliquyus	in	NaN	0	0	0	https://x.com/lennaliquyus/status/1758011963...	9020803059802268	lennaliquyus	Tue Jun 04 15:17:29

Gambar 3.1 Data Hasil *Crawling* Twitter

Dari gambar 3.1 data yang diperoleh terdapat beberapa atribut yakni *conversation_id_str*, *created_at*, *favorit_count*, *full_text*, *id_str*, *image_url*,

in reply to sreen name, lang, location, quote_count, reply_cpunt, retweet_count, tweet_url, user_id_str, username, tanggal dan waktu.

3.2 Alur Proses

Alur prose merupakan gambaran dari suatu proses yang akan dilakukan dalam sebuah pemecahan masalah, proses ini akan menjelaskan secara rinci mengenai tahapan dari penelitian. Dengan adanya proses ini maka akan memudahkan penelitian sehingga berjalan sesuai dengan rancangan yang diinginkan.

3.2.1 Identifikasi dan Analisis proses Bisnis

Pada bagian ii dilakukan identifikasi terkait denga objek penelitian yang sedang dilakukan yaitu data Tweet drama *Queen Of Tears*, pada tahap ini akan dijabarkan proses yang terkait sehingga data siap digunakan.

- 1) *Preprocessing data*
- 2) *Labeling data*
- 3) *Metode Naive bayes*
- 4) *Metode Particle Swarm optimization*

3.2.2 Analisis Proses Bisnis

Pada proses ini dilakaukan analisis sebelum proses pembahasan objek dilakukan. Pada umunya proses bisnis berkaitan dengan tahapan pengerjaan terkait obejek penelitian. Dengan analisis ini akan mempermudah penelitian dalam menjabarkan proses yang diperlukan

dalam menyelesaikan permasalahan objek penelitian. Berikut adalah analisis proses bisnis yang diperlukan dalam penelitian ini.

a. *Preprocessing Data*

Pada tahap ini, dilakukan *preprocessing* data yang bertujuan untuk menyeleksi data teks agar menjadi lebih terstruktur dengan melalui beberapa tahapan berikut, gambar 3.2 menjelaskan rangkaian proses *preprocessing* [26].



Gambar 3.2 Tahapan *Preprocessing*

- 1) *Cleaning*, merupakan proses dimana karakter dan tanda baca yang tidak diperlukan dihapus dari teks. Proses ini bertujuan untuk mengurangi gangguan pada dataset, tahapan *cleaning* adalah hapus *URL*, hapus *username*, hapus tanda RT dan hapus *Hashtag*.
- 2) *Case Folding*, merupakan proses untuk mengubah semua karakter teks menjadi huruf kecil serta menghilangkan tanda baca dan angka. Cara kerja *case folding* adalah memproses huruf alphabet dari “a” hingga “z” sehingga karakter selain huruf tersebut akan dihapus.
- 3) *Tokenizing*, adalah proses memecah kalimat menjadi kata-kata atau memutus urutan string menjadi potongan kata.

- 4) *Stopwords*, adalah proses menghilangkan kosakata yang bukan termasuk ciri pada suatu dokumen atau tidak menyampaikan pesan apapun secara signifikan pada teks atau kalimat.
- 5) *Stemming*, yakni suatu proses untuk mendapatkan kata dasar dengan cara menghilangkan awalan, akhiran, sisipan dan cofilixes.

b. Metode *Naive Bayes*

Berikut ini adalah langkah-langkah pengujian pada algoritma *Naive Bayes Classifier* :

- 1) Menghitung probabilitas dari V_j pada data training dengan persamaan (2.3) dimana V_j merupakan Kategori sentimen, yaitu $V_1 = \text{negatif}$ dan $V_2 = \text{positif}$.
- 2) Menghitung probabilitas kata a_i pada kategori V_j dengan persamaan (2.4)
- 3) Model probabilitas NBC disimpan dan digunakan untuk tahap testing.
- 4) Menghitung probabilitas tertinggi dari kategori sentimen yang diujikan (V_{map}) dengan persamaan (2.1)
- 5) Mencari nilai V_{map} paling maksimum dan masukkan tweet tersebut pada kategori dengan V_{map} maksimum.

c. Metode *Particle Swarm Optimization*

Berikut adalah langkah-langkah optimasi hasil dari metode *Naive Bayes* menggunakan metode optimasi *Particle Swarm Optimization (PSO)* :

- 1) Inisialisasi posisi awal partikel (C_i) dan kecepatan awal partikel (V_i), dengan $i = 1, 2, \dots, m$ dan m adalah ukuran Swarm.
- 2) Evaluasi nilai fungsi tujuan untuk setiap partikel ($f(C_i)$).
- 3) Tentukan P_{ibest} awal dan G_{ibest} awal
- 4) Update kecepatan dengan persamaan (2.12)
- 5) Update posisi partikel baru dengan persamaan (2.13)
- 6) Evaluasi $F(C_i)$, jika $F(C_i) \leq F(P_{ibest})$ maka $P_{ibest} = C_i$, setelah mendapatkan P_{ibest} baru, maka didapatkan g_{best} baru.
- 7) Jika iterasi sudah maksimum atau konvergen maka algoritma berhenti, jika belum maksimum maka kembali ke langkah 4.

d. *Confusion Matrix*

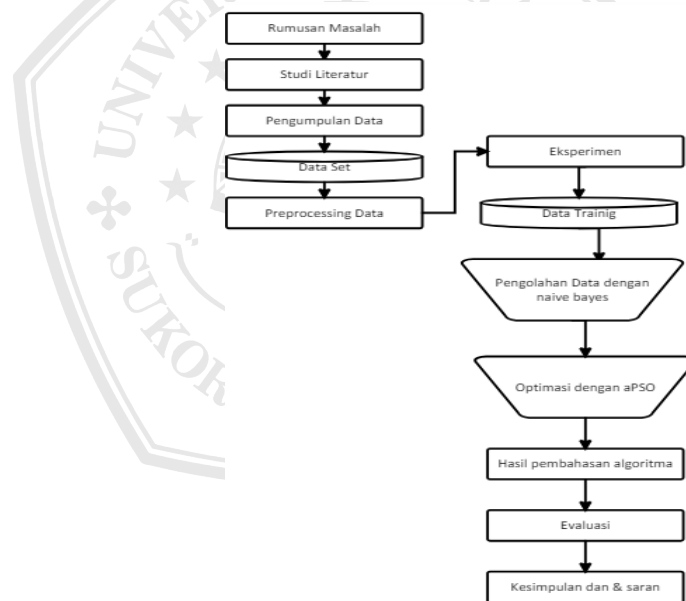
Menghitung ketepatan klasifikasi dan optimasi dari algoritma *Naive Bayes Classifier* dan optimasi PSO berdasarkan hasil *Confusion Matrix* yang diperoleh, berikut langkah-langkah perhitungannya :

- 1) Menghitung tingkat akurasi dengan persamaan (3.1)
- 2) Menghitung recall dengan persamaan (3.2)
- 3) Menghitung tingkat presisi dengan persamaan (3.3)
- 4) Menghitung F-measure dengan persamaan (3.4)

e. *Flowchart* Dokumen Data

1) Tahapan Analisis Sentimen

Proses ini merupakan tahapan dalam melakukan analisis sentimen pada data Twitter. Pada gambar 3.3 merupakan alur proses analisis dilakukan.



Gambar 3.3 Tahapan Analisi Sentimen

3.2.3 Identifikasi dan Analisis Kebutuhan

a. Identifikasi Kebutuhan Fungsional

Sebuah proses akan berjalan lancar jika kebutuhan fungsional terpenuhi, sehingga peran kebutuhan fungsional sangatlah penting dalam sebuah penelitian.

b. Analisis Kebutuhan Fungsional

Pada tahap ini data akan diolah menggunakan sebuah tahapan data yakni *Preprocessing* data. Data diolah melalui enam tahapan yaitu *Cleaning*, *case folding*, *normalisasi*, *tokenizing*, *stopword*, dan *steaming*. *Preprocessing* dilakukan agar data yang didapat setelah proses *crawling* data Twitter yang merupakan data mentah, setelah dilakukan *preprocessing* akan menjadi data yang siap digunakan untuk analisis sentimen menggunakan metode yang telah ditentukan, sehingga data tidak mengalami error karena adanya data yang tidak memiliki nilai. Setelah proses *preprocessing* selesai dilakukan data akan melewati proses pelabelan dalam penelitian ini ada 3 label yang digunakan yakni label yang bernilai positif, negatif dan netral. Setelah data melewati proses pelabelan data akan dibagi menjadi data training dan data testing, hal ini dilakukan untuk memudahkan proses implementasi metode algoritma yang akan digunakan yakni algoritma *Naive bayes* dan algoritma *Particle Swarm optimization*. Pada proses perhitungan menggunakan algoritma *Naive Bayes* akan mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat. Setelah mendapatkan nilai probabilitas tertinggi pada data testing maka tahap

selanjutnya adalah proses optimasi menggunakan metode *Particle Swarm Optimization*, algoritma ini akan mencari nilai maksimal dari jumlah populasi yang ada.

3.2.4 Identifikasi dan Analisis Alternatif Solusi

a. identifikasi Alternatif Solusi

Proses ini berfungsi untuk menganalisis alternatif apa saja yang digunakan dalam membuat program. Tabel 3.1 adalah tabel yang merupakan identifikasi alternatif solusi. Terdapat dua elternatif solusi dalam penelitian ini.

Tabel 3.1 Identifikasi Alternatif Solusi

Karkteristik	Alternatif 1 (Python)	Alternatif 2 (RapidMiner)
Bagian sistem yang terkomputerisasi.	Semua kebutuhan fungsional akan terpenuhi	Semua kebutuhan fungsional akan terpenuhi.
Keuntungan, deskripsi ringkas keuntungan yang akan direalisasikan	Pengembanagn lebih cepat dan mudah dilakukan bagi pemula	Proses lebih cepat bagi pemula

Tabel 3.1 (Lanjutan)

Perangkat lunak aplikasi	perangkat lunak yang akan digunakan adalah web browser Google Colab menggunakan bahas python	Perangkat lunaka yang digunakan yang akan digunakan adlah GUI, Java API dan data mining.
Metode pemrosesan data, pada biasanya kombinasi online, batch, deferred, remote batch dan reel time	Server gratis yang sudah disediakan oleh google	Data mining
Alat Output	Monitor dan Printer	Monitor dan Printer
Alat Input	Keybord dan Mouse	Keyboard, Mouse dan Microsoft excel
Alat penyimpanan	Data disimpan dalam google drive	Data disimpan dalam pabel repositori

b. Analisis Kelayakan Alternatif Solusi

Proses ini berfungsi menganalisis sistem yang akan dibuat berdasarkan dokumen yang didapatkan sebelumnya.

Tabel 3.2 menjelaskan identifikasi kelayakan alternatif.

Tabel 3.2 Analisis Kelayakan

Kriteria Kelayakan	Bobot	Alternatif 1 (Python)	Alternatif 2 (RapidMiner)
Kelayakan Operasional Fungsional	40%	Mendukung seluruh kebutuhan fungsional serta pengembangan yang lebih mudah dan bisa dikerjakan oleh banyak orang	Mendukung seluruh kebutuhan fungsional, mudah dilakukan oleh pemula.
Skor		40	30
Kelayakan teknis teknologi	30%	Teknologi yang digunakan sudah sangat mencukupi dalam keahlian	Teknologi yang digunakan dalam manajemen sudah memadai dan sudah cukup.
Skor		25	30

Tabel 3.2 (Lanjutan)

Kelayakan ekonomis biaya pengembangan	30%	Pengembangan sistem berbasis kolaborasi serta open source, tidak memerlukan biaya dalam pengoperasiannya	Tidak ada pengembangan dan bersifat komersial
Skor		30	20
Total	100%	95	80

3.2.5 Desain Proses

desain proses merupakan tahapan dalam merancang sebuah sistem. Tahapan ini dilakukan agar proses sistem berjalan sesuai dengan alur yang diinginkan.

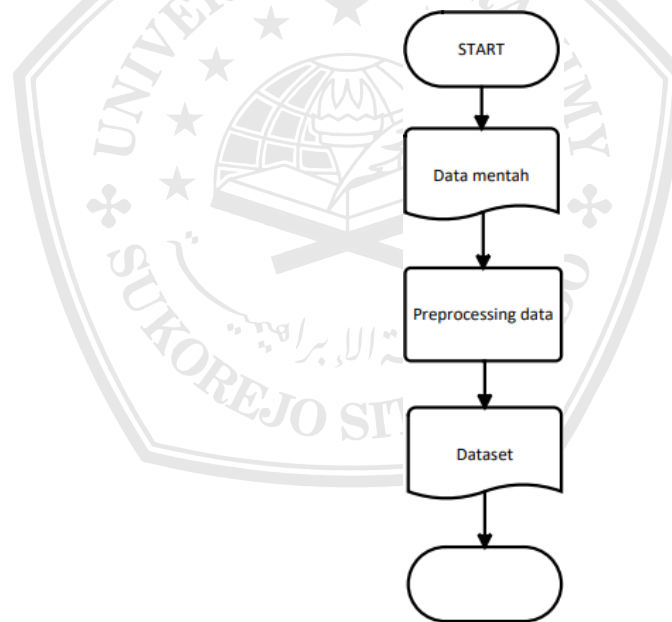
a. Identifikasi Desain Proses

Pada identifikasi desain proses menjelaskan tentang alur dari analisis yang dilakukan. Bagian ini fokus pada proses yang akan dilakukan yakni proses input data, perhitungan dengan metode

dan evaluasi hasil dari metode. Berikut adalah desain proses yang akan dilakukan.

1. *Preprocessing Data*

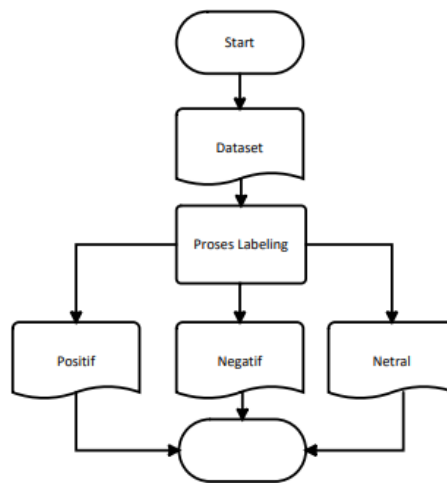
Proses ini dilakukan untuk mengolah data menjadi sebuah data set. Input data Excel yang diperoleh dari hasil crawling data twitter output yang dihasilkan berupa dataset yang berupa file Excel yang sudah siap dilakukan proses perhitungan dengan algoritma. Gambar 3.4 menjelaskan tentang alur proses *Preprocessing data*.



Gambar 3.4 Preprocessing data

2. *Labeling Data*

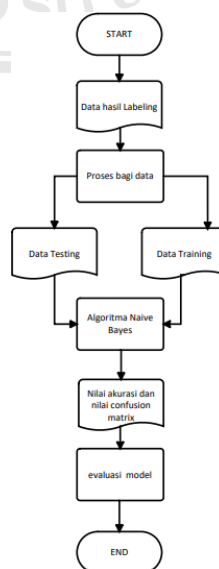
Input data yang sudah dilakukan preprocessing, hasil labeling, sentimen positif, negatif dan netral. Gambar 3.5 merupakan alur proses labeling data.



Gambar 3.5 Alur Labeling

3. Metode *Naive Bayes*.

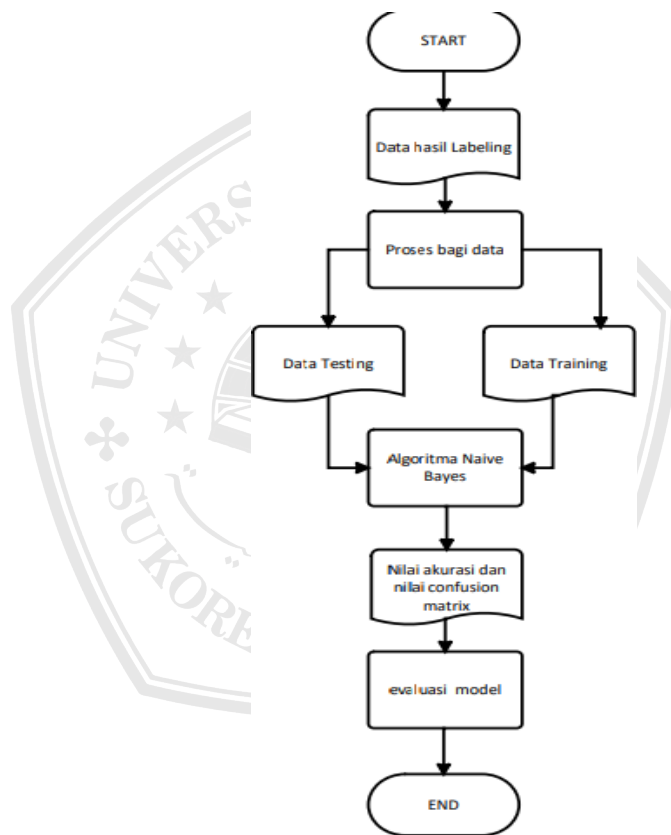
Input data hasil labeling, bagi menjadi data latih dan data uji untuk menghitung akurasi, diinput data Excel dari hasil labeling data. Output yang dihasilkan adalah Nilai akurasi dan *confusion matrik*. Gambar 3.6 menjelaskan alur metode *Naive Bayes*



Gambar 3.6 Alur metode NBC

4. Metode *Particle Swarm Optimization*

Input data hasil uji metode *Naive Bayes* untuk menguji nilai optimasi dari PSO, data yang diinput berupa Excel. Output yang dihasilkan adalah Nilai akurasi dari pso dan confusion matrix. Gambar 3.7 menjelaskan alur metode PSO.

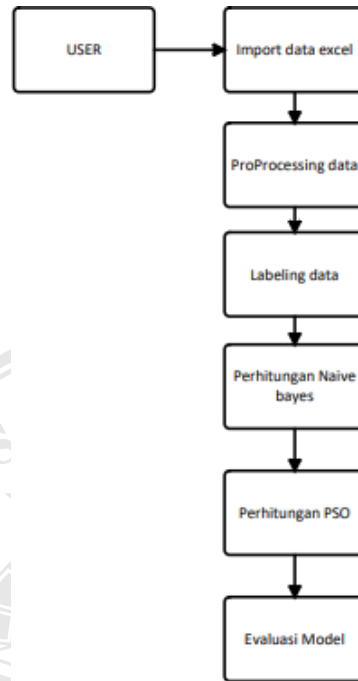


Gambar 3.7 Alur metode PSO

b. Arsitektur Sistem

Arsitektu sistem berfungsi sebagai penjelasan dari pola sistem analisis yang akan dilakukan, berikut adalah arsitektur

sistem yang akan dilakukan. Gambar 3.8 menjelaskan arsitektur sistem



Gambar 3.8 Arsitektur Sistem

BAB IV

IMPLEMENTASI SISTEM

4.1 Konstruksi Sistem

4.1.1 Kebutuhan Sistem

Dalam bahasa pemrograman *Python* sistem dapat dijalankan dengan memperhatikan beberapa perangkat yang bisa menunjang keberhasilan sistem. Berikut adalah perangkat yang digunakan.

1. *Software*

Software yang digunakan dalam mendukung sistem adalah sebagai berikut :

- a) Google Colab
- b) Microsoft excel

2. *Hardware*

- a) Laptop Hp
- b) Mouse dan Keyboard
- c) Processor AMD Ryzen 3
- d) Hard Disk dengan kapasitas 250 GB
- e) RAM 8 GB

4.2 Pengumpulan Data

Tahap awal dalam pengumpulan data yakni menggunakan cara *crawling* data pada aplikasi Twitter menggunakan *Tweet-Harvest*. *Tweet-harvets* adalah alat untuk *crawling* data dari aplikasi Twitter melalui

bahas pemrograman python. Dalam mengambil data twitter menggunakan *tweet-harvest* ini memiliki beberapa kelebihan yakni mudah digunakan serta tidak terbatas waktu *crawling* data serta tidak memiliki batasan data yang bisa diambil, berikut adalah langkah-langkah *crawling* data dengan *tweet-harvest* :

1. Masukkan *Twitter Auth Token* yang ada dihalaman Twitter
2. Install Node Js, hal ini diperlukan karena *tweet – harvest* diciptkan oleh Node.Js
3. Masukkan kata kunci yang ingin dicari atau bisa menggunakan hashtag yang bersangkutan dengan objek data yang ingin dicari.
4. Masukkan jumlah data yang ingin diambil
5. Menyimpan data kedalam bentuk file csv.

Berikut adalah gambaran dari proses *crawling* data twitter menggunakan *tweet-harvest* :

Segmen Program 4.1 *Twitter Auth Token*

```
#@title Twitter Auth Token

twitter_auth_token = '86b9146eec2fcc151cb3501160f59c4660fdd' # change this auth
token
```

Segmen Program 4.2 *Crawling* Data Twitter

```
# Crawl Data

filename = 'dramaqot2.csv'

search_keyword = '#queenoftears since:2024-03-09 until:2024-04-28 lang:id'

limit = 100

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -
l {limit} --token {twitter_auth_token}
```

4.3 *Text Preprocessing*

Dalam melakukan preprocessing ada beberapa tahapan yang perlu dilakukan. Adapun tahapan preprocessing adalah sebagai berikut :

1. *Cleaning*

Adalah tahap awal dari proses *preprocessing* tahap ini merupakan tahap pembersihan atau eliminasi aksara nonalfabetis hal ini dilakukan agar menurunkan noise atau error pada saat pengolahan data. Dalam tahap ini dilakukan penghapusan URL, *mention*, *link double*, *whitespace*, *emoticon*, duplikasi teks dan angka. Dalam proses ini menggunakan fungsi library NLTK (*natural Language Toolkit*) pada bahasa pemrograman *Python*. Tabel 4.1 menjelaskan teks sebelum dan sesudah *cleaning* data.

Tabel 4.1 Proses Cleaning Tweet

Full Text	Cleaning
yeontan_juseyo,Kayaknya aku lebih cocok romance sejenis lovely runner daripada romance queen of tears	Kayaknya aku lebih cocok romance sejenis lovely runner daripada romance queen of tears
Daniamee5,Trending di aku #queenoftears #kimsoohyun #kimjiwon	Trending di aku queenoftears kimsoohyun kimjiwon
Leuxninie,Dari dulu ngedrakor baru kali ini deh ni drama #QueenOfTears bikin se gamon itu maunya tetep lanjut aje Btw itu yg baju itemÂ² juga semoga lanjut ampe pelaminan yee hehe #KimSooHyun #KimJiwon	Dari dulu ngedrakor baru kali ini deh ni drama QueenOfTears bikin se gamon itu maunya tetep lanjut aje Btw itu yg baju item juga semoga lanjut ampe pelaminan yee hehe KimSooHyun KimJiwon

Segmen Program 4.3 menjelaskan tentang segmen program cleaning data yang ada pada sistem ini untuk menghapus kata yang tidak bernilai

Segmen Program 4.3 Proses cleaning

```
import re

import string

import nltk

# Fungsi untuk menghapus URL
def remove_URL(tweet):

    if tweet is not None and isinstance(tweet, str):

        url = re.compile(r'https?://\S+|www\.\S+')

        return url.sub(r'', tweet)

    else:

        return tweet

# Fungsi untuk menghapus HTML
def remove_html(tweet):

    if tweet is not None and isinstance(tweet, str):

        html = re.compile(r'<.*?>')

        return html.sub(r'', tweet)

    else:

        return tweet

# Fungsi untuk menghapus emoji
def remove_emoji(tweet):

    if tweet is not None and isinstance(tweet, str):

        emoji_pattern = re.compile("[

u"\U0001F600-\U0001F64F" # emoticons

u"\U0001F300-\U0001F5FF" # symbols & pictographs

u"\U0001F680-\U0001F6FF" # transport & map symbols

u"\U0001F700-\U0001F77F" # alchemical symbols

u"\U0001F780-\U0001F7FF" # Geometric Shapes Extended
```

Lanjutan Segmen Program(4.3)

```
u"\U0001F800-\U0001F8FF" # Supplemental Arrows-C
    u"\U0001F900-\U0001F9FF" # Supplemental Symbols and
Pictographs
    u"\U0001FA00-\U0001FA6F" # Chess Symbols
    u"\U0001FA70-\U0001FAFF" # Symbols and Pictographs Extended-
A
    u"\U0001F004-\U0001F0CF" # Additional emoticons
    u"\U0001F1E0-\U0001F1FF" # flags
    "+"], flags=re.UNICODE)
return emoji_pattern.sub(r", tweet)
else:
    return tweet
# Fungsi untuk menghapus simbol
def remove_symbols(tweet):
    if tweet is not None and isinstance(tweet, str):
        tweet = re.sub(r'^a-zA-Z0-9\s]', ", ", tweet) # Menghapus semua simbol
    return tweet
# Fungsi untuk menghapus angka
def remove_numbers(tweet):
    if tweet is not None and isinstance(tweet, str):
        tweet = re.sub(r'd', ", ", tweet) # Menghapus semua angka
    return tweet
def remove_username(text):
    import re
    return re.sub(r'@[^\s]+', ", ", text)
df['cleaning'] = df['full_text'].apply(lambda x: remove_URL(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_username(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_html(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_emoji(x))
df['cleaning'] = df['cleaning'].apply(lambda x: remove_symbols(x))
```

```
df['cleaning'] = df['cleaning'].apply(lambda x: remove_numbers(x))
df.head(5)
```

2. Case Folding

Case folding merupakan proses yang berfungsi untuk menyamaratakan penggunaan huruf kapital pada suatu text menjadi huruf kecil. Proses ini dilakukan menggunakan fungsi pandas yaitu `string.lower()`. Setelah itu data yang telah diproses akan menampilkan data dengan teks yang telah diubah ke huruf kecil. Tabel 4.2 menjelaskan tentang proses case folding dan Segmen Program 4.4 berisi tentang proses case folding.

Tabel 4.2 Proses Case Folding

Sebelum Case Folding	Case Folding
Kayaknya aku lebih cocok romance sejenis lovely runner daripada romance queen of tears	kayaknya aku lebih cocok romance sejenis lovely runner daripada romance queen of tears
Trending di aku queenoftears kimsoohyun kimjiwon	trending di aku queenoftears kimsoohyun kimjiwon
Dari dulu ngedrakor baru kali ini deh ni drama QueenOfTears bikin se gamon itu maunya tetep lanjut aje Btw itu yg baju item juga semoga lanjut ampe	dari dulu ngedrakor baru kali ini deh ni drama queenoftears bikin se gamon itu maunya tetep lanjut aje btw itu yg baju item juga semoga lanjut ampe

pelaminan yee hehe KimSooHyun KimJiwon	pelaminan yee hehe kimsoohyun kimjiwon
---	---

Segmen program 4.4 Proses Case folding

```
def case_folding(text):
    if isinstance(text, str):
        lowercase_text = text.lower()
        return lowercase_text
    else:
        return text
df['case_folding'] = df['cleaning'].apply(case_folding)
df.head(5)
```

3. Normalisasi

Pada tahap normalisasi ini kata yang sebelumnya mempunyai makna yang tidak baku atau tidak sesuai dengan ejaan maka akan diubah sesuai dengan kamus kata baku yang dibuat berdasarkan hasil dari pengamatan dari dataset yang akan diproses.

Tabel 4.3 menjelaskan Normalisasi dan segmen program 4.5 berisi tentang code program proses normalisasi.

Tabel 4.3 Proses Normalisasi

Sebelum Normalisasi	Setelah Normalisasi
kayaknya aku lebih cocok romance sejenis lovely runner daripada romance queen of tears	kayaknya aku lebih cocok romance sejenis lovely runner

	daripada romance queen of tears
--	------------------------------------

Lanjutan tabel 4.3

trending di aku queenoftears kimsoohyun kimjiwon	trending di aku queenoftears kimsoohyun kimjiwon,
dari dulu ngedrakor baru kali ini deh ni drama queenoftears bikin se gamon itu maunya tetep lanjut aje btw itu yg baju item juga semoga lanjut ampe pelaminan yee hehe kimsoohyun kimjiwon	dari dulu ngedrakor baru kali ini deh nih drama queenoftears bikin se gamon itu maunya tetap lanjut saja btw itu yang baju item juga semoga lanjut sampai pelaminan ya hehe kimsoohyun kimjiwon

SegmenProgram 4.5 Proses Normalisasi

```
import pandas as pd

# Fungsi penggantian kata tidak baku

def replace_taboo_words(text, kamus_tidak_baku):

    if isinstance(text, str):

        words = text.split()

    replaced_words = []

    kalimat_baku = []

    kata_diganti = []

    kata_tidak_baku_hash = []
```

```
for word in words:
```

Lanjutan Segmen Program 4.5

```
    if word in kamus_tidak_baku
        baku_word = kamus_tidak_baku[word]
        if isinstance(baku_word, str) and all(char.isalpha() for char
in baku_word):
            replaced_words.append(baku_word)
            kalimat_baku.append(baku_word)
            kata_diganti.append(word)
            kata_tidak_baku_hash.append(hash(word))
        # else:
        #     replaced_words.append('<karakter eksplisit>')
        #     kalimat_baku.append(word)
        else:
            replaced_words.append(word)
        replaced_text = ''.join(replaced_words)
    else:
        replaced_text = "
        kalimat_baku = []
        kata_diganti = []
        kata_tidak_baku_hash = []

return replaced_text, kalimat_baku, kata_diganti, kata_tidak_baku_hash
```

4. Tokenizing

Pada tahap ini merupakan proses pemecahan kalimat berdasarkan setiap kata dalam proses penyusunan suatu kalimat.

Tabel 4.6 menjelaskan tentang proses Tokenizing dan

Tabel 4.4 Proses Tokenizing

Sebelum Tokenizing	Setelah Tokenizing
kayaknya aku lebih cocok romance sejenis lovely runner daripada romance queen of tears	'kayaknya', 'aku', 'lebih', 'cocok', 'romance', 'sejenis', 'lovely', 'runner', 'daripada', 'romance', 'queen', 'of', 'tears'
trending di aku queenoftears kimsoohyun kimjiwon,	'trending', 'di', 'aku', 'queenoftears', 'kimsoohyun', 'kimjiwon'
dari dulu ngedrakor baru kali ini deh nih drama queenoftears bikin se gamon itu maunya tetap lanjut saja btw itu yang baju item juga semoga lanjut sampai pelaminan ya hehe kimsoohyun kimjiwon	'dari', 'dulu', 'ngedrakor', 'baru', 'kali', 'ini', 'deh', 'nih', 'drama', 'queenoftears', 'bikin', 'se', 'gamon', 'itu', 'maunya', 'tetap', 'lanjut', 'saja', 'btw', 'itu', 'yang', 'baju', 'item', 'juga', 'semoga', 'lanjut', 'sampai', 'pelaminan', 'ya', 'hehe', 'kimsoohyun', 'kimjiwon'

Tabel 4.7 berisi tentang segmen program proses tokenizing dimana proses ini merupakan proses keempat dari proses preprocessing yang berfungsi sebagai pemecah kalimat .

Segmen Program 4.6 Proses Tokenizing

```
def tokenize(text):
    tokens = text.split()
    return tokens

df['tokenize'] = df['normalisasi'].apply(tokenize)
df.head(5)
```

5. Stopword removal

Pada tahapan ini kata yang tidak baku akan dihapus dan dianggap tidak penting dalam dokumen, seperti kalo, bikin, masih, ke, atau dll. Tahapan ini dilakukan setelah proses Tokenizing. Tabel 4.5 menjelaskan tentang proses Stopword Removal dan segmen program 4.7 berisi tentang code program proses Stopword Removal.

Tabel 4.5 Proses Stopword Removal

Sebelum Stopword Removal	Setelah Stopword Removal
'kayaknya', 'aku', 'lebih', 'cocok', 'romance', 'sejenis', 'lovely', 'runner', 'daripada', 'romance', 'queen', 'of', 'tears'	'kayaknya', 'cocok', 'romance', 'sejenis', 'lovely', 'runner', 'romance', 'queen', 'of', 'tears'

'trending', 'di', 'aku',	'trending', 'queenoftears',
'queenoftears', 'kimsoohyun',	'kimsoohyun', 'kimjiwon'
'kimjiwon'	

Lanjutan Tabel 4.5

'dari', 'dulu', 'ngedrakor', 'baru', 'kali', 'ini', 'deh', 'nih', 'drama', 'queenoftears', 'bikin', 'se', 'gamon', 'itu', 'maunya', 'tetap', 'lanjut', 'saja', 'btw', 'itu', 'yang', 'baju', 'item', 'juga', 'semoga', 'lanjut', 'sampai', 'pelaminan', 'ya', 'hehe', 'kimsoohyun', 'kimjiwon'	'ngedrakor', 'kali', 'deh', 'nih', 'drama', 'queenoftears', 'bikin', 'gamon', 'maunya', 'btw', 'baju', 'item', 'semoga', 'pelaminan', 'ya', 'hehe', 'kimsoohyun', 'kimjiwon'
---	---

Segmen Program 4.7 Proses Stopword Removal

```
def remove_stopwords(text):
    return [word for word in text if word not in stop_words]

df['stopword removal'] = df['tokenize'].apply(lambda x:
remove_stopwords(x))
```

```
df.head(5)
```

6. Steaming

Tahapan ini merupakan proses terakhir dari preprocessing dalam tahap ini kata yang memiliki imbuhan akan dipecah menjadi kata dasar. Dalam proses ini menggunakan library sastrawi yang mana proses ini dilakukan setelah proses normalisasi, tokenizing dan stopword selesai dilakukan. Tabel 4.6 menjelaskan tentang proses Steaming dan segmen program 4.8.

Tabel 4.6 Proses Steaming

Sebelum Stopword Removal	Setelah Stopword Removal
'kayaknya', 'cocok', 'romance', 'sejenis', 'lovely', 'runner', 'romance', 'queen', 'of', 'tears'	kayak cocok romance jenis lovely runner romance queen of tears
'ngedrakor', 'kali', 'deh', 'nih', 'drama', 'queenoftears', 'bikin', 'gamon', 'maunya', 'btw', 'baju', 'item', 'semoga', 'pelaminan', 'ya', 'hehe', 'kimsoohyun', 'kimjiwon'	ngedrakor kali deh nih drama queenoftears bikin gamon mau btw baju item moga pelamin ya hehe kimsoohyun kimjiwon

Segmen Program 4.8 Proses Steaming

```
factory = StemmerFactory()
```

```

stemmer = factory.create_stemmer()

def stem_text(text):

    return [stemmer.stem(word) for word in text]

df['stemming_data'] = df['stopword removal'].apply(lambda
x: ''.join(stem_text(x)))

df.head(5)
    
```

Berikut adalah gambar 4.1 yang menampilkan hasil dari proses preprocessing pada google colab :

full_text	cleaning	case_folding	normalisasi	tokenize	stopword removal	stemming_data
gw nangisin ini dri tdi #QueenOfTears https://...	gw nangisin ini dri tdi QueenOfTears	gw nangisin ini dri tdi queenoftears	gue nangisin ini dari tadi queenoftears	[gue, nangisin, ini, dari, tadi, queenoftears]	[gue, nangisin, queenoftears]	gue nangisin queenoftears
serius nanya nonton interview yg queen of tear...	serius nanya nonton interview yg queen of tear...	serius nanya nonton interview yg queen of tear...	serius bertanya menonton interview yang queen ...	[serius, bertanya, menonton, interview, yang, ...]	[serius, menonton, interview, queen, of, tears...]	serius tonton interview queen of tears lengkap...
BAEK HYUNWOO #QueenOfTearsEp15 #QueenOfTears h...	BAEK HYUNWOO QueenOfTearsEp QueenOfTears	baek hyunwoo queenoftearsep queenoftears	baik hyunwoo queenoftearsep queenoftears	[baik, hyunwoo, queenoftearsep, queenoftears]	[hyunwoo, queenoftearsep, queenoftears]	hyunwoo queenoftearsep queenoftears
delulu is solululu #queenoftears #soowon	delulu is solululu queenoftears soowon	delulu is solululu queenoftears soowon	delulu is solululu queenoftears soowon	[delulu, is, solululu, queenoftears, soowon]	[delulu, is, solululu, queenoftears, soowon]	delulu is solululu queenoftears soowon
Cegil mana yang drama favorit nya uda tamat ta...	Cegil mana yang drama favorit nya uda tamat ta...	cegil mana yang drama favorit nya uda tamat ta...	cegil mana yang drama favorit nya sudah tamat ...	[cegil, mana, yang, drama, favorit, nya, sudah...]	[cegil, drama, favorit, nya, tamat, nangisin, ...]	cegil drama favorit nya tamat nangisin part sa...

Gambar 4.1 Hasil Preprocessing Data

4.4 Wordcloud

Dalam visualisasi data dengan memakai wordcloud secara otomatis yang dilakukan sebelum proses preprocessing didapatkan visualisasi data dengan frekuensi kata yang didominasi oleh kata Queen of Tears, Kim Soo Hyun, Drama, Kim Jiwoon. Gambar 4.2 adalah wordcloud data sebelum dilakukan preprocessing data.

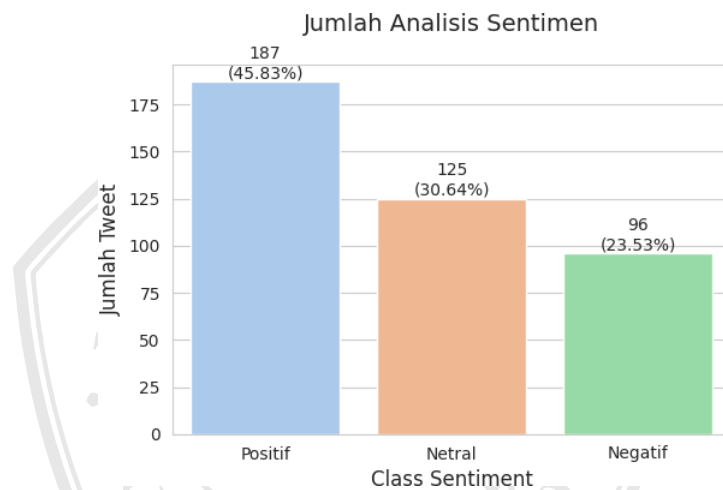
1. Text Mining : tahapan pertama adalah Preprocessing dataset dengan melalui beberpa tahapan yaitu cleaning, case folding, normalisasi, tokenizing, stopword, dan steaming.
2. Setelah melalui preprocessing langkah selanjutnya adalah mengekstraksi kata yang memiliki sentimen positif dan sentimen negatif dengan membandingkannya dengan kamus lexicon positif dan kamus lexicon negatif.
3. Menentukan sentimen dengan menghitung kemunculan dari kata yang dihitung, dengan melakukan perbandingan jumlah kata terbanyak.

Gambar 4.4 adalah contoh data yang sudah dilabeli dan akan melalui proses analisis, data tersebut memiliki skor dari hasil labeling data, data yang memiliki skor lebih dari 0 maka akan masuk kedalam kategori sentimen positif, data yang memiliki skor kurang dari 0 maka masuk kedalam kategori sentimen negatif, sedangkan data yang memiliki skor sebesar 0 maka akan masuk kedalam kategori sentimen netral.

	tanggal	waktu	username	steming_data	Score	Sentiment
0	Tue Jun 04	15:29:58	fairywoopie	queen of tears	1	Positif
1	Tue Jun 04	15:27:51	ruru_sgf	lu lihat ngedit pls wkwk queenoftears kimsoohy...	0	Netral
2	Tue Jun 04	15:27:26	stormiismykid	ho kokam sarbat peete masta	0	Netral
3	Tue Jun 04	15:17:38	savageloeoy	queen of tears sedih banget orang nikah	1	Positif
4	Tue Jun 04	15:17:29	lemmelopeuyou	tinggal jejak anabul	0	Netral
5	Tue Jun 04	15:15:45	mirrorballsszz	queen of tears anjgg wkwkwk	2	Positif
6	Tue Jun 04	15:11:23	Beauty_1619	latih mukul kimsoohyun kimjiwon queenoftears	0	Netral
7	Tue Jun 04	15:10:40	hehengdeng	ku mua nya shipper kimsoohyun kimjiwon	0	Netral
8	Tue Jun 04	15:09:59	Beauty_1619	baekhong khawatir kimsoohyun kimjiwon queenoft...	0	Netral
9	Tue Jun 04	15:09:57	awanbulan0_0	saing berat haein teteplah eunsong lihat tanga...	1	Positif

Gambar 4.4 Hasil Labeling Data

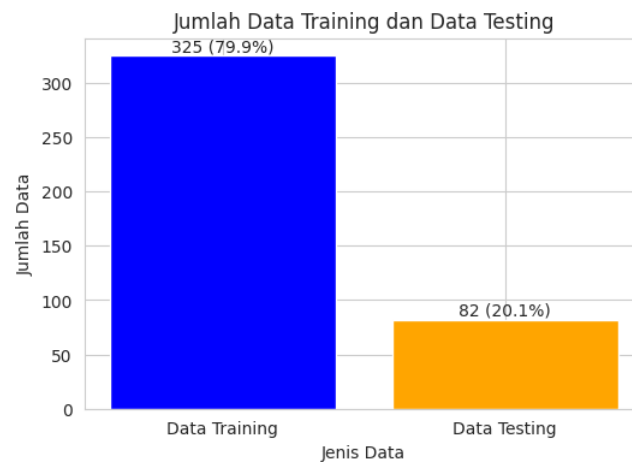
Gambar 4.5 dibawah ini merupakan visualisasi dari pelabelan menggunakan Lexicon Based, terdapat tiga kategori dengan warna yang berbeda, warna biru menunjukkan label positif, orange menunjukkan label netral dan hijau menunjukkan label negatif. Pelabelan ini berisi sebanyak 408 data dan berhasil memperoleh nilai positif sebanyak 187, negatif sebanyak 96 dan netral sebanyak 125.



Gambar 4.5 Diagram hasil labeling

4.6 Split Data Training dan Data Testing

Pada proses ini dataset dibagi menjadi dua bagian yakni data training dan data testing, hal ini bertujuan untuk mempermudah proses klasifikasi, karna pada proses klasifikasi membutuhkan data training dan data testing. Gambar 4.6 adalah diagram hasil dari split data.



Gambar 4.6 Split Data

Fungsi yang digunakan pada proses pemisahan ini adalah fungsi operator split data, setelah dilakukan proses split data tweet yang berisi opini penonton drama Queen Of Tears data terbagi menjadi 79,9% data training atau sebanyak 325 data menjadi data training dan 20,1% atau sebanyak 82 data menjadi data testing.

4.7 Metode Naive Bayes Classifier

Langkah pertama dalam mengklasifikasi data tweet adalah dengan melatih model menggunakan data training. Data training dari dataset yang sudah dipreprocessing, data training ini digunakan sebagai melatih model menggunakan software python. Model algoritma yang sudah dilatih menggunakan data training akan digunakan untuk mengklasifikasi data testing kedalam tiga kategori sentimen positif, negatif dan netral. Pembagian data training dan data testing ini berdasarkan metode 10-fold cross validation. Klasifikasi dengan metode naive bayes menghasilkan probabilitas yang digunakan untuk menentukan apakah tweet masuk kedalam kategori positif,

negatif atau netral. Probabilitas tersebut didapatkan dengan menggunakan persamaan (2.3) dan persamaan (2.4).

Nilai pada suatu probabilitas menunjukkan peluang untuk masuk kedalam suatu kategori sentimen dengan nilai yang ada pada sentimen tersebut, tweet akan masuk kedalam kategori sesuai dengan nilai yang paling besar, sehingga jika sentimen positif lebih besar dari probabilitas tweet maka akan masuk kedalam kategori positif begitu juga sebaliknya. Pengukuran ketepatan klasifikasi dibuktikan dengan adanya evaluasi dari confusion matrix yang diambil berdasarkan hasil dari klasifikasi, dibawah ini adalah confusion matrix dengan kategori aktual dan prediksi pada data training tweet sentimen.

	tanggal	waktu	username	steming_data	Score	Sentiment
0	Tue Jun 04	15:29:58	fairywoopie	queen of tears	1	Positif
1	Tue Jun 04	15:27:51	ruru_sgf	lu lihat ngedit pls wkwk queenoftears kimsoohy...	0	Netral
2	Tue Jun 04	15:27:26	stormiismykid	ho kokam sarbat peete masta	0	Netral
3	Tue Jun 04	15:17:38	savageloey	queen of tears sedih banget orang nikah	1	Positif
4	Tue Jun 04	15:17:29	lemmelopeuyou	tinggal jejak anabul	0	Netral
5	Tue Jun 04	15:15:45	mirrorballsszz	queen of tears anjgg wkwkwk	2	Positif
6	Tue Jun 04	15:11:23	Beauty_1619	latih mukul kimsoohyun kimjiwon queenoftears	0	Netral
7	Tue Jun 04	15:10:40	hehengdeng	ku mua nya shipper kimsoohyun kimjiwon	0	Netral
8	Tue Jun 04	15:09:59	Beauty_1619	baekhong khawatir kimsoohyun kimjiwon queenoft...	0	Netral
9	Tue Jun 04	15:09:57	awanbulan0_0	saing berat haein teteplah eunsong lihat tanga...	1	Positif

Gambar 4.7 Input hasil Labeling data

Score tweet sentimen pada gambar 4.6 menunjukkan bahwa tweet memiliki peluang untuk masuk kedalam kategori sentimen sesuai dengan score yang dihasilkan yakni jika score yang dihasilkan > 0 maka tweet masuk pada kategori sentimen positif, jika score yang dihasilkan 0 maka akan masuk pada kategori sentimen netral, sedangkan tweet yang

menghasilkan score < 0 . Langkah selanjutnya adalah mengukur ketepatan klasifikasi pada data training dan data testing dari setiap sentimen. Pengukuran ini dilakukan menggunakan confusion matrik berdasarkan hasil dari klasifikasi, tabel 4.7 berikut adalah hasil dari confusion matrik pada data training.

Tabel 4.7 Confusion Matrix NB

Kelas Aktual	Kelas Prediksi		
	Positif	Negatif	Netral
Positif	9	0	13
Negatif	5	3	8
Netral	5	2	37

Setelah nilai dari confusion matrix ditemukan langkah selanjutnya adalah melakukan perhitungan ketepatan klasifikasi menggunakan persamaan (3.1) sampai dengan persamaan berikut adalah cara menghitung akurasi.

$$\text{Akurasi} = \frac{3+9+37}{3+5+8+9+13+2+5+37} = \frac{49}{82} = 0,59$$

Hasil akurasi dari metode klasifikasi Naive Bayes yang dievaluasi berdasarkan rumus confusion matrix menunjukkan bahwa akurasi yang didapat dari metode ini adalah sebesar 59%. Sedangkan untuk mencari hasil dari precision, recall dan f-measure diambil dari nilai rata-rata dari jumlah

yang dihasilkan berdasarkan rumus confusion matrix (3.2) sampai (3.4) .

Tabel 4.8 berikut adalah hasil evaluasi dari metode Naive Bayes.

Tabel 4.8 Hasil Evaluasi NB

Label	Judul Kolom Tabel		
	Precision	Recall	F-measure
Positif	64%	84%	73%
Negatif	60%	19%	29%
Netral	47%	41%	44%

Pada Segmen progra, 4.9 dibawah ini terdapa code program untuk menerapkan algoritma Naive Bayes

Segmen Progra 4.9 Naive Bayes

```
import numpy as np

import pandas as pd

from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer

from sklearn.naive_bayes import MultinomialNB,

from sklearn.metrics import accuracy_score, classification_report

from sklearn.preprocessing import LabelEncoder

from sklearn.pipeline import make_pipeline

from sklearn import metrics
```

4.8 Particle Swarm Optimization

Setelah melakukan uji dengan menggunakan metode algoritma Naive Bayes , kemudian akan dilakukan uji optimasi menggunakan metode Particle Swarm Optimization, hal ini bertujuan untuk mengetahui perbandingan akurasi antara hasil dari metode algoritma Naive Bayes dengan hasil algoritma Naive Bayes setelah dioptimasi menggunakan metode Particle Swarm Optimization. Berikut adalah segmen program 4.10 dari PSO.

Segmen Program 4.10 PSO

```
import numpy as np
import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import confusion_matrix, accuracy_score,
classification_report

# Fungsi untuk menghitung akurasi berdasarkan prediksi dan target
sebenarnya

def calculate_accuracy(y_true, y_pred):

    return accuracy_score(y_true, y_pred)

# Fungsi untuk menghitung nilai fungsi objektif (cost) dari model

def calculate_cost(y_true, y_pred):

    return 1 - calculate_accuracy(y_true, y_pred)

# Fungsi untuk menghitung confusion matrix dari model

def calculate_confusion_matrix(y_true, y_pred):

    return confusion_matrix(y_true, y_pred)
```

Pada code program diatas menjelaskna tentang fungsi dari beberapa proses import data yang akan dioptimasi :

- a. Langkah pertama adalah melakukan import library terlebih dahulu, hal ini dilakukan untuk memanggil fungsi-fungsi yang digunakan dalam proses optimasi menggunakan PSO
- b. Fungsi untuk menghitung akurasi berdasarkan prediksi dan target sebenarnya, data yang digunakan adalah data testing dan data training.
- c. Fungsi untuk menghitung nilai fungsi objektif dari model yang digunakan.
- d. Selanjutnya fungsi confusion matrix yang akan melakukan evaluasi pada model untuk menghitung ketepatan dari model.
- e. Fungsi untuk menghitung fitness dari partikel, hal ini dilakukan untuk mengetahui inisialisasi nilai kecepatan partikel.

Segmen program 4.11 berikut ini adalah code dari pemrograman algoritma PSO .

Segmen Program 4.11 Inisialisasi

```
# Algoritma Particle Swarm Optimization
def pso(X_train, X_test, y_train, y_test, n_particles, n_iterations):
    # Inisialisasi partikel dan kecepatan
    dimensions = X_train.shape[1]
    particles_position = np.random.rand(n_particles, dimensions)
    particles_velocity = np.random.rand(n_particles, dimensions)
```

```
# Inialisasi pbest dan gbest
pbest_position = particles_position.copy()
pbest_fitness = np.full(n_particles, np.inf)
gbest_position = None
gbest_fitness = np.inf

# Training
for _ in range(n_iterations):
    for i in range(n_particles):
# Evaluasi partikel
        classifier = RandomForestClassifier()
        classifier.fit(X_train, y_train)
        y_pred = classifier.predict(X_test)
        fitness = fitness_function(y_test, y_pred)
```

Lanjutan Segmen Program 4.11

```
# Update pbest
if fitness < pbest_fitness[i]:
    pbest_fitness[i] = fitness

    pbest_position[i] = particles_position[i].copy()

# Update gbest
if fitness < gbest_fitness:
    gbest_fitness = fitness
    gbest_position = particles_position[i].copy()

# Update kecepatan dan posisi partikel
inertia_weight = 0.7
cognitive_coefficient = 1.5
```

```
social_coefficient = 1.5
```

Penjelasan dari code program diatas sebagai berikut :

- a. Inisialisasi partikel dan kecepatan adalah langkah awal untuk proses PSO, setiap partikel mewakili satu nilai dalam ruang pencarian sedangkan lokasi dan kecepatan setiap partikel ditentukan secara acak.
- b. Inisialisasi *Pbest* adalah penentuan nilai awala terbaik yang pernah ditemukan selama iterasai, nilai *Pbest* diperoleh dari nilai objektif terbaik dari suatu partikel.
- c. Inisialisasi *Gbest* menyimpan nilai awal terbaik yang pernah ditemukan dari seluruh populasi.
- d. Klasifikasi data menggunakan fungsi dari *Randomforestclasifier*, data yang digunakan adalah data training.
- e. Update nilai *Pbest* jika nilai *Pbest* memiliki nilai objektif yang lebh baik dari sebelumnya maka nilai *Pbest* akan diupdate dengan nilai yang saat ini.
- f. Update *Gbest* jika partikel saat ini menemukan nilai yang lebih baik dari sebelumnya maka *Gbest* akan update nilai dengan nilai yang saat ini.
- g. Update Kecepatan dan posisi kondis ini tidak akan berhenti sebelum terpenuhi, maka setiap partikel mengubah kecepatan

dan lokasinya berdasarkan rumus dari P_{best} dan G_{best} dan kecepatan sebelumnya.

- h. Output yang dihasilkan adalah nilai terbaik yang ditemukan dan disimpan dalam G_{best} sehingga dijadikan output dari algoritma.
- i. Setelah memprediksi hasil yang sudah dibandingkan dengan label dari data uji untuk menghitung *confusion matrix*, selanjutnya adalah menampilkan hasil evaluasi dal algoritma *Particle Swarm Optimization*.
- j. Nilai yang dihasilkan dari perhitungan *Confusion Matrix* merupakan hasil akhir dari perhitungan metode

Tabel 4.9 Hasil confusion matrix

Kelas	Kelas Prediksi		
	Positif	Negatif	Netral
Aktual			
Positif	16	0	6
Negatif	9	3	4
Netral	6	1	37

Setelah nilai dari *confusion matrix* ditemukan langkah selanjutnya adalah melakukan perhitungan ketepatan klasifikasi menggunakan persamaan (3.1) sampai dengan persamaan berikut adalah cara menghitung akurasi.

$$\text{Akurasi} = \frac{16+3+37}{16+3+37+6+9+3+4+6+1} = \frac{49}{85} = 0,68 = 68\%$$

Hasil akurasi dihitung dari jumlah TP dibagi dengan TP + FP sehingga menghasilkan nilai akurasi sebesar 68%, hasil ini merupakan hasil dari algoritma *Naive bayes* setelah dilakukan optimasi menggunakan algoritma *Particle Swarm Optimization*. Selanjutnya adalah hasil dari nilai precision, recall dan F-measur yang masing-masing dihitung menggunakan rumus *confusion matrix* (3.2) sampai rumus (3.4).

Tabel 4.1 Hasil dari metode PSO

Label	Precision	Recall	F-Measur
Positif	80%	80%	80%
Negatif	83%	31%	45%
Netral	50%	73%	59%

Dari hasil optimasi metode *Naive Bayes* menggunakan *Particle Swarm optimization* dalam analisis sentimen opini penonton drama *Queen Of Taers* yang diambil dari percakapan di Twitter mendapatkan nilai akurasi sebesar 68%, nilai precision pada sentimen positif sebesar 80%, nilai recall sebesar 80% dan nilai f-measur sebesar 80%, sedangkan untuk sentimen Negatif nilai precision yang dihasilkan sebesar 83%, nilai recall sebesar 31% dan nilai f-measur sebesar 45%, dan yang terakhir untuk sentimen yang netral nilai precision yang dihasilkan sebesar 50%, recall sebesar 73% dan f-measur sebesar 59%.

BAB V

PENUTUP

5.1 Kesimpulan

Algoritma yang digunakan dalam penelitian ini adalah Algoritma *Naïve Bayes Clasifier* (NBC) untuk menganalisis sentimen tweet opini dari penonton, dan menggunakan algoritma optimasi *Particle Swarm Optimization* (PSO). Hasil dari penelitian ini menunjukkan bahwa algoritma PSO dapat mengoptimalkan nilai Accuracy dari hasil klasifikasi algoritma NBC, yang mana nilai dari hasil NBC sebesar 59% lalu setelah dioptimasi menggunakan PSO nilai accuracy meningkat menjadi 68%.

Dari hasil penelitian ini dapat disimpulkan bahwa analisis sentimen yang dihasilkan dalam memetakan opini penonton diklasifikasi menjadi tiga kategori yakni, sentimen positif sebanyak 45,83% atau sebanyak 187 tweet, sentimen negatif sebanyak 23,53% atau sebanyak 96 tweet dan sentimen netral 30,64% atau sebanyak 125 tweet.

5.2 Saran

Dari hasil penelitian ini klasifikasi sentimen dibagi menjadi tiga label sentimen yakni Sentimen positif, sentimen Netral dan sentimen negatif, jumlah data yang digunakan dalam penelitian ini adalah sebanyak 408 data

tweet yang mana setelah melakukan proses analisis data tersebut terkalsifikasi kedalam tiga label sesuai dengan bobot sentimen yang dikandung. Dalam penelitian ini terdapat model algoritma NBC yang mana dalam hasil klasifikasinya menunjukkan hasil akurasi sebesar 60%, lalu angka tersebut melalui proses optimasu menggunakan PSO yang mana hasil accuracy setelah dioptimasi menjadi 68%. Hal ini menunjukkan bahwa dengan menggunakan metode PSO dapat meningkatkan hasil accuracy dari metode NBC



DAFTAR PUSTAKA

- [1] Asrumi, D. Suharijadi, A. Setiari, and Daih Putri Wulanda, *Analisis Sentimen dan Penggalian Opini*, 1st ed. Jawa Tengah: Eureka Media Aksara, 2023.
- [2] Akbar Maulana, “*Twitter Sentimen Analysis wiyh python*,” exsight.
- [3] Cindy mutia annur, “jumlah pengguna twitter indonesia duduki peringkat ke-4 dunia per juli 2023,” databoks.katadata.co.id.
- [4] M. ayu Rizaty, “Wabah K-Pop Makin Mneyebar di Dunia,” katadata.co.id.
- [5] D. Pamugari, “*Review Queen Of Tears*: Drama dengan rating tertinggi sejauh 2024,” CXO media.
- [6] T. T. Widowati, “analisis sentimen twitter terhadap tokoh publik dengan algoritma *Naive bayes* dan *Support Vector Machine*,” *Simetris*, vol. 11, no. 2, Nov. 2020.
- [7] H. Muhamad, C. A. Prasojo, N. A. Sugianto, L. Surtiningsih, and I. Cholissodin, “*Optimasi Naive Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris*,” vol. 4, no. 3, pp. 180–184, 2017.
- [8] R. Aulianita and A. Rifai, “31-40 Mandiri Jakarta; Jl. Damai No. 8 Warung Jati Barat (Margasatwa), Pasar Minggu Jakarta Selatan 12540, (021) 78839513; Mandiri Jakarta,” vol. 3, no. 1, p. 78839513, 2018, [Online]. Available: www.amazon.com
- [9] karimuddin Abdullah, misbahul jannah, and U. Aiman, *metodologi penelitian kuantitatif*, 1st ed. aceh: yayasan penerbit Muhammad Zaini , 2022.
- [10] R. Abubakar, *Metodologi Penelitian*, 1st ed., vol. 1. Yogyakarta: Suka-Pres UIN Sunan Kalijaga, 2021.
- [11] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif dan R&D*, 25th ed., vol. 1. Bandung: Alfabeta, 2017.
- [12] Team, “apa itu crawling data? pengertian, proses dan caranya,” codingstudio.id.
- [13] A. Watratan, E. Utami, and A. Hartanto, “*a. Comparison of Naive Bayes and PSO-Based Naive Bayes Algorithm for prediction of Covid-19 Patient Recovery Data in Indonesia*,” *Jurnal resti*, vol. 7, Apr. 2023.

- [14] T. Putra, E. Utamai, and M. Kurnaiawan, "Analisis Sentimen Pemilu 2024 dengan Naive Bayes Berbasis *Particle Swarm Optimization* (PSO)," *Explore*, vol. 13, Feb. 2023.
- [15] Y. K. Putra and M. Sadali, "Perbandingan Algoritma *Naive Bayes* dan *Naive Bayes* Berbasis PSO untuk Analisis Kredit pada PT. BPR Syariah Paokmotong," *Jurnal Informatika dan Teknologi*, vol. 2, no. 2, pp. 61–69, 2019.
- [16] asrumi, D. suharijadi, A. D. Setiari, and D. putri Wulanda, analisis sentimen dan Penggalian opini, 1st ed. Purbalingga: Eureka Media Aksara, 2023.
- [17] R. D. Arifin, "pengertian Twitter beserta sejarah, fitur, fungsi dan manfaat," dianisa.com.
- [18] Z. B. ram Ghani, "tagar atau hashtag: pengertian, kegunaan dan contoh," duniaku.com.
- [19] Suyanto, *Machine Learning* tingkat dasar dan lanjut, Isbn., vol. 1. bandung: infomatika bandung, 2018.
- [20] A. A. Permana and S. Wahyudin, *Machine Learning*, 1st ed., vol. 1. Padang, Sumatra Utara: PT. Global Eksekutif Teknologi , 2023.
- [21] fajar Ratnawati, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film pada Twitter," *Jurnal Inovtek Polbeng*, vol. 3, no. 1, Jun. 2018.
- [22] jubilee Enterprise, Otodidak Perograman *Python*, 1st ed., vol. 1. jakarta: PT. Elex Media Komputindo, 2017.
- [23] D. Satria, Analisis dan Perancangan Sistem Informasi, 1st ed., vol. 1. Purbalingga, Jawa Tengah: Eureka Media Aksara, 2023.
- [24] ahmad Mufid, "Google colab: pengertian, keuntungan dan cara menggunakan," Rumah Web.
- [25] christopher lee, Belajar Microsoft excel mahrir Step by Step, 1st ed., vol. 1. jakarta: PT. Elex media Komputindo, 2016.
- [26] O. W. Purbo, text mining-analisis medsos, kekuatan bran dan intelijen di internet. yogyakarta: Andi publisher, 2019.

LAMPIRAN-LAMPIRAN

1. Curriculum Vitae



IKFINA AZIZAH
SISTEM INFORMASI

✉ finamannan05@gmail.com
☎ 0881 036 649 477
🏠 Arjasa, Jember, Jawa Timur

Data Pribadi

- Nama : IKFINA AZIZAH
- Tempat, Tanggal Lahir : Jember, 05 Januari 2002
- Alamat : Arjasa - Jember - Jawa Timur
- Jenis Kelamin : Perempuan
- Kewarganegaraan : Indonesia
- Status : Belum Menikah



PENDIDIKAN

- SDN Arjasa 01 Jember
- SMP Ibrahimi 03 Sukorejo
- SMA Ibrahimi 01 Sukorejo
- S1 Sistem Informasi Universitas Ibrahimi Sukorejo

PENGALAMAN

- Anggota IKSASS Jember
- Anggota Desain & Layout LPMI
- Pimred LPMI 2023-2024

KEMAMPUAN

- Komunikasi 
- Manajemen waktu 
- Bahasa Inggris 

HOBI

- Membaca
- Menulis
- Travelling

2. Data Twitter

No	Data Twitter
1	1797675071135457757,"Tue Jun 04 15:29:58 +0000 2024","0","@nawomiha @kdrama_menfess queen of tears gasih","1798014338172874847","","nawomiha","in","","0","1","0","https://x.com/fairywoopie/status/1798014338172874847","1205193906754703362","fairywoopie"
2	1798009300813766713,"Tue Jun 04 15:09:57 +0000 2024","0","Saingan terberat HAEIN teteplah eunsong liat tangannya eunsong dia gak mau buat HYUNWOO cemburu karena udah megang tangan HAEIN #QueenOfTears https://t.co/Q6ZvUtcoFX","1798009300813766713","https://pbs.twimg.com/media/GPPP7ShakAALG-h.jpg","","in","","0","0","0","https://x.com/awanbulan0_0/status/1798009300813766713","1254428092484878346","awanbulan0_0"
3	1798011236778709066,"Tue Jun 04 15:17:38 +0000 2024","0","queen of tears sedih bgttt padahal masalah org yg udah nikah","1798011236778709066","","in","ua'23","0","0","0","https://x.com/savageloey/status/1798011236778709066","1295321899229048834","savageloey"
4	1798009136954933274,"Tue Jun 04 15:09:18 +0000 2024","1","Kayaknya aku lebih cocok romance sejenis lovely runner daripada romance queen of tears","1798009136954933274","","in","Magic ShopðŸ™®","0","1","0","https://x.com/yeontan_juseyo/status/1798009136954933274","541333667","yeontan_juseyo"
5	1798007691903910096,"Tue Jun 04 15:03:33 +0000 2024","1","Queen of tears reunion Satu moment bareng gini aja bahagia bangett aku #QueenOfTears https://t.co/ZkbPsVUD0L","1798007691903910096","https://pbs.twimg.com/media/GPPOdwJb0AATEez.jpg","","in","","0","0","0","https://x.com/Jonquilip/status/1798007691903910096","865843292","Jonquilip"
6	1797996757428973603,"Tue Jun 04 14:20:06 +0000 2024","1","Dari dulu ngedrakor baru kali ini deh ni drama #QueenOfTears bikin se gamon itu maunya tetep lanjut aje Btw itu yg baju itemÂ² juga semoga lanjut ampe pelaminan yee hehe #KimSooHyun #KimJiwon https://t.co/pG7RGI7cju","1797996757428973603","https://pbs.twimg.com/media/GPPEg4AaIAAX01y.jpg","","in","Indonesia","0","0","1","https://x.com/Leuxninie/status/1797996757428973603","1150266629101477888","Leuxninie"