



**ANALISIS EMOSI MENGGUNAKAN METODE SUPPORT VECTOR
MACHINE (STUDI KASUS : TEKS KOMENTAR VIDEO REWIND
INDONESIA 2021 DI YOUTUBE)**

SKRIPSI

Oleh :

Bagas Tetuko

NIM 182410102021

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS JEMBER
2023**



**ANALISIS EMOSI MENGGUNAKAN METODE SUPPORT VECTOR
MACHINE (STUDI KASUS : TEKS KOMENTAR VIDEO REWIND
INDONESIA 2021 DI YOUTUBE)**

SKRIPSI

Diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat untuk menyelesaikan Pendidikan Sarjana (S1) Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Jember dan mencapai gelar Sarjana Komputer

Oleh :

Bagas Tetuko

NIM 182410102021

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS JEMBER**

2023

PERSEMBAHAN

Skripsi ini saya persembahkan kepada:

1. Allah SWT yang selalu memberi rahmat dan hidayah-Nya. Mempermudah segala urusan dan memberi kelancaran dalam proses pengerjaan Skripsi.
2. Kedua orang tua yaitu Ayah Rudi Cahyono dan juga Ibu Lilik Andayani tercinta. Terima kasih karena berkat doa Ayah dan Ibu, pengerjaan Skripsi ini dapat terlaksana dengan baik.
3. Ibu Oktalia Juwita S.Kom., M.MT selaku Dosen Pembimbing Utama dan Bapak Priza Pandunata S.Kom., M.Sc selaku Dosen Pembimbing Pendamping yang telah sabar membimbing, memberikan semangat dan juga dukungan dalam proses penyelesaian Skripsi.
4. Almamater Program Studi Teknologi Informasi Fakultas Ilmu Komputer Universitas Jember.

MOTTO

“Tiada hidup yang dijalani tanpa adanya masalah, dan setiap masalah selalu ada solusi, karena hidup ini selalu ada sebab dan akibat”

-BT-

PENGESAHAN PEMBIMBING

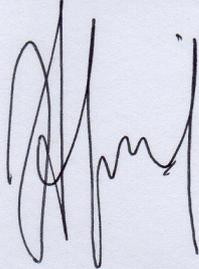
Skripsi yang berjudul “Analisis Emosi Menggunakan Metode Support Vector Machine (Studi Kasus : Teks Komentar Video Rewind Indonesia 2021 Di Youtube)”, telah diuji dan disahkan pada :

hari, tanggal : Rabu, 11 Januari 2023

tempat : Fakultas Ilmu Komputer Universitas Jember

Disetujui oleh:

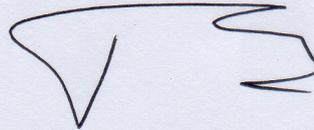
Pembimbing I



Oktalia Juwita S.Kom., M.MT

NIP. 198110202014042001

Pembimbing II



Priza Pandunata S.Kom., M.Sc

NIP. 198301312015041001

PENGESAHAN PENGUJI

Skripsi yang berjudul “Analisis Emosi Menggunakan Metode Support Vector Machine (Studi Kasus : Teks Komentar Video Rewind Indonesia 2021 Di Youtube)”, telah diuji dan disahkan pada:

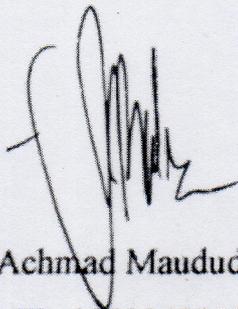
hari, tanggal : Rabu, 11 Januari 2023

tempat : Fakultas Ilmu Komputer Universitas Jember

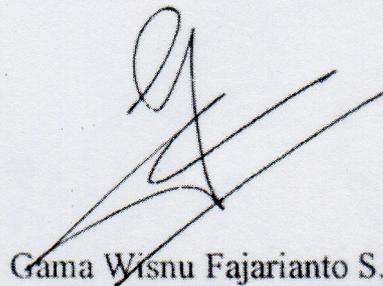
Tim Penguji,

Penguji I,

Penguji II



Achmad Maududie ST, M.Sc
NIP. 197004221995121001



Gama Wisnu Fajarianto S.Kom., M.Kom
NRP. 760015717

Mengesahkan

Dekan Fakultas Ilmu Komputer,



Drs. Antonius Cahya P, M. App., Sc., Ph.D

NIP. 196909281993021001

PERNYATAAN

Saya yang bertandatangan dibawah ini :

Nama : Bagas Tetuko

NIM : 182410102021

Menyatakan dengan sesungguhnya bahwa karya ilmiah yang berjudul “Analisis Emosi Menggunakan Metode Support Vector Machine (Studi Kasus : Teks Komentar Video Rewind Indonesia 2021 Di YouTube)”, adalah benar-benar hasil karya sendiri, kecuali jika dalam pengutipan substansi disebutkan sumbernya, belum pernah diajukan pada institusi mana pun, dan bukan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenarnya, tanpa adanya tekanan dan paksaan dari pihak manapun serta bersedia mendapat sanksi akademik jika di kemudian hari pernyataan ini tidak benar.

Jember, 03 September 2022

Yang menyatakan,



Bagas Tetuko

NIM 182410102021

RINGKASAN

Analisis Emosi Menggunakan Metode Support Vector Machine (Studi Kasus : Teks Komentar Video Rewind Indonesia 2021 Di YouTube). Oleh Bagas Tetuko dengan NIM 182410102021 yang terdiri dari 75 halaman dari Fakultas Ilmu Komputer Universitas Jember.

Youtube rewind Indonesia, diluncurkan untuk merayakan epidemik yang terdapat didalam media sosial yang viral selama satu tahun, adalah sebuah acara kolaborasi untuk pembuat konten yang lagi hangat maupun viral dalam setahun. Video rewind Indonesia 2021 dapat disaksikan melalui channel Indonesian Youtubers. Sebelumnya video youtube rewind Indonesia melakukan unggahan pertama kali dalam channel milik Eka Gustiwana. Youtube rewind Indonesia tidak melakukan berkolaborasi yakni inisiatif sendiri yakni pengguna media sosial atau pembuat konten dimana video tersebut menjadi viral maupun booming .

Penelitian ini bertujuan melakukan analisis emosi pada komentar youtube Rewind Indonesia 2021 untuk mengetahui emosi komentar pro dan kontra dari viewers sebagai bahan evaluasi content creator video indonesia rewind kedepannya. Penelitian analisis emosi memanfaatkan salah satu metode dalam text mining, yakni metode support vector machine. Penelitian ini diawali dengan melakukan scraping data komentar pada Youtube Rewind Indonesia 2021. Kemudian dilanjutkan dengan melakukan pelabelan data, melakukan tahap preprocessing, pembobotan kata, penyusunan model svm, dan melakukan evaluasi hasil.

Hasil dari penelitian ini yaitu Berdasarkan hasil analisis emosi didapatkan bahwa uji analisis emosi dari komentar yang mengandung emosi bahagia didapatkan sebanyak 85 komentar, komentar yang mengandung emosi marah didapatkan sebanyak 5 komentar, komentar yang mengandung emosi sedih didapatkan sebanyak 7 komentar, komentar yang mengandung emosi jijik didapatkan sebanyak 1 komentar, komentar yang mengandung emosi terkejut didapatkan sebanyak 2 komentar dan komentar yang mengandung emosi takut didapatkan sebanyak 0 komentar. Sehingga dari hasil uji analisis emosi tersebut rata - rata komentar yang dituliskan mengandung emosi bahagia.

DAFTAR ISI

SKRIPSI.....	i
PENGESAHAN PEMBIMBING.....	iv
PENGESAHAN PENGUJI.....	v
PERNYATAAN.....	vi
RINGKASAN	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	4
BAB 2 TINJAUAN PUSTAKA.....	5
2.1 Penelitian Terdahulu.....	5
2.2 Analisis Emosi.....	7
2.3 Support Vector Machine	7
2.4 Scraping Data	8
2.5 Text Mining	9
2.5.1 Convert Emoticon	9
2.5.2 Case Folding	9
2.5.3 Cleansing.....	9
2.5.4 Stemming	10
2.5.5 Tokenisasi	9
2.5.6 Stopword Removal.....	10
2.5.7 Algoritma Term Frequency Inverse Document Frequency (TF-IDF)	
10	
2.6 Confusion Matrix	11

2.6.1	Precision.....	12
2.6.2	Recall.....	12
2.6.3	Akurasi	12
2.7	Pengertian Emosi.....	13
BAB 3	METODOLOGI PENELITIAN	14
3.1	Jenis Penelitian	14
3.2	Objek Penelitian	14
3.3	Tempat dan Waktu Penelitian	14
3.4	Tahapan Penelitian	15
3.4.1	Scraping Data Komentar Youtube	15
3.4.2	Pelabelan Data.....	16
3.4.3	Preprocessing	16
3.5	Pembobotan Kata TF-IDF	18
3.6	Penyusunan Model SVM.....	18
3.7	<i>Confusion Matrix</i>	18
3.8	Hasil Analisis Emosi	19
BAB 4	HASIL DAN PEMBAHASAN	19
4.1	Hasil Dataset.....	19
4.2	Hasil Pembuatan Kamus Emoticon.....	21
4.3	Hasil Implementasi Sistem	22
4.3.1	Implementasi Scraping data	22
4.3.2	Implementasi preprocessing.....	23
4.3.3	Implementasi TF IDF	26
4.3.4	Implementasi penyusunan model.....	27
4.3.5	Implementasi confusion matrix.....	27
4.3.6	Implementasi hasil analisis emosi	28
4.4	Hasil Preprocessing	29
4.4.1	Hasil Convert emoticon.....	29
4.4.2	Hasil Case Folding	30
4.4.3	Hasil Cleansing	31
4.4.4	Hasil stemming	34

4.4.5	Hasil tokenisasi	32
4.4.6	Hasil Stopword Removal	35
4.5	Hasil TF IDF	36
4.6	Hasil Confusion Matrix	46
4.7	Hasil Uji Analisis Emosi	47
BAB 5	KESIMPULAN DAN SARAN	58
5.1	Kesimpulan.....	58
5.2	Saran.....	58
BAB 6	DAFTAR PUSTAKA	59
BAB 7	DAFTAR LAMPIRAN.....	61

DAFTAR TABEL

Tabel 1 Confusion Matrix	12
Tabel 2 Contoh Dataset.....	20
Tabel 3 Contoh Hasil Convert emoticon.....	29
Tabel 4 Contoh Hasil Case Folding	30
Tabel 5 Contoh hasil Cleansing	31
Tabel 6 Contoh Hasil Stemming	34
Tabel 7 Contoh Hasil Tokenisasi	32
Tabel 8 Contoh Hasil Remvoval	35
Tabel 9 Contoh Hasil Nilai TF Bahagia.....	36
Tabel 10 Contoh Hasil Nilai TF Jijik.....	36
Tabel 11 Contoh Hasil Nilai TF Marah	37
Tabel 12 Contoh Hasil Nilai TF Sedih.....	37
Tabel 13 Contoh Hasil Nilai TF Takut	37
Tabel 14 Contoh Hasil Nilai TF Terkejut	38
Tabel 15 Contoh Hasil Nilai IDF	39
Tabel 16 Contoh Hasil Nilai TF-IDF Bahagia.....	42
Tabel 17 Contoh Hasil Nilai TF-IDF Jijik	43
Tabel 18 Contoh Hasil Nilai TF-IDF Marah.....	43
Tabel 19 Contoh Hasil Nilai TF-IDF Sedih.....	44
Tabel 20 Contoh Hasil Nilai TF-IDF Takut.....	44
Tabel 21 Contoh Hasil Nilai TF-IDF Terkejut	45
Tabel 22 Hasil Uji Analisis Emosi.....	47
Tabel 23 Analisis Emosi Komentar yang tidak sesuai.....	56

DAFTAR GAMBAR

Gambar 1 Contoh Model SVM.....	8
Gambar 2 Tabel Kernel SVM	8
Gambar 3 Tahapan Penelitian	15
Gambar 4 Split Data Tes 90:10.....	20
Gambar 5 Split Data Tes 80:20.....	20
Gambar 6 Split Data Tes 70:30.....	20
Gambar 7 Split Data Tes 60:40.....	20
Gambar 8 Contoh Kamus Emoticon	21
Gambar 9 Kode Program Scraping	23
Gambar 10 Kode Program Convert Emoticon	23
Gambar 11 Kode Program Case Folding	24
Gambar 12 Kode Program Cleansing	24
Gambar 13 Kode Program Tokenisasi	25
Gambar 14 Kode Program Normalization	25
Gambar 15 Kode Program Stemming	26
Gambar 16 Kode Program Stopword.....	26
Gambar 17 Kode Program TF-IDF.....	27
Gambar 18 Kode Program Penyusunan Model.....	27
Gambar 19 Kode Program Implementasi Confusion Matrix	28
Gambar 20 Kode Program Hasil Analisis Emosi.....	28
Gambar 21 implementasi visualisasi data	28
Gambar 22 Nilai akurasi Presisi dan Recal 80:20%	46
Gambar 23 Hasil Confusion Matrix 80:20%	47
Gambar 24 Hasil Total Analisis Emosi.....	55
Gambar 25 Wordcloud Kata Emosi Bahagia	58
Gambar 26 5 Kata Terbanyak pada Wordcloud.....	59

BAB 1 PENDAHULUAN

Bab ini merupakan langkah awal dari penulisan tugas akhir ini yang memuat uraian latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, dan batasan masalah.

1.1 Latar Belakang

Serial video Indonesia yang merangkum peristiwa-peristiwa yang menjadi topik hangat maupun viral di sosial media termasuk youtube disebut youtube Rewind Indonesia. Youtube rewind Indonesia, diluncurkan untuk merayakan epidemik yang terdapat didalam media sosial yang viral selama satu tahun, adalah sebuah acara kolaborasi untuk pembuat konten yang lagi hangat maupun viral dalam setahun. Video rewind Indonesia 2021 dapat disaksikan melalui channel Indonesian Youtubers. Sebelumnya video youtube rewind Indonesia melakukan unggahan pertama kali dalam channel milik Eka Gustiwana. Youtube rewind Indonesia tidak melakukan berkolaborasi yakni inisiatif sendiri yakni pengguna media sosial atau pembuat konten dimana video tersebut menjadi viral maupun booming (*Youtube Rewind Indonesia, 2022*).

Youtube adalah situs web berbasis video yang diluncurkan pada tahun 2005. Youtube menempati tempat pertama pada media sosial waktu ini. Youtube berkembang pesat, dengan pernyataan misi youtube adalah memberikan suara kepada semua orang dan menunjukkan dunia kepada mereka. Melalui youtube, orang dapat mengekspresikan suara mereka atau menunjukkan kreativitas mereka dengan mudah. Orang-orang dapat terlibat secara tidak langsung di youtube dengan melihat video, memberikan komentar, dan membagikan video tersebut kepada orang lain (*Viny Christanti et al., 2020*).

Komentar merupakan tempat mengekspresikan suara bagi pengguna dalam menikmati sebuah video. Komentar yang ditujukan terhadap sebuah video bisa berupa komentar yang menenujukan kesukaan, kesedihan, ketakutan, dan lain lain (*Muktafin et al., 2020*). Maka dari itu pengapresian emosi melalui

komentar yang dilakukan oleh pengguna perlu adanya tindakan lebih lanjut yaitu berupa analisis emosi dalam rangka memecahkan bagaimana emosi yang digunakan di dalam komentar tersebut. Alasan melaksanakan sebuah analisis emosi dikarenakan peneliti dapat menganalisis lebih dalam tentang respon orang-orang yang melakukan komentar karena dapat mengetahui 6 emosi dasar yaitu bahagia, marah, takut, sedih, jijik, dan terkejut (Rohman et al., 2021), juga dapat mempelajari perubahan suasana hati orang-orang atau mengetahui emosi dalam merespon sesuatu melalui kalimat komentar. Analisis emosi ini menggunakan sebuah metode yakni support vector machine.

Support vector machine pada tahun 1992 diluncurkan oleh Vapnik. Support vector machine sebuah algoritma machine learning klasifikasi yang menggunakan structural risk minimization yang bertujuan untuk menemukan hyperlane dengan dua class pada input space yang dipisahkan (Irawan et al., 2021). Pengertian support vector machine adalah sebuah sistem pembelajaran dengan fungsi yang linier didalam ruang hipotesa dengan dimensi tinggi yang diturunkan dalam teori belajar statistik menggunakan algoritma pembelajaran optimasi (Nello & Shawe, n.d.).

Pada video rewind Indonesia 2021 banyak netizen yang membandingkan video rewind Indonesia 2021 dengan video rewind sebelumnya. Hal tersebut karena banyak komentar dari netizen yang menyebutkan video Indonesia rewind 2021 banyak berisi moment viral dari platform TikTok daripada moment dari youtube sendiri. Selain itu kemunculan video rewind Indonesia 2021 ini mendapat berbagai respond pro dan kontra dari netizen. Hal tersebut karena video Rewind Indonesia 2021 mendapatkan banyak komentar maupun kritikan dari netizen namun juga tidak sedikit netizen yang memberikan apresiasi karena video tersebut masuk trending 1 di media sosial youtube dan banyak juga apresiasi dari sederet selebriti dan youtuber pada video rewind 2021 ini (Terkenal.co.id, 2022).

Sehingga dari permasalahan diatas peneliti ingin melakukan analisis emosi pada komentar youtube Rewind Indonesia 2021 untuk mengetahui emosi komentar pro dan kontra dari viewers sebagai bahan evaluasi content

creator video indonesia rewind kedepannya. Penelitian analisis emosi memanfaatkan salah satu metode dalam text mining, yakni metode support vector machine.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang diatas, maka rumusan masalah dari penelitian ini adalah sebagai berikut:

1. Berapa akurasi hasil dari implementasi metode support vector machine dalam menganalisis emosi.
2. Bagaimana hasil dari analisis emosi menggunakan metode support vector machine?

1.3 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang telah disampaikan, maka tujuan dari penelitian ini sebagai berikut:

1. Mengetahui hasil akurasi implementasi metode support vector machine terhadap analisis emosi teks komentar video rewind Indonesia 2021.
2. Mengetahui hasil emosi dasar manusia terhadap analisis emosi teks komentar video rewind Indonesia 2021 menggunakan metode support vector machine.

1.4 Manfaat Penelitian

Berdasarkan tujuan penelitian diatas, maka manfaat penelitian ini adalah sebagai berikut:

1. Bagi peneliti sebagai bahan penyusunan tugas akhir dan pengembangan keilmuan di Universitas Jember serta memberikan wawasan dan pengembangan keilmuan bagi peneliti.
2. Bagi pengguna youtube untuk menilai sebuah konten video berdasarkan penyampaian emosi melalui sebuah komentar.
3. Bagi konten creator untuk mengetahui reaksi dari viewers berupa komentar yang telah diketahui emosinya agar kedepannya unggahan video lebih memenuhi harapan dari masyarakat dan tidak hanya berupa like dan dislike dari video tersebut.

1.5 Batasan Masalah

Berdasarkan uraian latar belakang diatas, maka batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Penelitian hanya menggunakan komentar bahasa Indonesia.
2. Emosi dikategorikan menjadi 6 emosi dasar yaitu marah, bahagia, sedih, terkejut, jijik dan takut.
3. Sumber data yang dimanfaatkan dalam penelitian ini adalah data komentar pada unggahan akun YouTube Indonesian Youtuber dengan video yang berjudul Rewind Indonesia 2021 pada tanggal 30 Desember 2021 – 31 Oktober 2022

BAB 2 TINJAUAN PUSTAKA

Bab ini menjelaskan tentang tinjauan yang berkaitan dengan penelitian terdahulu serta kajian teori yang berkaitan dengan penelitian yang dilakukan.

2.1 Penelitian Terdahulu

Dwi Ardiada, Made Sudarma dan Dwi Giriantari pada tahun 2019 melaksanakan riset yang berjudul “Text Mining pada Sosial Media untuk Mendeteksi Emosi Pengguna Menggunakan Metode Support Vector Machine dan K-Nearest Neighbour”. Beberapa penelitian mengenai deteksi emosi telah dilakukan contohnya pada penelitian yang dilakukan Chaitail G. Patil dan Sandip S. Patil menyebutkan penggunaan metode Support Vector Machine dan dataset ISEAR memiliki akurasi tertinggi yaitu 71.64% sedangkan Metode Naive Bayes Classifier akurasinya 60.8%. Penelitian ini dipublikasikan pada majalah ilmiah teknologi elektro yang terindeks sinta 3(SINTA, n.d.-a). Pada penelitian ini melakukan penelitian text mining pada Sosial Media untuk mendeteksi emosi pengguna. Riset ini menganalisis dengan menunjukkan hasil dari klasifikasi yang menunjukkan emosi menggunakan sebuah metode support vector machine dan metode k-nearest neighbor. Hasil yang diperoleh dari metode support vector machine meningkat lebih tinggi nilai accuracy, recall, presisi dari metode k-nearest neighbor. Hal ini terlihat dari perhitungan waktu dalam mengklasifikasi emosi.

Ainurrochman, Derry Pramono Adi, dan Agustinus Bimo Gumelar pada tahun 2020 melaksanakan riset yang berjudul “Deteksi Emosi Wicara On-Demand menggunakan SVM dan LSTM”. Penelitian ini dipublikasikan pada jurnal resti yang terindeks sinta 2 (SINTA, n.d.-b). Riset ini menganalisis dengan menggunakan dataset HENLO untuk menunjukkan proses identifikasi emosi wicara manusia. Dataset HENLO menggunakan MOD yang memungkinkan data data film yang diakusisi dengan jumlah yang banyak yang sangat berpengaruh pada klip suara. Dalam menggunakan metode SVM dan LSTM memberikan hasil yang memuaskan dalam mendeteksi sebuah emosi. Dengan menggunakan kedua metode tersebut pada penelitian ini hasil yang diperoleh melalui metode SVM dan LSTM memberikan hasil yang cukup

dalam pendeteksian emosi. SVM yang diberi trained model dalam LSTM ternyata memberi nilai akurasi yang lebih tinggi. Hasil training pertama pada SVM dan LSTM memberi angka 30%, sedangkan hasil training kelima pada SVM dan LSTM memberi angka 47,4%. Peningkatan dalam angka akurasi ini berarti performa SVM dan LSTM cukup besar serta berpengaruh klasifikasi emosi multi kelas

Arif Nur Rohman, Riska Dwi Handayani, Ryan Dwi Y.P, dan Kusriani pada tahun 2019 melaksanakan riset yang berjudul "Deteksi Emosi Media Sosial Menggunakan Term Frequency-Inverse Document Frequency". Penelitian ini dipublikasikan pada CSRID yang terindeks sinta 3 (SINTA, n.d.-c). Riset ini menganalisis dengan menggunakan data yang diambil melalui media social facebook yakni update status pengguna. Pengambilan data menggunakan metode scrapping dengan melakukan pra-pemrosesan dengan membuang simbol dan ikon yang tidak diperlukan. Selain itu dilakukan juga normalisasi teks dikarenakan Bahasa yang tidak baku yang digunakan oleh pengguna, dan dengan stopword membuang kata yang tidak diperlukan. Algoritma TF-IDF dipakai untuk menilai sebuah bobot dalam kata agar memudahkan dalam pelabelan emosi. Dengan menggunakan TF-IDF meningkatkan hasil akurasi daripada pelabelan emosi manual. Pelabelan emosi yang dilakukan secara manual diperoleh akurasi sebesar 59%, sedangkan yang sudah dilakukan TF-IDF diperoleh akurasi sebesar 87,23%.

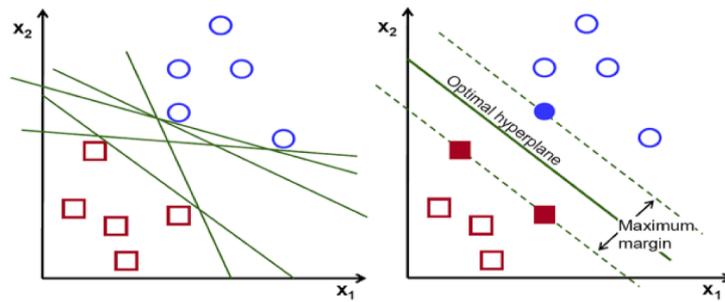
Hasil penelitian dari penelitian terdahulu diatas dapat digunakan sebagai bahan pertimbangan peneliti dalam melakukan analisis emosi. Adapun kontribusi penelitian yang digunakan yaitu metode yang digunakan support vector machine, pemanfaatan pembobotan kata untuk algoritma term frequency-inverse document frequency (TF-IDF), jenis kategori emosi yaitu 6 emosi dasar manusia yaitu bahagia, takut, marah, sedih, jijik dan kaget, teknik preprocessing dalam teks, penentuan tingkat akurasi dengan penggunaan confusion matrix.

2.2 Analisis Emosi

Analisis emosi yaitu mengumpulkan data tentang bagaimana seseorang berkomunikasi secara verbal dan non-verbal untuk memahami suasana hati dan sikap seseorang. Dalam teknologi, juga dikenal sebagai emosi sentimen, yang memberikan wawasan tentang bagaimana pelanggan memandang suatu produk, presentasi produk, atau interaksi. Teknologi analisis sentimen yang umum digunakan adalah analisis teks, yang mengkaji pilihan kata untuk menentukan emosi tentang suatu produk. Data emosional digunakan untuk mengembangkan strategi untuk meningkatkan manajemen. Pengenalan emosi saat ini biasanya mengklasifikasikan emosi sebagai marah, jijik, takut, bahagia, sedih, dan terkejut (Rohman et al., 2021).

2.3 Support Vector Machine

Support vector machine adalah sebuah algoritma machine learning klasifikasi yang memakai structural risk minimization yang bertujuan untuk menemukan hyperlane dengan dua class pada input space yang dipisahkan (Irawan et al., 2021). Pengertian support vector machine ialah sebuah sistem pembelajaran dengan fungsi linier didalam ruang hipotesa dengan dimensi tinggi yang diturunkan dalam teori belajar statistik menggunakan algoritma pembelajaran optimasi (Nello & Shawe, n.d.). Model Algoritma learning tinggi dalam hal analisis data untuk melakukan klasifikasi dan regresi data (Ainurrochman et al., 2020). SVM merupakan algoritma yang dikembangkan oleh Boser, Guyon dan Vapnik dimana algoritma ini untuk melakukan klasifikasi untuk memisahkan data berbeda melalui garis optimal yaitu hyperplane. Hyperplane merupakan pemisah antara kelas berdasarkan jarak margin. Dimana margin inilah merupakan jarak terdekat antar pola terdekat dengan garis optimal hyperplane untuk setiap kelas, dan pola terdekat inilah disebut sebagai support vector (Widayani & Harliana, 2021).



Gambar 1 Contoh Model SVM

Persamaan yang digunakan dalam SVM dapat dilihat pada persamaan 1 dan 2.

$$[(w^t \times x_i) + b] \geq 1 \text{ untuk } y_i = +1 \dots\dots\dots (1)$$

$$[(w^t \times x_i) + b] \leq -1 \text{ untuk } y_i = -1 \dots\dots\dots (2)$$

Ketika data tersebut bersifat liner maka persamaan 1 dan 2 dapat dioptimalkan dengan fungsi lagrange multiplier, sedangkan jika data tersebut bersifat non-linier maka dapat dioptimalkan dengan memakai fungsi kernel (Widayani & Harliana, 2021). Fungsi kernel dapat dilihat pada gambar 2.

Type of SVM	Mercer Kernel
Gaussian or Radial Basis Function (RBF)	$K(x_1, x_2) = \exp\left(-\frac{\ x_1 - x_2\ ^2}{2\sigma^2}\right)$
Linear	$K(x_1, x_2) = x_1^T x_2$
Polynomial	$K(x_1, x_2) = (x_1^T x_2 + 1)^p$
Sigmoid	$K(x_1, x_2) = \tanh(\beta_0 x_1^T x_2 + \beta_1)$

Gambar 2 Tabel Kernel SVM

2.4 Scraping Data

Scraping data web adalah sebuah metode untuk mengambil beberapa informasi dari situs web dengan mengekstraksi data dan menyimpan dalam sebuah database. Data yang didapatkan pada website menggunakan format HTML dengan data yang tak terstruktur. Lalu data yang tidak terstruktur tersebut disimpan didalam sebuah database maupun file dengan format yang sudah terstruktur (Viny Christanti et al., 2020).

2.5 Text Mining

Text Mining ialah suatu proses pencarian data teks yang sumber data biasanya diperoleh dari dokumen, dan dapat menganalisis keterhubungan antar dokumen dengan tujuan mencari kata atau kalimat yang dapat mewakili isi dari dokumen tersebut. Text mining adalah pengaplikasian dari konsep dan teknik dalam data mining agar menemukan pola dalam sebuah teks, atau cara penguraian teks untuk mengekstrak informasi yang berguna dalam tujuan tertentu. Text mining bisa didefinisikan dalam upaya penggalian suatu informasi dan berinteraksi pada sekumpulan dokumen (Ardiada et al., 2019). Dengan menggunakan komponen analisis dalam text mining untuk melakukan preprocessing teks adalah sebagai berikut:

2.5.1 Convert Emoticon

Convert emoticon adalah sebuah proses yang digunakan dalam melakukan konversi bentuk emoticon kedalam bentuk string yang sesuai (Alita et al., 2020). Convert emoticon digunakan untuk membantu keakuratan emosi dalam sebuah teks.

2.5.2 Case Folding

Case Folding adalah sebuah proses untuk mengubah atau menyeragamkan semua huruf besar didalam dokumen menjadi karakter dengan huruf kecil(Rustiana & Rahayu, 2017). Penggunaan huruf besar atau kapital dan huruf kecil bisa memiliki makna yang berbeda.

2.5.3 Cleansing

Proses dalam membersihkan data. Tujuan dari cleansing supaya data bersih dari noise. Penghapusan karakter yang tidak perlu contohnya : symbol, tanda baca, dan mention(@)(Rustiana & Rahayu, 2017).

2.5.4 Tokenisasi

Tokenisasi adalah pemotongan sebuah dokumen menjadi bagian-bagian yang disebut token(Indraloka & Santosa, 2017). N-gram adalah token yang merupakan urutan kata berurutan dengan

panjang n . N-gram menyimpan informasi tentang kemunculan kata, karena mengelompokkan kata-kata yang berdekatan ke dalam token yang sama (Anandarajan et al., 2019). Misalnya unigram ialah token yang terdiri dari satu kata, bigram ialah token yang terdiri dari dua kata yang berdekatan.

2.5.5 Normalisasi

Tahap Normalisasi adalah tahap yang merubah kalimat tidak baku menjadi kalimat baku. Tahap ini menggunakan sebuah kamus yang didapat dari github *Colloquial Indonesia Lexicon*. Konsep dari tahap ini ialah apabila pada data komentar youtube terdapat didalam kamus maka kata tersebut akan berubah ke bentuk bakunya (Zuhad & Wilantika, 2022).

2.5.6 Stemming

Stemming adalah sebuah proses untuk menemukan kata dasar dari sebuah kata dengan menghilangkan kata imbuhan baik dari yang terdiri dari awalan, sisipan, akhiran (Indraloka & Santosa, 2017). Proses stemming menggunakan library sastrawi.

2.5.7 Stopword Removal

Stopword removal adalah sebuah proses penghilangan kata kata yang tidak berkontribusi banyak pada isi dokumen (Indraloka & Santosa, 2017). Proses stopwords removal dibantu dengan menggunakan kamus yang dibuat oleh peneliti serta menggunakan library sastrawi.

2.5.8 Algoritma Term Frequency Inverse Document Frequency (TF-IDF)

Algoritma term frequency inverse document frequency ialah sebuah metode untuk menentukan bobot setiap kata untuk menentukan seberapa jauh keterhubungan kata terhadap dokumen (Deolika et al., 2019). TF-IDF melakukan pembobotan kata terhadap sebuah dokumen dengan menggunakan nilai TF dan IDF yang sudah

dikalikan. TF dapat dihitung dengan banyaknya suatu kata yang muncul pada 1 dokumen. DF dapat dihitung dengan jumlah dokumen dimana kata tersebut muncul (Sya'bani & Umilasari, 2018). Rumus untuk IDF dapat dilihat pada persamaan 3:

$$IDF = \log \left(\frac{D}{DF} \right) \dots\dots\dots (3)$$

Rumus untuk TF-IDF dapat dilihat pada persamaan 4:

$$W = TF \times IDF \dots\dots\dots (4)$$

Keterangan :

IDF = Inverse document frequency

D = Jumlah dokumen

DF = Frekuensi dokumen

W = Bobot

TF = Frekuensi kata

2.6 Confusion Matrix

Confusion matrix ialah metode untuk mengukur akurasi, *recall*, dan *precesion*. Sederhananya, *confusion matrix* digunakan untuk mengukur bagaimana hasil performa dari sebuah metode klasifikasi. Pada confusion matrix memiliki 4 istilah dalam representasi klasifikasi yakni false positif, true negatif, true positif, dan false negative (Arini et al., 2020). Tabel confusion matrix dapat dilihat pada tabel 1:

Tabel 1 Confusion Matrix

		True Value	
		True	False
Prediction	True	TP Correct Result	FP Unexpected Result
	False	FN Missing Result	TN Correct Absence of Result

2.6.1 Precision

Presisi adalah rasio nilai benar yang diprediksi positif dengan total seluruh nilai prediksi positif (Arini et al., 2020). Rumus presisi adalah dapat dilihat pada persamaan 5:

$$Precision = \left(\frac{TP}{TP + FP} \right) * 100\% \dots\dots\dots (5)$$

2.6.2 Recall

Recall bekerja dengan cara mengukur seberapa mampu sebuah system dalam menemukan data yang relevan dalam sebuah dokumen (Arini et al., 2020). Rumus recall dapat dilihat pada persamaan 6:

$$Recall = \left(\frac{TP}{TP + FN} \right) * 100\% \dots\dots\dots (6)$$

2.6.3 Akurasi

Akurasi untuk mengukur seberapa akurat sebuah kasus dimana yang prediksinya benar dengan jumlah keseluruhan data (Arini et al., 2020). Rumus akurasi dapat dilihat pada persamaan 7:

$$Akurasi = \left(\frac{TP + TN}{TP + TN + FP + FN} \right) * 100\% \dots\dots\dots (7)$$

2.7 Pengertian Emosi

Emosi dijelaskan sebagai keadaan sebuah perasaan yang mempengaruhi perilaku, dan pikiran. Emosi dapat menular pada seseorang yang memperhatikan. Berikut adalah gambaran dari 6 emosi dasar. Penentuan emosi marah digambarkan pada isi sebuah komentar yang memiliki makna menentang, merasa dirugikan oleh sebuah pernyataan. Penentuan emosi sedih digambarkan pada isi komentar yang memiliki makna iba, penyesalan atau kegagalan oleh pernyataan tersebut. Penentuan emosi jijik digambarkan pada komentar yang memiliki makna kotor, tidak nyaman atau risih dalam sebuah pernyataan. Penentuan emosi takut digambarkan pada isi komentar yang memiliki makna kekhawatiran menghadapi sebuah pernyataan. Penentuan emosi bahagia digambarkan pada isi komentar yang memiliki makna senang, tenang terhadap sebuah pernyataan. Penentuan emosi terkejut digambarkan pada isi komentar yang memiliki makna kaget atau sesuatu yang mendadak terhadap sebuah pernyataan (Agastya & Aripin, 2020).

2.8 Visualisasi Data

Visualisasi Data adalah sebuah bentuk dalam mempresentasikan sebuah data menjadi bentuk grafis. Digunakan untuk mempercepat memahami sebuah data yang ingin disampaikan. Visualisasi data menggunakan *library WordCloud* (Julianto, 2022).

BAB 3 METODOLOGI PENELITIAN

Bab ini menjelaskan tentang metode penelitian yang digunakan untuk menganalisa data.

3.1 Jenis Penelitian

Jenis Penelitian ini dilakukan secara kuantitatif yang merupakan sebuah penelitian berupa pengolahan dataset dan numerik berdasarkan metode yang digunakan yakni support vector machine.

3.2 Objek Penelitian

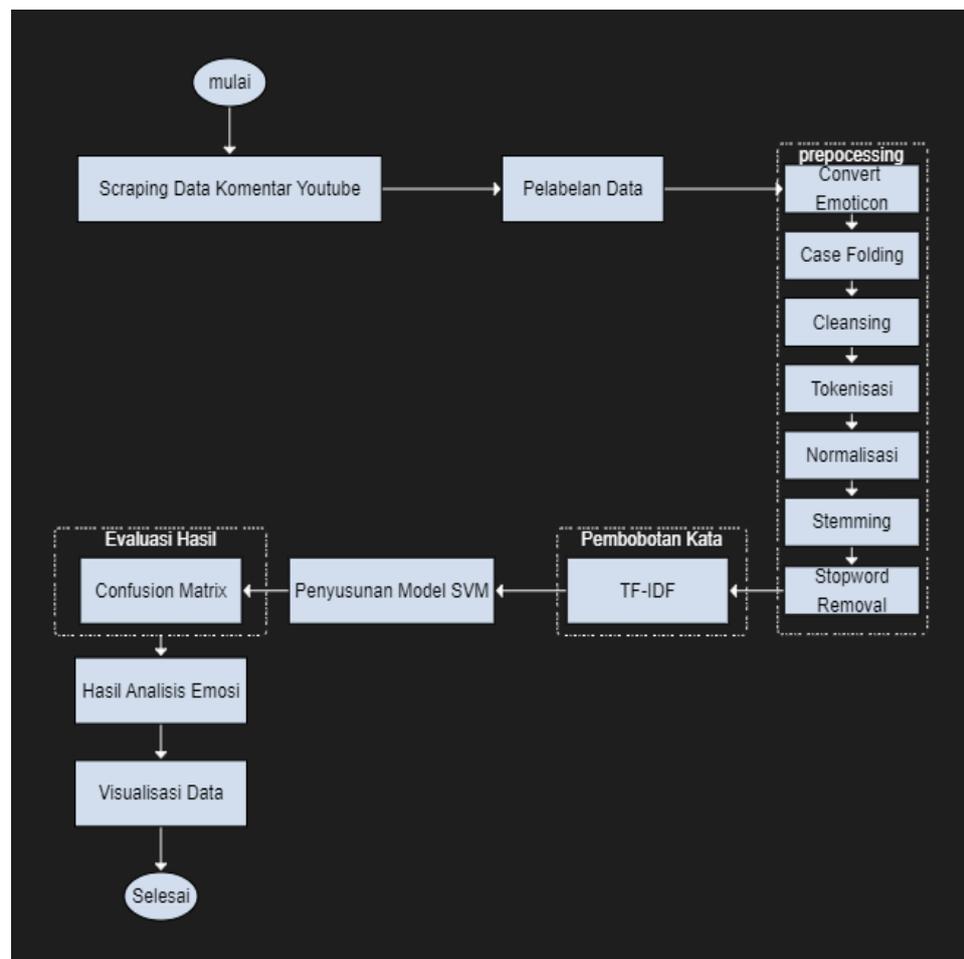
Objek yang dipakai dalam penelitian ini merupakan sebuah pendapat masyarakat yang berupa komentar untuk menyampaikan reaksi-reaksi emosi dalam teks melalui sosial media youtube. Data yang digunakan diambil melalui sosial media YouTube dengan channel Indonesain youtuber dengan video yang berjudul Rewind Indonesia 2021.

3.3 Tempat dan Waktu Penelitian

Penelitian ini dilakukan di Media Sosial Youtube Rewind Indonesia 2021. Waktu penelitian sekitar bulan April - November 2022.

3.4 Tahapan Penelitian

Pada tahapan penelitian dipaparkan mengenai urutan penelitian yang akan dikerjakan peneliti. Tahapan yang digunakan oleh peneliti diilustrasikan pada gambar 3:



Gambar 3 Tahapan Penelitian

3.4.1 Scraping Data Komentar Youtube

Pada tahap Scraping data melakukan proses pengumpulan data berupa komentar yang diekstraksi dari video yang berjudul rewind Indonesia 2021 di media social youtube. Dataset yang digunakan harus bersifat unique atau tidak memiliki data yang sama, dan data tersebut berjumlah sebanyak 2.000 data. Untuk menentukan data yang bersifat unique serta data yang hanya menggunakan Bahasa Indonesia dilakukan dengan memilah secara

manual. Pada tahap ini menggunakan bahasa pemrograman python serta menggunakan library `youtube_comment_scraper_python`.

3.4.2 Pelabelan Data

Pada tahap pelabelan data proses pengerjaan dilakukan secara manual oleh ahli bahasa. Tahap ini menentukan data yang masuk tergolong dalam 6 emosi dasar manusia yaitu marah, sedih, jijik, takut, bahagia, dan terkejut.

3.4.3 Preprocessing

Preprocessing data merupakan sebuah proses untuk mengolah suatu data mentah yang telah dilakukan pelabelan data terlebih dahulu untuk menjadi dataset bersih. Dalam tahap ini menggunakan Bahasa pemrograman python. Langkah-langkah dalam melakukan preprocessing data adalah sebagai berikut:

1. Convert Emoticon

Proses diawali dengan memasukan data komentar, lalu jika terdapat sebuah emoticon, maka dilakukan konversi emoticon ke dalam bentuk string yang sesuai, untuk mengkonversi emoticon kedalam bentuk string menggunakan sebuah kamus yang telah dibuat oleh peneliti, lalu mendapatkan data komentar yang sudah dilakukan convert emoticon dan selesai. Convert emoticon digunakan untuk membantu keakuratan emosi dalam teks.

2. Case Folding

Proses case folding diawali dengan memasukan data komentar yang sudah dilakukan convert emoticon, lalu dilakukan proses dengan mengubah berbagai huruf besar menjadi huruf kecil, mendapatkan hasil dari data komentar yang sudah dilakukan case folding. Case folding digunakan karena perbedaan huruf kapital dan huruf kecil memiliki makna yang berbeda. Proses case folding menggunakan library string dengan function `lower()`.

3. Cleansing

Proses cleansing diawali dengan memasukan data komentar yang sudah dilakukan case folding, memproses dengan membuang bagian-bagian dari data komentar, seperti membuang angka, membuang tanda baca atau karakter, dan hashtag untuk mengurangi noise pada data komentar. Proses cleansing menggunakan library string dan re dengan fungsi `string.punctuation` dan `re.sub()`.

4. Tokenisasi

Tokenisasi diawali dengan memasukan data komentar yang sudah dilakukan cleansing, memproses dengan memecah kalimat menjadi token. Proses tokenisasi ini menggunakan library `nltk.tokenize` dengan fungsi `word_tokenize()`.

5. Normalisasi

Normalisasi diawali dengan memasukan kamus yang sudah didapat melalui github *Colloquial Indonesia Lexicon*, memproses dengan kata yang sudah diperoleh dari komentar youtube yang sudah dilakukan tokenisasi, jika mengandung kata yang tidak baku maka kata tersebut akan berubah menjadi bentuk baku atau formal.

6. Stemming

Proses stemming diawali dengan memasukan data komentar yang sudah dilakukan normalisasi, jika data komentar memiliki kata berimbuhan maka dikembalikan menjadi sebuah kata dasar. Proses stemming ini menggunakan sebuah library `sastrawi` dengan fungsi `StemmerFactory().create_stemmer()`.

7. Stopword Removal

Proses stopword removal dilakukan memasukan data komentar yang sudah dilakukan stemming, memproses

dengan mengeliminasi kata-kata umum yang tidak memiliki makna dan tidak diperlukan. Untuk menyaring kata menggunakan sebuah kamus yang telah dibuat oleh peneliti dan menggunakan library sastrawi dengan fungsi `StopWordRemover().remove()`.

3.5 Pembobotan Kata TF-IDF

Pada tahap untuk pembobotan kata melakukan proses algoritma (TF-IDF). Dalam menaksir nilai suatu bobot pada setiap kata dalam kalimat perlu memanfaatkan Algoritma TF-IDF. Proses ini dilakukan terlebih dahulu dengan mencari nilai Term Frequency (TF), kemudian ditetapkan nilai IDFnya dengan mencari nilai dari logaritma nilai dokumen frekuensi (DF). Selanjutnya, bobot dari setiap kata didalam kalimat dilakukan perhitungan dengan mengalikan nilai yang sudah didapat dari TF dan IDF. Proses TF-IDF menggunakan library `sklearn.feature_extraction.text` dengan fungsi `TfidfVectorizer()`. Setelah melakukan pembobotan kata dengan TF-IDF dilakukan proses analisis emosi menggunakan support vector machine.

3.6 Penyusunan Model SVM

Penyusunan Model SVM dilakukan untuk mengetahui emosi publik dalam teks komentar terhadap Indonesia rewind dengan cara menginputkan data uji komentar yang didapatkan dari hasil scraping data. Dalam metode yang digunakan mengambil kernel linear dalam support vector machine dan menggunakan library `sklearn` fungsi `svm.SVC()` yang menggunakan Bahasa pemrograman python. Hasil dari analisis emosi akan mendapatkan luaran komentar yang memiliki 6 emosi dasar manusia yaitu marah, sedih, jijik, takut, bahagia, dan terkejut dengan memanfaatkan library `sklearn.pipeline` fungsi `Pipeline().fit().predict()`.

3.7 Confusion Matrix

Performa hasil analisis emosi dilakukan evaluasi hasil dengan melakukan confusion matrix untuk menentukan nilai accuracy, presisi, dan recall. Berdasarkan analisis emosi yang dilakukan, hasil dari penerapan confusion matrix digunakan untuk ukuran keberhasilan.

3.8 Hasil Analisis Emosi

Evaluasi analisis emosi adalah untuk dapat mengetahui emosi dari setiap komentar yang akan di uji. Proses tahap analisis emosi dilakukan dengan hasil data testing yang dilakukan prediksi oleh sistem sehingga data komentar pada data testing muncul emosi tiap komentar.

3.9 Visualisasi Data

Visualisasi adalah untuk dapat mengetahui berbagai data yang dipresentasikan menjadi sebuah bentuk grafis agar lebih memudahkan dalam penyampaian informasi yang ingin disampaikan. Karena kinerja dari visualisasi data yang mengkonversikan kumpulan data menjadi lebih sederhana. Visualisasi data ini akan menampilkan sebuah data komentar youtube dari data uji yang sudah dilakukan pada proses analisis emosi. Proses visualisasi data menggunakan *library WordCloud*.

BAB 4 HASIL DAN PEMBAHASAN

Bab ini menjelaskan tentang data yang telah diperoleh akan disajikan, dibahas dan dianalisis. Pembahasan dilakukan guna memaparkan dan menjelaskan bagaimana penelitian ini menjawab rumusan masalah yang telah ditentukan di awal penelitian.

4.1 Hasil Dataset

Pada tahap ini untuk mendapatkan dataset dilakukan scraping data guna mengambil data komentar pada Youtube Indonesia Rewind 2021. Jumlah dataset yang terkumpul sebanyak 2000 data dengan bermacam-macam komentar. Pengambilan dataset yang dilakukan pada komentar youtube Indonesia Rewind 2021 digunakan peneliti untuk data training dan data testing. Jumlah dataset yang terkumpul sebanyak 2000 data dengan bermacam-macam komentar. Data training yang digunakan bertujuan sebagai bahan pembelajaran di tahap klasifikasi sehingga dari hasil klasifikasi dapat digunakan untuk menguji data testing. Dataset diperoleh dari scraping data dan diberikan label secara manual oleh ahli bahasa yaitu Anna Dianlat Olinda, S.Pd selaku Dosen Program Studi Bahasa Indonesia Fakultas Keguruan dan Ilmu Pendidikan Universitas Negeri Malang. Dataset pada penelitian ini dibagi menjadi data training dan data testing dengan perbandingan 90%:10%, 80%:20%, 70%:30%, 60:40%. Kelas klasifikasi pada penelitian ini menggunakan 6 emosi dasar manusia yaitu Jijik, sedih, takut, marah, terkejut, dan bahagia. Jumlah kelas yang bersifat jijik sebanyak 7 komentar, kelas yang bersifat sedih sebanyak 287 komentar, kelas yang bersifat takut sebanyak 26 komentar, kelas yang bersifat marah sebanyak 139 komentar, kelas yang bersifat terkejut sebanyak 96 komentar dan kelas yang bersifat bahagia sebanyak 1445 komentar. Contoh kode program yang digunakan pada penelitian ini dapat dilihat pada gambar 4 sampai 7 sebagai berikut :

```

#----- split data tes 90:10
input_data = df["stopword"]
output_data = df["data"] Loading...
train_data, test_data, train_output, test_output = train_test_split(input_data, output_data, test_size=0.10, random_state=0)

```

Gambar 4 Split Data Tes 90:10

```

#----- split data tes 80:20
input_data = df["stopword"]
output_data = df["dataemosi"]
train_data, test_data, train_output, test_output = train_test_split(input_data, output_data, test_size=0.20, random_state=0)

```

Gambar 5 Split Data Tes 80:20

```

#----- split data tes 70:30
input_data = df["stopword"]
output_data = df["dataemosi"]
train_data, test_data, train_output, test_output = train_test_split(input_data, output_data, test_size=0.30, random_state=0)

```

Gambar 6 Split Data Tes 70:30

```

#----- split data tes 60:40
input_data = df["stopword"]
output_data = df["dataemosi"]
train_data, test_data, train_output, test_output = train_test_split(input_data, output_data, test_size=0.40, random_state=0)

```

Gambar 7 Split Data Tes 60:40

Adapun contoh dataset yang digunakan pada penelitian ini dapat dilihat pada tabel 2 sebagai berikut :

Tabel 2 Contoh Dataset

No	Data komentaremosi	Dataemosi
1	1/3 awal rada jijik apa-apaan ada sponsor ih jijik banget sih	Jijik
2	setelah menonton ini saya takut ditinggal orang tersayang yakni ibu saya sedih	Takut

3	2020 suram. 2021 Dah mulai ada aktivitas keren	Bahagia
4	Dosa kah kita punya kulit yang hitam	Sedih
5	1/3 video cuma joget, dan lu meneteskan air mata	Terkejut
6	Bisa membuat Rewind Indonesia 2022 kah	Marah

4.2 Hasil Pembuatan Kamus Emoticon

Dalam pembuatan kamus Emoticon didapat melalui hasil dari pengertian yang terdapat didalam kamus emoji yang dapat dicari melalui web search yaitu google. Dengan cara memilah emoji yang digunakan pada kalimat komentar dari hasil pengumpulan dataset. Daftar pembuatan kamus emoticon yang telah dibuat oleh penulis dilampirkan pada gambar 8 :

```
{
  ">.<": "gemas",
  "T_T": "nangis",
  "<3": "cinta",
  ":)": "nangis",
  ":)": "senyum",
  ":(": "sedih",
  ":(": "sedih",
  "ID": "indonesia",
  "😂": "lucu",
  "😊": "bahagia",
  "😭": "nangis banget",
  "❤️": "cinta",
  "♥️": "cinta",
  "🙌": "angkat tangan",
  "💪": "kuat",
  "🔥": "membara",
  "💜": "cinta",
  "😄": "senang",
  "🗿": "patung batu",
  "😞": "sedih",
  "😭": "nangis",
  "😭": "mau menangis",
  "😍": "cinta senang",
  "🥶": "dingin",
  "🥵": "panas",
  "😂": "tertawa",
  "👍": "bagus",
  "😄": "sangat senang",
  "😡": "sangat marah",
  "🌟": "bintang",
  "♥️": "cinta",
}
```

Gambar 8 Contoh Kamus Emoticon

4.3 Hasil Pembuatan Kamus Stopword

Dalam pembuatan kamus Stopword didapat melalui hasil dari data training. Daftar pembuatan kamus stopwords yang telah dibuat oleh penulis dilampirkan pada tabel 3 :

Tabel 3 Contoh Kamus Stopword

No	Kata Stopword
1	ada
2	adalah
3	awal
4	awalnya
5	bisa
6	bisakah
7	boleh
8	itu
9	Dah
10	Ini

4.4 Hasil Implementasi Sistem

Tahap Implementasi sistem merupakan tahapan yang telah dilaksanakan oleh penulis pada penelitian ini menggunakan bahasa pemrograman python 3.7.14 . Hasil dari implementasi sistem dijelaskan sebagai berikut:

4.4.1 Implementasi Scraping data

Pada tahap ini untuk mendapatkan dataset dilakukan scraping data guna mengambil data komentar pada Youtube Indonesia Rewind 2021. Proses scraping data dilakukan menggunakan bahasa pemrograman python 3.7.14. Untuk melakukan scraping data diperlukan link youtube Indonesia Rewind 2021 untuk diambil komentarnya. Selain itu proses scraping data menggunakan library scraping comment (youtube_comment_scraper_python). Berikut merupakan scraping yang digunakan dapat dilihat pada gambar 9 :

```

1  from youtube_comment_scraper_python import *
2  import pandas as pd
3
4  youtube.open("https://youtu.be/IPZO85HFM4w")
5  youtube.keypress("pagedown")
6
7  all_data = []
8  for i in range(0, 480):
9      response = youtube.video_comments()
10     data = response['body']
11     all_data.extend(data)
12
13     df = pd.DataFrame(data)
14
15     df = df.replace('\n', ' ', regex=True)
16
17     df = df[['Comment']].drop_duplicates(keep="first")
18
19     df.to_csv('YoutubeCommentScrapper.csv', index=False)

```

Gambar 9 Kode Program Scraping

4.4.2 Implementasi preprocessing

Implementasi tahap preprocessing dimulai dengan proses convert emoticon. Proses tersebut dilakukan dengan mengubah emoticon menjadi teks. Proses ini menggunakan kamus emoji manual yang dibuat penulis karena jika menggunakan library emoji yang tersedia pada python output tersebut menghasilkan kalimat dengan bahasa inggris sehingga tidak sesuai dengan batasan masalah peneliti. Kode pada tahap convert emoticon dapat dilihat pada gambar 10.

```

# -----CONVERT EMOTICON-----#
def convert_emoticon(data):
    openemoticon = eval(open("/content/drive/MyDrive/skripsi_final/kamus_emot.txt").read())
    for emoticon, emot in openemoticon.items():
        data = data.replace(emoticon, "+emot+ ")
    return data

df["convert_emoticon"]=df["datakomentar"].apply(lambda data: convert_emoticon(data))
df.head(2001)

```

Gambar 10 Kode Program Convert Emoticon

Setelah tahap tersebut selesai, dilanjutkan yaitu tahap Case Folding. Pada tahap ini dilakukan proses mengubah semua huruf pada data komentar menjadi huruf kecil / case folding dengan menggunakan kode program `text = text.lower()`. Kode program pada tahap case folding dapat dilihat pada gambar 11.

```
[6] # -----CASE FOLDING-----#
df["case_folding"] = df["convert_emoticon"].str.lower()
df.head()
```

Gambar 11 Kode Program Case Folding

Apabila proses tersebut sudah selesai dilanjutkan dengan tahap cleansing. Pada tahap cleansing dilakukan dengan menghapus tanda baca yang tidak diperlukan dengan kode program `remove = string.punctuation`, menghapus ASCII dan Unicode, dan menghapus Newline. Library yang digunakan pada tahap ini yaitu Library String. Berikut kode program pada tahap cleansing dapat dilihat pada gambar 12.

```
def cleansing(data):
    #Remove Punctuation
    remove=string.punctuation
    translator=str.maketrans(remove, '*len(remove))
    data=data.translate(translator)

    #Remove ASCII & UNICODE
    data=data.encode('ascii','ignore').decode('utf-8')
    data=re.sub(r'^\x00-\x7f',r'', data)

    #Remove Newline
    data=data.replace('\n',' ')

    return data

df["cleansing"] = df["case_folding"].apply(cleansing)
```

Gambar 12 Kode Program Cleansing

Apabila setelah proses tersebut selesai kemudian dilanjutkan dengan tahap tokenisasi. Pada tahap tokenisasi dilakukan proses memecah

atau memisahkan kalimat menjadi per kata / Token. Library yang digunakan yaitu `from nltk.tokenize import word_tokenize`. Kode program yang digunakan pada tahap tokenisasi dapat dilihat pada gambar 13.

```
def tokenisasi(t):
    t = word_tokenize(t)
    return t
df["tokenisasi"]=df["cleansing"].apply(tokenisasi)
```

Gambar 13 Kode Program Tokenisasi

Selanjutnya setelah dilakukan tokenisasi dilanjutkan dengan tahap normalization. Pada tahap ini dilakukan merubah kalimat tidak baku menjadi kalimat baku. Pada tahap ini menggunakan kamus normalization yang telah didapat pada *Colloquial Indonesia Lexicon* yang didapat melalui github. Kode program pada tahap normalization dapat dilihat pada gambar 14.

```
normalization_word = pd.read_csv('/content/drive/MyDrive/final_skripsi_fix/colloquial-indonesian-lexicon.csv')
normalization_word_dict = {}

for index, row in normalization_word.iterrows():
    if row[0] not in normalization_word_dict:
        normalization_word_dict[row[0]] = row[1]

def normalization(document):
    return [normalization_word_dict[term] if term in normalization_word_dict else term for term in document]

df['Normalization'] = df['tokenisasi'].apply(normalization)
df['Normalization']=df['tokenisasi'].apply(lambda x: TreebankWordDetokenizer().detokenize(x))
```

Gambar 14 Kode Program Normalization

Selanjutnya dilanjutkan dengan tahap stemming. Pada tahap ini dilakukan mengembalikan kata yang berimbuhan menjadi kata dasar. Library yang digunakan pada tahap ini yaitu menggunakan Library Sastrawi. Kode program pada tahap stemming dapat dilihat pada gambar 15.

```
def stem_words(text):
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    return " ".join([stemmer.stem(word) for word in text.split()])

df["stemming"] = df["Normalization"].apply(stem_words)
```

Gambar 15 Kode Program Stemming

Setelah tahap tokenisasi selesai kemudian dilanjutkan dengan tahap stopword removal. Pada tahap stopword removal dilakukan dengan mengeliminasi kata-kata yang tidak memiliki makna. Pada tahap ini menggunakan kamus yang dibuat oleh peneliti. Kode program yang digunakan pada tahap stopword removal dapat dilihat pada gambar 16.

```
def stopword(text):
    with open('/content/drive/MyDrive/final_skripsi_fix/kamus_2.txt') as txt_file:
        content = txt_file.readlines()
        ls_stopword = list(map(lambda x: x.replace("\n", ""), content))
        dictionary = ArrayDictionary(ls_stopword)
        stopword = StopWordRemover(dictionary)
        text = stopword.remove(text)
    return (text)

df["stopword"] = df["stemming"].apply(stopword)
```

Gambar 16 Kode Program Stopword

4.4.3 Implementasi TF IDF

Proses ini memanfaatkan algoritma TF-IDF (Term Frequency-Inverse Document Frequency) digunakan untuk melakukan sebuah pembobotan kata. Dalam memanfaatkan algoritma TF-IDF memanfaatkan library dari `sklearn.feature_extraction.text` import `TfidfVectorizer`. Digunakan oleh beberapa mesin pencari untuk membantu mendapatkan hasil yang lebih baik yang lebih relevan dengan query tertentu. Kode program pada implementasi TF-IDF dapat dilihat pada gambar 17:

```

MAX_FEATURES = 2000

tfidf = TfidfVectorizer(max_features=MAX_FEATURES, ngram_range=(1,2))
tfidf_text_train = tfidf.fit_transform(train_data.values.astype('U'))
tfidf_text_test = tfidf.transform(test_data.values.astype('U'))

```

Gambar 17 Kode Program TF-IDF

4.4.4 Implementasi penyusunan model

Penyusunan model klasifikasi diimplementasikan dengan metode Support Vector Machine. Pada tahap implementasi model yaitu dilakukan penentuan kelas emosi. Langkah awal pada tahap ini menentukan kernel yang dipakai dalam SVM. Kernel yang digunakan dalam penelitian yaitu Linear. Library yang digunakan untuk kernel yaitu sklearn import svm. Selanjutnya menentukan prediksi kelas emosi dengan memanfaatkan kernel tersebut serta memanfaatkan library pipeline dari TF-IDF atau Pembobotan kata dengan kernel. Kode program pada penyusunan model yang digunakan dapat dilihat pada gambar 18.

```

from sklearn import svm
linier_svc = svm.SVC(kernel = 'linear', decision_function_shape = 'ovo',C=1)
linier_svc.fit(tfidf_text_train,train_output)
predictions = linier_svc.predict(tfidf_text_test)
emotion_ = Pipeline([('tfidf', TfidfVectorizer(max_features=MAX_FEATURES, ngram_range=(1,1))),('linier_svc', svm.SVC(kernel = 'linear', decis:

emotion_.fit(train_data,train_output)
pred = emotion_.predict(test_data)

```

Gambar 18 Kode Program Penyusunan Model

4.4.5 Implementasi confusion matrix

Pada tahap ini dilakukan untuk mengetahui nilai persentase akurasi yang sudah dilakukan oleh sistem dengan mengacu dari hasil akurasi. Library yang digunakan pada tahap ini yaitu from sklearn.metrics. Implementasi dari confusion matrix pada sistem menggunakan akurasi, presisi dan recall. Kode program yang digunakan pada confusion matrix dapat dilihat gambar 19.

```
print(classification_report(test_output,pred))
acc = accuracy_score(test_output,pred)
print(acc)
plot_confusion_matrix(linier_svc,tfidf_text_test,test_output)
```

Gambar 19 Kode Program Implementasi Confusion Matrix

4.4.6 Implementasi hasil analisis emosi

Implementasi emosi dilaksanakan guna menganalisa emosi dari data komentar yang telah diuji mengandung 6 emosi dasar yaitu bahagia, jijik, marah, sedih, takut dan terkejut dengan kode program analisis emosi yang dapat dilihat pada gambar 20:

```
[82] def prediksiemosi(t):
      return emotion_.predict([t])
```

Gambar 20 Kode Program Hasil Analisis Emosi

4.4.7 Implementasi visualisasi data

Implementasi visualisasi data dilaksanakan guna mendapatkan kata terbanyak yang terdapat dari data komentar yang telah diuji dengan menggunakan library wordcloud dengan kode program visualisasi data yang dapat dilihat pada gambar 21:

```
bahagia = df[df['data_emosi_baru'].str.contains('bahagia')]
bahagia = bahagia['DataKomentaremosi']

data = ["ini", "yg", "yang", "yang", "dan", "ga", "di", "jadi", "utk", "ke", "iya", "intm", "gw", "kalo", "gak", "itu", "lah", "tapi", "udah", "you", "nyaaaa", "kek", "u",
"tdk", "nyiapin", "adlh", "ala2", "ngga", "step", "ferry", "si", "d", "nya", "sama", "terakhir", "tahun", "jgn", "jg", "coba", "this", "benar", "berharap", "mata", "mal",
"ya", "ada", "dari", "aja", "aku", "lagi", "sih", "iniii", "saya", "menjadi", "gini", "kan", "dll", "gitu", "anti", "lbh", "kenapa", "pahami", "jika", "namanya", "klo", "p",
"para", "kok", "mba", "bang", "banget", "bgt", "kita", "selalu", "so", "benar2", "dgn", "lalu", "buat", "mana", "pun", "setiap", "kah", "harus", "pokok", "tuh", "karna",
"semua", "bisa", "juga", "bnget", "liat", "dapat", "gua", "tunggu", "kaya", "helm", "telinga", "mulu", "hampa", "masuk", "semuanya", "muda", "gasil", "full", "bangett",
"diliat", "besar", "apaa", "blak", "nggak", "sebuah", "sangat", "anak", "hal2", "ngetik", "dritadi", "kayak", "dikemas", "masa", "bangett", "the", "sekali", "mu", "ne",
"antar", "apa", "tetep", "menurut", "lari", "kali", "atau", "bian", "lebih", "satu", "makin", "untuk", "anak2", "ngerti", "tidak", "bukan", "dapat", "ganti", "untuk",
"sampe", "banyak", "hari", "memberi", "sm", "saat", "beda", "aduk", "sekarang", "luar", "baru", "orang", "guys", "punya", "sampai", "memberikan", "masih", "cuma", "kaka",
"cuuyy", "lihat", "terima", "deh", "bagian tema", "catur", "bagian bangsa", "kasih", "bagian", "biasa", "sekilas", "kalian"]

plt.figure(figsize = (20,20))
wc = WordCloud(max_words = 60 , width = 800 , height = 400,
               collocations=False, stopwords=data).generate(" ".join(bahagia))
plt.imshow(wc)
```

Gambar 21 Kode Program Implementasi Visualisasi Data

4.5 Hasil Preprocessing

Pada Tahap Preprocessing diawali dengan tahap convert emoticon yaitu mengubah emoticon menjadi teks. Convert emoticon menggunakan kamus yang dibuat secara manual oleh peneliti dan menggunakan Bahasa pemrograman python. Hasil convert emoticon dapat dilihat pada tabel 4 sebagai berikut :

4.5.1 Hasil Convert emoticon

Tabel 4 Contoh Hasil Convert emoticon

No	Data komentaremosi	convert _emoticon
1	1/3 awal rada jijik apa-apaan ada sponsor ih jijik banget sih	1/3 awal rada jijik apa-apaan ada sponsor ih jijik banget sih
2	20:9 maksudnya bang 🤔 biasa kan kalau layar yang masih ada atasnya itu 16:9, nah ini yang bener2 full layar, kece banget tim nya 🤔	20:9 maksudnya bang lucu biasa kan kalau layar yang masih ada atasnya itu 16:9, nah ini yang bener2 full layar, kece banget tim nya nangis banget
3	2020 suram. 2021 Dah mulai ada aktivitas keren	2020 suram. 2021 Dah mulai ada aktivitas keren
4	iya yak, lebih ke tiktok rewind sih. Tapi keren sih efek2 visualnya menurut gua	iya yak, lebih ke tiktok rewind sih. Tapi keren sih efek2 visualnya menurut gua
5	oh iya kah!	oh iya kah!
6	tetap mengapresiasi dengan tema dan nilai di dalamnya.	tetap mengapresiasi dengan tema dan nilai di dalamnya.
7	setelah menonton ini saya takut ditinggal orang tersayang yakni ibu saya :(setelah menonton ini saya takut ditinggal orang tersayang yakni ibu saya sedih
8	"Dosa kah kita punya kulit yang hitam?"	"Dosa kah kita punya kulit yang hitam?"
9	"Meraup keuntungan dari kesempatan"Money heist	"Meraup keuntungan dari kesempatan"Money heist
10	Semangat gotong royong 🤝	Semangat gotong royong kuat

Setelah tahap convert emoticon selesai maka dilanjutkan dengan tahap case folding. Pada tahap case folding dilakukan mengubah semua

huruf pada data komentar menjadi huruf kecil. Hasil Case folding dapat dilihat pada tabel 5 sebagai berikut :

4.5.2 Hasil Case Folding

Tabel 5 Contoh Hasil Case Folding

No	Data komentar emosi	case_folding
1	1/3 awal rada jijik apa-apaan ada sponsor ih jijik banget sih	1/3 awal rada jijik apa-apaan ada sponsor ih jijik banget sih
2	20:9 maksudnya bang 🤔 biasa kan kalau layar yang masih ada atasnya itu 16:9, nah ini yang bener2 full layar, kece banget tim nya 🤔	20:9 maksudnya bang lucu biasa kan kalau layar yang masih ada atasnya itu 16:9, nah ini yang bener2 full layar, kece banget tim nya nangis banget
3	2020 suram. 2021 Dah mulai ada aktivitas keren	2020 suram. 2021 dah mulai ada aktivitas keren
4	iya yak, lebih ke tiktok rewind sih. Tapi keren sih efek2 visualnya menurut gua	iya yak, lebih ke tiktok rewind sih. tapi keren sih efek2 visualnya menurut gua
5	oh iya kah!	oh iya kah!
6	tetap mengapresiasi dengan tema dan nilai di dalamnya.	tetap mengapresiasi dengan tema dan nilai di dalamnya.
7	setelah menonton ini saya takut ditinggal orang tersayang yakni ibu saya :(setelah menonton ini saya takut ditinggal orang tersayang yakni ibu saya sedih
8	"Dosa kah kita punya kulit yang hitam?"	"dosa kah kita punya kulit yang hitam?"
9	"Meraup keuntungan dari kesempatan" Money heist	"meraup keuntungan dari kesempatan" money heist
10	Semangat gotong royong 🙌	semangat gotong royong kuat

Setelah selesai dilakukannya tahap case folding maka dilanjutkan pada tahap cleansing. Pada tahap cleansing dilakukan menghapus tanda baca yang tidak diperlukan. Hasil cleansing dapat dilihat pada tabel 6 sebagai berikut :

4.5.3 Hasil Cleansing

Tabel 6 Contoh hasil Cleansing

No	Data komentaremosi	Cleansing
1	1/3 awal rada jijik apa-apaan ada sponsor ih jijik banget sih	13 awal rada jijik apaapaan ada sponsor ih jijik banget sih
2	20:9 maksudnya bang 🤔 biasa kan kalau layar yang masih ada atasnya itu 16:9, nah ini yang bener2 full layar, kece banget tim nya 🤔	209 maksudnya bang lucu biasa kan kalau layar yang masih ada atasnya itu 169 nah ini yang bener2 full layar kece banget tim nya nangis banget
3	2020 suram. 2021 Dah mulai ada aktivitas keren	2020 suram 2021 dah mulai ada aktivitas keren
4	iya yak, lebih ke tiktok rewind sih. Tapi keren sih efek2 visualnya menurut gua	iya yak lebih ke tiktok rewind sih tapi keren sih efek2 visualnya menurut gua
5	oh iya kah!	oh iya kah
6	tetap mengapresiasi dengan tema dan nilai di dalamnya.	tetap mengapresiasi dengan tema dan nilai di dalamnya
7	setelah menonton ini saya takut ditinggal orang tersayang yakni ibu saya :(setelah menonton ini saya takut ditinggal orang tersayang yakni ibu saya sedih
8	"Dosa kah kita punya kulit yang hitam?"	dosa kah kita punya kulit yang hitam
9	"Meraup keuntungan dari kesempatan" Money heist	meraup keuntungan dari kesempatan money heist
10	Semangat gotong royong 🙌	semangat gotong royong kuat

Setelah selesai dilakukannya tahap cleansing maka dilanjutkan pada tahap tokenisasi. Pada tahap tokenisasi memproses dengan memecah kalimat menjadi token. Hasil tokenisasi dapat dilihat pada tabel 7 sebagai berikut :

4.5.4 Hasil tokenisasi

Tabel 7 Contoh Hasil Tokenisasi

No	Data komentaremosi	Tokenisasi
1	1/3 awal rada jijik apa-apaan ada sponsor ih jijik banget sih	['13', 'awal', 'rada', 'jijik', 'apaapaan', 'ada', 'sponsor', 'ih', 'jijik', 'banget', 'sih']
2	20:9 maksudnya bang 🤔 biasa kan kalau layar yang masih ada atasnya itu 16:9, nah ini yang bener2 full layar, kece banget tim nya 🤔	['209', 'maksud', 'bang', 'lucu', 'biasa', 'kan', 'kalau', 'layar', 'yang', 'masih', 'ada', 'atas', 'itu', '169', 'nah', 'ini', 'yang', 'bener2', 'full', 'layar', 'kece', 'banget', 'tim', 'nya', 'nang', 'banget']
3	2020 suram. 2021 Dah mulai ada aktivitas keren	['2020', 'suram', '2021', 'dah', 'mulai', 'ada', 'aktivitas', 'keren']
4	iya yak, lebih ke tiktok rewind sih. Tapi keren sih efek2 visualnya menurut gua	['iya', 'yak', 'lebih', 'ke', 'tiktok', 'rewind', 'sih', 'tapi', 'keren', 'sih', 'efek2', 'visual', 'turut', 'gua']
5	oh iya kah!	['oh', 'iya', 'kah']
6	tetap mengapresiasi dengan tema dan nilai di dalamnya.	['tetap', 'apresiasi', 'dengan', 'tema', 'dan', 'nilai', 'di', 'dalam']
7	setelah menonton ini saya takut ditinggal orang tersayang yakni ibu saya :(['telah', 'tonton', 'ini', 'saya', 'takut', 'tinggal', 'orang', 'sayang', 'yakni', 'ibu', 'saya', 'sedih']
8	"Dosa kah kita punya kulit yang hitam?"	['dosa', 'kah', 'kita', 'punya', 'kulit', 'yang', 'hitam']
9	"Meraup keuntungan dari kesempatan" Money heist	['raup', 'untung', 'dari', 'kesempatan', 'money', 'heist']
10	Semangat gotong royong 🙌	['semangat', 'gotong', 'royong', 'kuat']

Setelah selesai dilakukannya tahap tokenisasi maka dilanjutkan pada tahap normalization. Pada tahap normalization dilakukan merubah kalimat tidak baku menjadi kalimat baku menggunakan kamus *Colloquial Indonesia Lexicon*. Hasil normalization dapat dilihat pada tabel 8 sebagai berikut :

4.5.5 Hasil Normalization

Tabel 8 Contoh Hasil Normalization

No	Data komentaremosi	Normalization
1	1/3 awal rada jijik apa-apaan ada sponsor ih jijik banget sih	1 3 awal rada jijik apa apaan ada sponsor ih jijik banget sih
2	20:9 maksudnya bang 🤔 biasa kan kalau layar yang masih ada atasnya itu 16:9, nah ini yang bener2 full layar, kece banget tim nya 😊	20 9 maksudnya bang lucu biasa kan kalau layar yang masih ada atasnya itu 16 9 nah ini yang bener2 full layar kece banget tim nya nangis banget
3	2020 suram. 2021 Dah mulai ada aktivitas keren	2020 suram 2021 dah mulai ada aktivitas keren
4	iya yak, lebih ke tiktok rewind sih. Tapi keren sih efek2 visualnya menurut gua	iya yak lebih ke tiktok rewind sih tapi keren sih efek2 visualnya menurut gua
5	oh iya kah!	oh iya kah
6	tetap mengapresiasi dengan tema dan nilai di dalamnya.	tetap mengapresiasi dengan tema dan nilai di dalamnya
7	setelah menonton ini saya takut ditinggal orang tersayang yakni ibu saya :(setelah menonton ini saya takut ditinggal orang tersayang yakni ibu saya sedih
8	"Dosa kah kita punya kulit yang hitam?"	dosa kah kita punya kulit yang hitam
9	"Meraup keuntungan dari kesempatan" Money heist	meraup keuntungan dari kesempatan money heist
10	Semangat gotong royong 🦷	semangat gotong royong kuat

Setelah selesai dilakukannya tahap normalization maka dilanjutkan pada tahap stemming. Pada tahap stemming dilakukan mengembalikan kata yang berimbuhan menjadi kata dasar. Hasil stemming dapat dilihat pada tabel 9 sebagai berikut :

4.5.6 Hasil stemming

Tabel 9 Contoh Hasil Stemming

No	Data komentaremosi	Stemming
1	1/3 awal rada jijik apa-apaan ada sponsor ih jijik banget sih	13 awal rada jijik apaapaan ada sponsor ih jijik banget sih
2	20:9 maksudnya bang 🤔 biasa kan kalau layar yang masih ada atasnya itu 16:9, nah ini yang bener2 full layar, kece banget tim nya 🤔	209 maksud bang lucu biasa kan kalau layar yang masih ada atas itu 169 nah ini yang bener2 full layar kece banget tim nya nang banget
3	2020 suram. 2021 Dah mulai ada aktivitas keren	2020 suram 2021 dah mulai ada aktivitas keren
4	iya yak, lebih ke tiktok rewind sih. Tapi keren sih efek2 visualnya menurut gua	iya yak lebih ke tiktok rewind sih tapi keren sih efek2 visual turut gua
5	oh iya kah!	oh iya kah
6	tetap mengapresiasi dengan tema dan nilai di dalamnya.	tetap apresiasi dengan tema dan nilai di dalam
7	setelah menonton ini saya takut ditinggal orang tersayang yakni ibu saya :(telah tonton ini saya takut tinggal orang sayang yakni ibu saya sedih
8	"Dosa kah kita punya kulit yang hitam?"	dosa kah kita punya kulit yang hitam
9	"Meraup keuntungan dari kesempatan" Money heist	raup untung dari kesempatan money heist
10	Semangat gotong royong 🙌	semangat gotong royong kuat

Setelah selesai dilakukannya tahap tokenisasi maka dilanjutkan pada tahap stopword removal. Pada tahap stopword removal memproses dengan mengeliminasi kata-kata umum yang tidak memiliki makna dan tidak diperlukan. Hasil stopword removal dapat dilihat pada tabel 10 sebagai berikut:

4.5.7 Hasil Stopword Removal

Tabel 10 Contoh Hasil Remvoal

No	Data komentaremosi	Stopword
1	1/3 awal rada jijik apa-apaan ada sponsor ih jijik banget sih	['13', 'awal', 'rada', 'jijik', 'apaapaan', 'ada', 'sponsor', 'ih', 'jijik', 'banget', 'sih']
2	20:9 maksudnya bang 🤔 biasa kan kalau layar yang masih ada atasnya itu 16:9, nah ini yang bener2 full layar, kece banget tim nya 🤔	['209', 'maksud', 'bang', 'lucu', 'biasa', 'kan', 'kalau', 'layar', 'yang', 'masih', 'ada', 'atas', 'itu', '169', 'nah', 'ini', 'yang', 'bener2', 'full', 'layar', 'kece', 'banget', 'tim', 'nya', 'nang', 'banget']
3	2020 suram. 2021 Dah mulai ada aktivitas keren	['2020', 'suram', '2021', 'dah', 'mulai', 'ada', 'aktivitas', 'keren']
4	iya yak, lebih ke tiktok rewind sih. Tapi keren sih efek2 visualnya menurut gua	['iya', 'yak', 'lebih', 'ke', 'tiktok', 'rewind', 'sih', 'tapi', 'keren', 'sih', 'efek2', 'visual', 'turut', 'gua']
5	oh iya kah!	['oh', 'iya', 'kah']
6	tetap mengapresiasi dengan tema dan nilai di dalamnya.	['tetap', 'apresiasi', 'dengan', 'tema', 'dan', 'nilai', 'di', 'dalam']
7	setelah menonton ini saya takut ditinggal orang tersayang yakni ibu saya :(['telah', 'tonton', 'ini', 'saya', 'takut', 'tinggal', 'orang', 'sayang', 'yakni', 'ibu', 'saya', 'sedih']
8	"Dosa kah kita punya kulit yang hitam?"	['dosa', 'kah', 'kita', 'punya', 'kulit', 'yang', 'hitam']
9	"Meraup keuntungan dari kesempatan" Money heist	['raup', 'untung', 'dari', 'kesempatanmoney', 'heist']
10	Semangat gotong royong 🙌	['semangat', 'gotong', 'royong', 'kuat']

4.6 Hasil TF IDF

Perhitungan nilai tf dihitung dengan menggunakan banyaknya kata dalam 1 dokumen. Tabel 11 merupakan kelas bahagia, tabel 12 kelas jijik, tabel 13 kelas marah, tabel 14 kelas sedih, tabel 15 kelas takut dan tabel 16 kelas terkejut.

Tabel 11 Contoh Hasil Nilai TF Bahagia

Kata	Nilai TF
2020	1
Suram	1
2021	1
dah	1
mulai	1
ada	1
aktivitas	1
Keren	1

Tabel 12 Contoh Hasil Nilai TF Jijik

Kata	Nilai TF
1/3	1
awal	1
rada	1
jijik	2
Apa	1
apaan	1
Ada	1
sponsor	1

lh	1
banget	1
Sih	1

Tabel 13 Contoh Hasil Nilai TF Marah

Kata	Nilai TF
Bisa	1
Membuat	1
Rewind	1
Indonesia	1
2022	1
Kah	1

Tabel 14 Contoh Hasil Nilai TF Sedih

Kata	Nilai TF
Dosa	1
kah	1
kita	1
Punya	1
kulit	1
Yang	1
Hitam	1

Tabel 15 Contoh Hasil Nilai TF Takut

Kata	Nilai TF
setelah	1

menonton	1
ini	1
Saya	2
takut	1
ditinggal	1
orang	1
Tersayang	1
yakni	1
ibu	1
Sedih	1

Tabel 16 Contoh Hasil Nilai TF Terkejut

Kata	Nilai TF
1/3	1
video	1
Cuma	1
Joget	1
dan	1
lu	1
meneteskan	1
air	1
Mata	1

Bobot IDF dihitung melalui $\log(N/df)$, untuk menghitung idf diperlukan nilai N dan df. Hasil nilai IDF dapat dilihat pada tabel 17.

Tabel 17 Contoh Hasil Nilai IDF

Kata	Df	N	idf=log(N/df)
1/3	2	6	0,477121
2020	1	6	0,778151
2021	1	6	0,778151
2022	1	6	0,778151
Ada	2	6	0,477121
Air	1	6	0,778151
Aktivitas	1	6	0,778151
Apa	1	6	0,778151
Apaan	1	6	0,778151
Awal	1	6	0,778151
banget	1	6	0,778151
Bisa	1	6	0,778151
Cuma	1	6	0,778151

dah	1	6	0,778151
dan	1	6	0,778151
ditinggal	1	6	0,778151
Dosa	1	6	0,778151
Hitam	1	6	0,778151
ibu	1	6	0,778151
Ih	1	6	0,778151
Indonesia	1	6	0,778151
ini	1	6	0,778151
jijik	2	6	0,477121
Joget	1	6	0,778151
Kah	2	6	0,477121
Keren	1	6	0,778151
kita	1	6	0,778151
kulit	1	6	0,778151
lu	1	6	0,778151

Mata	1	6	0,778151
Membuat	1	6	0,778151
meneteskan	1	6	0,778151
menonton	1	6	0,778151
mulai	1	6	0,778151
orang	1	6	0,778151
Punya	1	6	0,778151
rada	1	6	0,778151
Rewind	1	6	0,778151
Saya	2	6	0,477121
Sedih	1	6	0,778151
setelah	1	6	0,778151
Sih	1	6	0,778151
sponsor	1	6	0,778151
Suram	1	6	0,778151
takut	1	6	0,778151

Tersayang	1	6	0,778151
video	1	6	0,778151
yakni	1	6	0,778151
Yang	1	6	0,778151

Bobot TF-IDF dihitung melalui perkalian antara nilai TF dan nilai IDF. tabel 18 merupakan kelas bahagia, tabel 19 kelas jijik, tabel 20 kelas marah, tabel 21 kelas sedih, tabel 22 kelas takut dan tabel 23 kelas terkejut.

Tabel 18 Contoh Hasil Nilai TF-IDF Bahagia

Kata	Nilai TF	Nilai IDF	Nilai TF-IDF
2020	1	0,778151	0,778151
Suram	1	0,778151	0,778151
2021	1	0,778151	0,778151
dah	1	0,778151	0,778151
mulai	1	0,778151	0,778151
ada	1	0,778151	0,778151
aktivitas	1	0,778151	0,778151
Keren	1	0,778151	0,778151

Tabel 19 Contoh Hasil Nilai TF-IDF Jjrik

Kata	Nilai TF	Nilai IDF	Nilai TF-IDF
1/3	1	0,778151	0,778151
awal	1	0,778151	0,778151
rada	1	0,778151	0,778151
jjrik	2	0,477121	0,954242
Apa	1	0,778151	0,778151
apaan	1	0,778151	0,778151
Ada	1	0,778151	0,778151
sponsor	1	0,778151	0,778151
lh	1	0,778151	0,778151
banget	1	0,778151	0,778151
Sih	1	0,778151	0,778151

Tabel 20 Contoh Hasil Nilai TF-IDF Marah

Kata	Nilai TF	Nilai IDF	Nilai TF-IDF
Bisa	1	0,778151	0,778151

Membuat	1	0,778151	0,778151
Rewind	1	0,778151	0,778151
Indonesia	1	0,778151	0,778151
2022	1	0,778151	0,778151
Kah	1	0,778151	0,778151

Tabel 21 Contoh Hasil Nilai TF-IDF Sedih

Kata	Nilai TF	Nilai IDF	Nilai TF-IDF
Dosa	1	0,778151	0,778151
kah	1	0,778151	0,778151
kita	1	0,778151	0,778151
Punya	1	0,778151	0,778151
kulit	1	0,778151	0,778151
Yang	1	0,778151	0,778151
Hitam	1	0,778151	0,778151

Tabel 22 Contoh Hasil Nilai TF-IDF Takut

Kata	Nilai TF	Nilai IDF	Nilai TF-IDF
setelah	1	0,778151	0,778151

menonton	1	0,778151	0,778151
ini	1	0,778151	0,778151
Saya	2	0,477121	0,954242
takut	1	0,778151	0,778151
ditinggal	1	0,778151	0,778151
orang	1	0,778151	0,778151
Tersayang	1	0,778151	0,778151
yakni	1	0,778151	0,778151
ibu	1	0,778151	0,778151
Sedih	1	0,778151	0,778151

Tabel 23 Contoh Hasil Nilai TF-IDF Terkejut

Kata	Nilai TF	Nilai IDF	Nilai TF-IDF
1/3	1	0,778151	0,778151
video	1	0,778151	0,778151
Cuma	1	0,778151	0,778151
Joget	1	0,778151	0,778151

dan	1	0,778151	0,778151
lu	1	0,778151	0,778151
meneteskan	1	0,778151	0,778151
air	1	0,778151	0,778151
Mata	1	0,778151	0,778151

4.7 Hasil Confusion Matrix

Tahap Confusion matrix bertujuan mengetahui tingkat akurasi dari proses klasifikasi yang telah dikerjakan oleh sistem yang mengacu pada nilai akurasi. Pada penelitian ini, penulis melakukan satu kali percobaan dengan perbandingan antara data training dan data testing sebesar 90:10%, 80:20%, 70:30% dan 60:40. Pada tahap confusion matrix ini menggunakan kernel Linear. Pada implementasi kernel linear, penulis menggunakan perhitungan akurasi, presisi dan recall. Adapun hasil perhitungan confusion matrix dengan kernel linear SVC dapat dilihat pada gambar 22 dan 23.

	precision	recall	f1-score	support
bahagia	0.81	0.97	0.88	296
jijik	1.00	0.33	0.50	3
marah	0.73	0.40	0.52	20
sedih	0.57	0.32	0.41	53
takut	1.00	0.20	0.33	5
terkejut	0.40	0.09	0.14	23
accuracy			0.79	400
macro avg	0.75	0.38	0.46	400
weighted avg	0.76	0.79	0.75	400

Gambar 22 Nilai akurasi Presisi dan Recal 80:20%



Gambar 23 Hasil Confusion Matrix 80:20%

4.8 Hasil Analisis Emosi

Evaluasi analisis emosi dilakukan agar dapat mengetahui emosi dari setiap komentar yang telah di uji. Proses tahap analisis emosi dilakukan dengan hasil data testing yang dilakukan prediksi oleh sistem sehingga data komentar pada data testing muncul emosi tiap komentar. Hasil analisis emosi pada perbandingan tersebut dapat dilihat pada tabel 23:

Tabel 7 Hasil Uji Analisis Emosi

No	Data komentaremosi	data_emosi_baru
1	hah ? Kalimat mana yang jelekin karya orang ?	['bahagia']
2	iya jatohnya yang bikin sensasi yang viral orangnya tetep dimasukin, yang penting viral 😊.	['marah']
3	Bangga banget bisa menikmati video rewind 2021 ini	['bahagia']
4	Lah malah ini lebih keren cuyyy, beda tema sm konsep Sm yg 2020 lalu yg lbh Fun, coba liat deh ini lbh banyak Kesan & Pesan moralnya, jgn cm diliat sekilas doang tapi pahami jg makna nya...	['bahagia']

5	"coba konten indo kaya gini semua" dari sini aja gw ngerti dek menjurus ke konten rewind	['bahagia']
6	2022 mengangkat issue terhot polisi tembak polisi	['bahagia']
7	Ada abang henokh wkwk,nostalgia banget liat konten jawaban lucu anak sd	['bahagia']
8	Aku berharap bnget sih bakal ada scene yang menyiratkan olimpiade paralimpiade dan Thomas cup	['sedih']
9	anjay 😊😊😊 aku pun serasa nostalgia	['sedih']
10	Artinya sebuah perjalanan bersama YouTube bukan sekedar merekam setiap moment menjadi sebuah video yang bisa menjadi cuan, namun bersama YouTube kita bisa memberi manfaat positif lagi buat sekitar, jika tujuannya tak hanya fokus di cuan.	['bahagia']
11	Bagus banget syiii 🐱 di tunggu rewind 2022 nya ❤️❤️	['bahagia']
12	Baru liat sekarang tapi ini keren banget sampe bikin nangis wah gila si .	['bahagia']
13	Baru liat sekarang tapi ini keren bgt sampe bikin nangis wah gila si 😊❤️	['bahagia']
14	baru liat, makna ceritanya dalem banget!!! semoga yg nonton bisa sama2 ngerti kondisi negeri kita yg sekarat. sukses terus berkarya influencer Indonesia	['bahagia']
15	Baru nonton. Durasi bikin puuaaass, memanjakan mata dan telinga. Part terakhir selalu mengharukan ... GOOD JOBB	['bahagia']
16	beneran nangis ?	['bahagia']
17	Bentar lagi komen mu dihapus sama si lord Liow wmwmm	['bahagia']

18	Baru nonton. Durasi bikin puas, memanjakan mata dan telinga. Part terakhir selalu mengharukan ... Good Job	['bahagia']
19	Bikin mewek smpe akhir, sumpah keren abis🤔,trima kasih untk karya karyanya👏	['bahagia']
20	GILA INI KEREN BANGET MAU NANGIS IM REALLY PROUD OF YOU KAKAK KAKAK YOU , ALL THE BEST LOVE IT SO MUCH !!!	['bahagia']
21	bisa bedain sinematografi sama visual effect gasi?wkwkw kalo visual effect 2020 kalo sinematografi ini jauh lebih lebih bagus	['sedih']
22	Gila, this is masterpiece, dari awal cerita smpai habis benar2 kita terbawa perasaan duka dan sedih beda sama yang 2020 perasaan yang dicampur aduk, konsepnya tertata rapi, benar2 konsepnya diambil dari pengalaman Chandra saat terkena covid , 2nd the best setelah Rewind 2020, kerja bagus buat Para Youtubers disana	['bahagia']
23	Btw ga nonton full kah?	['bahagia']
24	buat aulion,chandra liow,step by step id,jovial, dan ferry irwandi, YOU GUYS ROCK! kalian sangat berani dan kreatif pake tema ala2 hutan gini. pasti ribet bgt nyiapin properti,editing,dan macem2nya. aku hargai bgt kerja keras kalian. yg ditunggu2 akhirnya hari ini dateng juga👏👏👏	['bahagia']
25	Catur juga, dritadi gw liat catur mulu 🤔	['marah']
26	channel nya ganti Indonesian Youtubers & Tiktokers 🤔	['bahagia']
27	coba konten indo kayak gini semua, jadi ikutan semangt lagi kayak dulu terharu bnget lihat nyaaaa makasi bang chandra dan bang aulion	['bahagia']
28	ini namanya karya luar biasa kerja bagus semuanya di tunggu tahun ini meledak lagi	['bahagia']
29	dek dek kalau ngetik harus tepat kepointnya	['bahagia']

30	Demi apaa bagus banget, aku terharu, endingnya bikin nangis 😭😭😭	['bahagia']
31	Di video ini jg udh dikemas sedemikian mungkin terutama di dialog yg bareng Mba Nana, kenapa ga diangkat secara blak blakan ya karna terlalu Frontal dan Membekas, yg ada malah membuka luka lama	['sedih']
32	Gak sabar lihat next REWIND INDONESIA 2022..❤️	['bahagia']
33	gangerti lagi waktu nonton rewind iniii !! Semua campur aduk sedih,senang,kagum,bangga KEREN BANGET POKOKNYA !!! 🤩🤩🤩	['bahagia']
34	gua gak nyangka kaget banget ada Helm hiu intm c2	['terkejut']
35	Gilaaak hebat banget, Kak, maduin semua yang terjadi/pernah viral di taun 2021 jadi satu... Keren asli👏👏👏	['bahagia']
36	Gw kira konsepnya seperti tahun lalu yg peduli antar sesama atau ttg covid yg gak ada habis-habisnya eh ternyata biasa aja malah gw bingung ini ceritanya lari kemana kok hampa banget, padahal udah nunggu-nunggu alur ceritanya😞😞🤔	['marah']
37	Ending bikin merinding	['bahagia']
38	Hebat salut yg ikut berpartisipasi ini video kaya nya ngena banget banyak pesan moral nya, selain adegan yg musik viral makanan viral joged viral , ada juga tragedi dan solusi best 👏👏👏❤️❤️ love indonesia ❤️	['bahagia']
39	Ikoy-Ikoyan legend sekali 🔥🔥	['bahagia']
40	Ini menurut gua udah cukup. Ga lebay, ga kurang.	['marah']
41	Iya saya rasa ada yg kurang.. Video 19detik ga di cantumin padahal heboh bgt ya ngab haha.. Sama drama kazakstan tuh ya kan	['sedih']

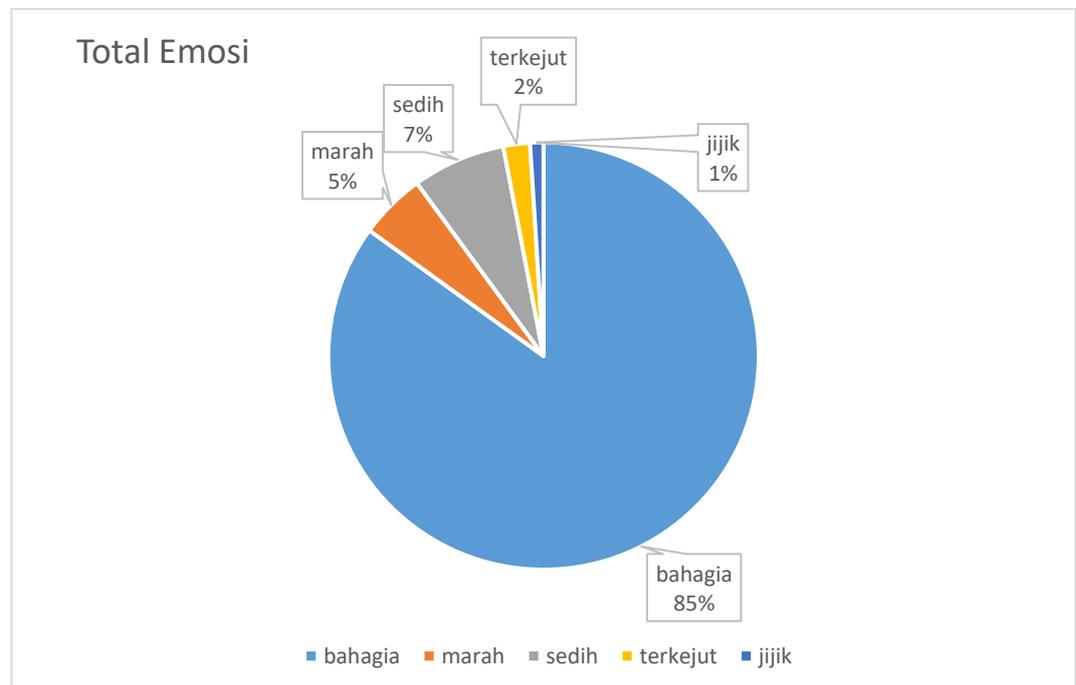
42	Ini yang bikin aku makin cinta dengan Indonesia, banyak banget insan2 muda kreatifnya, there's so much to explore about apalagi selama 2021, pandemi covid-19 tidak menjadi penghalang, malah banyak banget bermunculan karya anak2 muda Indonesia yang membanggakan, yang bikin gua bersyukur jadi bagian dari Indonesia	['bahagia']
43	jijik banget ini koreonya jijik ih gajelas	['jijik']
44	Kalo ada paling fuji,ghozali,timnas,noah yg masuk kan lagi viral bgt	['bahagia']
45	kalo Indonesian creator sangat banyak yakin bisa di tampung 😊	['bahagia']
46	kan tokyo olimpics masuk (gatau ejaan yg bener, maaf huehue)	['bahagia']
47	Keren banget saya aja merinding ini keren banget gila pesan moral dll nya banyak bgt bagus plis harus support bgt Indonesia biar jadi negara anti rasis dan semoga sembuh 🙏🙏	['bahagia']
48	Keren banget the best si ini video	['bahagia']
49	Keren banget. Pesannya sampai ke hati. ❤️❤️	['bahagia']
50	Keren bangettt ahhh bangga banget aku jadi anak indonesia I LOVE YOU INDONESIA	['bahagia']
51	KEREN BANGETTT!! pesan yg aku dapet adlh:	['bahagia']
52	Bangsa kita Indonesia butuh kita sbg anak2 muda utk sama2 berjuang memberi harapan. Anak2 muda itu punya energi yg besar, punya kreativitas yg luarbiasa buat kerjain hal2 baik utk banggain bangsa ini.	['bahagia']
53	Kerenn, semangat para kreator konten dgn memberikan konten konten yg positif yg guys, jadi RI satu kesatuan yg TDK dapat dipisahkan, maju teruss dgn hasil hasil yg positif, spirittt 🙌	['bahagia']

54	Kerennnn jarang2 dan sangat sedikit yg bikin konten seperti ini 👍👍👍 makin sukses pokoknya	['bahagia']
55	klo bagus semua mana ada monontn nya cuk, kan jadi pembelajar klo jelek langsung lewat	['bahagia']
56	klo kek gini semua, uploadnya setaun sekali ntar:v	['bahagia']
57	kok lah ? Emang kenyataan gitu kalau mampu ya silahkan bikin sendiri	['sedih']
58	konten bagus tu bukan berarti konten nya harus sama tapi kualitas nya yang di perbaiki	['bahagia']
59	Aku kaget banget loh ada windah	['terkejut']
60	Karya yang luar biasa salam Indonesia Youtuber	['bahagia']
61	lebih ke Indonesian Creator sih	['bahagia']
62	Keren bang bagus penasaran yang 2022 nanti	['bahagia']
63	Keren banget gak nyangka dalam situasi pandemi kaya gini bisa bikin rewind sekeren ini	['bahagia']
64	Menurutku setiap lirik yg di pilih itu ada arti dan masuk k konsep yg d maksud.di pintu itu(inet) d tawarkan dan d tampilkan banyak hal, uang, cewe, kesempatan, dan show dance, makan dll. Dan tragedi yg terjadi sdah d masukkan dalam lukisan. Tragedi besarnya menurutku lebih ke covid ke orang n negara sbgai ibu, dan anak as children dan warga.	['sedih']
65	Virus yg menyerg tubuh, bsa jdi org yg mencari kesempatan keuangan d saat sdg crisis.	['bahagia']
66	monoton dari mana, editan super keren gini, ada kerja keras yg keluar	['bahagia']
67	Nama nya juga youtube rewind ya yg viral di youtube	['bahagia']

68	Keren banget Nih , Rewind Tidak Pernah Mengecewakan	['bahagia']
69	nih cuk kalau ada konten bagus harus ada konten jelek kalau semua konten bagus jadi g bakal wow	['bahagia']
70	NONTONNYA SAMBIL NANGISSSSS 😭😭 keren banget karya anak INDONESIA 🤔🤔🤔🤔🤔🤔🤔 bangga jadi anak Indoonesiaaaaa	['bahagia']
71	Keren banget.....Maju terus Indonesia	['bahagia']
72	point nya konten konten di indo itu harus konten yang berkualitas udah gitu doing	['bahagia']
73	rada heran gitu ga sih? pas baru pertama liat? cerita nya indo banget, tapi gw yang liat lebih condong ke culture barat nya wkwk 🤔🤔	['bahagia']
74	Rewind berantakan sepanjang masa konsepnya apa gw masih bingung Jalan cerita ngga dapat kebanyakan joget dan drama lebay.. kalau lebih kreatif nama nama Youtuber dicantumkan biar lebih hidup..tapi ini kebanyakan gw gak kenal semuanya lebih ke Tiktok rewind kali 😭. Maaf tahun ini nggak.	['marah']
75	Rewind kali ini rasanya kyk tiktoker rewind ya bang 🤔	['bahagia']
76	Sambo termasuk juga nggak?	['bahagia']
77	Saran nama Channelnya ganti jadi Indonesian Youtubers & Tiktokers 🤔🤔🤔🤔🤔	['bahagia']
78	Saran untuk tahun depan, perbanyak issue yang diangkat "YANG HAPPENING DI 2022". Jadi ga cuma fokus ke lagu viral, dance viral. Tapi dengan pencapaian Indonesia, tragedi di Indonesia dll.	['bahagia']
79	sedih thn 2021 ini bg , banyak musibah .. semoga tahun 2022 di awali dengan hal2 yang baik	['bahagia']
80	Selalu bikin merinding setiap taunnya	['bahagia']

81	Selalu nangissss... makin cinta indonesia, makin ingin terlibat untuk memajukan bangsa. ❤️	['bahagia']
82	Seni Teater terbaik yg pernah Guaa tontonnn.. merinding sumpahh..bangga banga	['bahagia']
83	Setelah 8 bulan berlalu pun,ini masih menggetarkan hati,semoga tahun 2022 ini bisa jadi lebihbaik	['bahagia']
84	sinematogeafinya udh bagus kok, cuma banyak blink2nya bikin puyeng wkwkw	['bahagia']
85	Sinematografi-nya, Koreografi-nya, KELAAAAAS!!! KALIAN KEREEEEEN PARAAHH!! 🔥🔥🔥	['bahagia']
86	smngat ka untuk msuk rewind 2022 :(['bahagia']
87	Suka banget!!! soalnya ada subtitlenyaa. Jadi berguna bgt untuk teman-teman tuli semuanya 🥰	['bahagia']
88	Sumpah rewind nya keren banget salut dah pokok nya!	['bahagia']
89	Super kerenn. Kalian luar biasa...acting memukau..sukses dan jaya selalu	['bahagia']
90	Telat 15 hari buat nonton ini, tapi merindingnya masih dapet bgt	['bahagia']
91	Telat nontonnya tapi nangis dr awl smpai akhir.. terharu luar biasaaaaaa	['bahagia']
92	Terharu aku. Keren bangeett 🔥🥰	['bahagia']
93	Terima kasih anak2 muda dengan positive vibes dan optimisme tinggi.. 🔥🔥🔥🔥🔥	['bahagia']
94	Terima kasih Orang Indonesia,	['bahagia']

95	Karya Mu Luar Biasa id	['bahagia']
96	Terimakasih atas karyanya yg luar biasa!!! Bangga menjadi bagian dari Indonesia khususnya industri content creating. Semoga banyak tren seru dan positif di 2022!!!	['bahagia']
97	Terimakasih sudah memberikan karya yg luar biasa... Benar2 kagum..	['bahagia']
98	Tiktokers Rewind sih	['bahagia']
99	Udah balance ini bang, tragedi ada copid, pesawat, kapal selam, kebakaran, bansos. 🙏	['bahagia']
100	Like komen aku bang biar mama aku bangga😊😊	['bahagia']



Gambar 24 Hasil Total Analisis Emosi

Berdasarkan hasil dari tabel 23 didapatkan bahwa uji analisis emosi dari komentar yang mengandung emosi bahagia didapatkan sebanyak 85 komentar, komentar yang mengandung emosi marah didapatkan sebanyak 5 komentar,

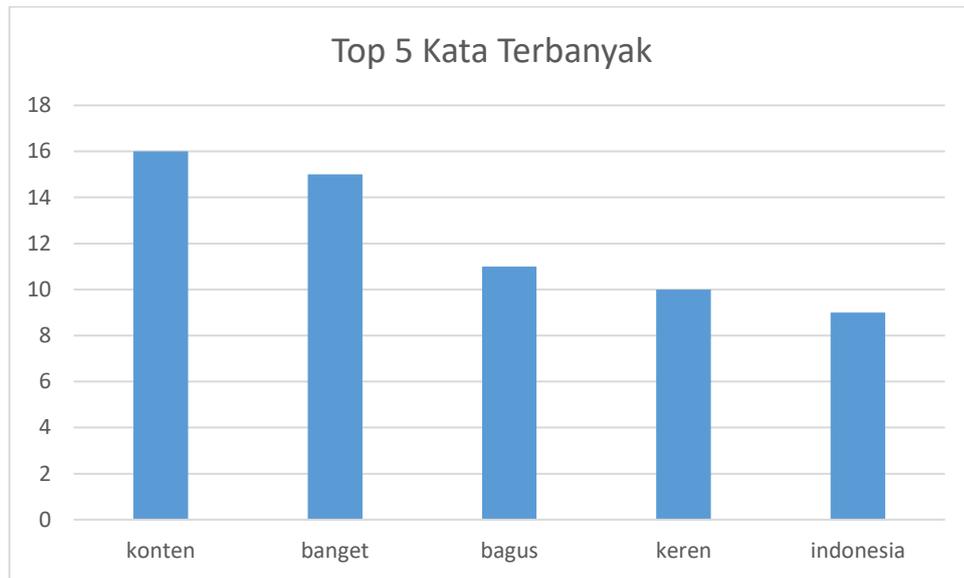
komentar yang mengandung emosi sedih didapatkan sebanyak 7 komentar, komentar yang mengandung emosi jijik didapatkan sebanyak 1 komentar, komentar yang mengandung emosi terkejut didapatkan sebanyak 2 komentar dan komentar yang mengandung emosi takut didapatkan sebanyak 0 komentar. Sehingga dari hasil uji analisis emosi tersebut rata - rata komentar yang dituliskan mengandung emosi bahagia.

Berdasarkan hasil emosi bahagia sebanyak 85 komentar terdapat 7 komentar yang menurut peneliti hasil emosi tidak sesuai dari yang didapatkan oleh sistem. Berikut merupakan analisis emosi dari komentar yang tidak sesuai dapat dilihat pada tabel 24.

Tabel 8 Analisis Emosi Komentar yang tidak sesuai

No	Data koementaremosi	data_emosi_baru	Alasan
1	beneran nangis ?	['bahagia']	Komentar tersebut menurut sistem menghasilkan emosi bahagia karena terdapat kata beneran dan nangis yang memiliki arti terharu, sesuai dengan rata - rata contoh dataset yang mengandung kata nangis sehingga memiliki arti terharu, sedangkan menurut peneliti komentar tersebut lebih pantas menghasilkan emosi sedih.
2	kalo Indonesian creator sangat banyak yakin bisa di tampung 😊	['bahagia']	Komentar tersebut menurut sistem menghasilkan emosi bahagia karena terdapat kata Indonesian creator sangat banyak, sesuai dengan rata-rata kata tersebut dalam dataset yang memiliki arti bahagia, sedangkan menurut peneliti komentar tersebut lebih pantas menghasilkan emosi sedih.

3	klo bagus semua mana ada monontn nya cuk, kan jadi pembelajar klo jelek langsung lewat	['bahagia']	Komentar tersebut menurut sistem menghasilkan emosi bahagia karena terdapat kata bagus semua mana ada monontn nya, sesuai dengan rata-rata kata tersebut dalam dataset yang memiliki arti bahagia, sedangkan menurut peneliti komentar tersebut lebih pantas menghasilkan emosi marah.
4	rada heran gitu ga sih? pas baru pertama liat? cerita nya indo banget, tapi gw yang liat lebih condong ke culture barat nya wkwk 😊🤔	['bahagia']	Komentar tersebut menurut sistem menghasilkan emosi bahagia karena terdapat kata pas baru pertama liat ceritanya indo banget, wkwk, dan terdapat emoji bahagia, sesuai dengan rata-rata kata tersebut dalam dataset yang memiliki arti bahagia, sedangkan menurut peneliti komentar tersebut lebih pantas menghasilkan emosi sedih
5	Rewind kali ini rasanya kyyk tiktoker rewind ya bang 🤔	['bahagia']	Komentar tersebut menurut sistem menghasilkan emosi bahagia karena terdapat kata tiktoker rewind, sesuai dengan rata-rata kata tersebut dalam dataset yang memiliki arti bahagia, sedangkan menurut peneliti komentar tersebut lebih pantas menghasilkan emosi sedih
6	smngat ka untuk msuk rewind 2022 :(['bahagia']	Komentar tersebut menurut sistem menghasilkan emosi bahagia karena terdapat kata smngat, msuk, rewind 2022, sesuai dengan rata-rata kata tersebut dalam



Gambar 26 5 Kata Terbanyak pada Wordcloud

BAB 5 KESIMPULAN DAN SARAN

Bab ini menjelaskan tentang kesimpulan dan saran yang diberikan oleh peneliti berdasarkan penelitian yang telah dilakukan. Kesimpulan dan saran dapat digunakan sebagai masukan untuk penelitian selanjutnya.

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan dapat ditarik kesimpulan sebagai berikut :

1. Berdasarkan hasil pengujian model menggunakan confusion matriks menghasilkan nilai akurasi yang tinggi sebesar 79% dengan model split data 80:20%, sedangkan menghasilkan nilai akurasi yang rendah sebesar 77% dengan model split data 90:10%.
2. Berdasarkan hasil analisis emosi didapatkan bahwa uji analisis emosi dari komentar yang mengandung emosi bahagia didapatkan sebanyak 85 komentar, komentar yang mengandung emosi marah didapatkan sebanyak 5 komentar, komentar yang mengandung emosi sedih didapatkan sebanyak 7 komentar, komentar yang mengandung emosi jijik didapatkan sebanyak 1 komentar, komentar yang mengandung emosi terkejut didapatkan sebanyak 2 komentar dan komentar yang mengandung emosi takut didapatkan sebanyak 0 komentar. Sehingga dari hasil uji analisis emosi tersebut rata - rata komentar yang dituliskan mengandung emosi bahagia.

5.2 Saran

Jumlah emosi bahagia, marah, sedih, takut, terkejut, dan jijik yang tidak seimbang dalam data training juga menjadi sebuah kendala dalam penelitian ini. Jumlah yang tidak seimbang ini mempengaruhi kinerja sistem dalam memprediksi emosi. Untuk kedepannya dalam mengatasi masalah ini dapat dilakukan dengan mencari dataset yang memiliki jumlah emosi yang sama.

DAFTAR PUSTAKA

- Agastya, W., & Aripin. (2020). Pemetaan Emosi Dominan pada Kalimat Majemuk Bahasa Indonesia Menggunakan Multinomial Naïve Bayes. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 9(2), 171–179. <https://doi.org/10.22146/jnteti.v9i2.157>
- Ainurrochman, Adi, D. P., & Gumelar, A. B. (2020). Deteksi Emosi Wicara pada Media On-Demand menggunakan SVM dan LSTM. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(5), 799–804. <https://doi.org/10.29207/resti.v4i5.2073>
- Alita, D., Fernando, Y., & Sulistiani, H. (2020). Implementasi Algoritma Multiclass Svm Pada Opini Publik Berbahasa Indonesia Di Twitter. *Jurnal Tekno Kompak*, 14(2), 86. <https://doi.org/10.33365/jtk.v14i2.792>
- Anandarajan, M., Hill, C., & Nolan, T. (2019). *Text Preprocessing* (pp. 45–59). https://doi.org/10.1007/978-3-319-95663-3_4
- Ardiada, I. M. D., Sudarma, M., & Giriantari, D. (2019). Text Mining pada Sosial Media untuk Mendeteksi Emosi Pengguna Menggunakan Metode Support Vector Machine dan K-Nearest Neighbour. *Majalah Ilmiah Teknologi Elektro*, 18(1), 55. <https://doi.org/10.24843/mite.2019.v18i01.p08>
- Arini, A.-, Wardhani, L. K., & Octaviano, D.-. (2020). Perbandingan Seleksi Fitur Term Frequency & Tri-Gram Character Menggunakan Algoritma Naïve Bayes Classifier (Nbc) Pada Tweet Hashtag #2019gantipresiden. *Kilat*, 9(1), 103–114. <https://doi.org/10.33322/kilat.v9i1.878>
- Deolika, A., Kusriani, K., & Luthfi, E. T. (2019). Analisis Pembobotan Kata Pada Klasifikasi Text Mining. *Jurnal Teknologi Informasi*, 3(2), 179. <https://doi.org/10.36294/jurti.v3i2.1077>
- Indraloka, D. S., & Santosa, B. (2017). Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia. *Jurnal Sains Dan Seni ITS*, 6(2), 6–11. <https://doi.org/10.12962/j23373520.v6i2.24419>
- Irawan, D., Perkasa, E. B., Yurindra, Y., Wahyuningsih, D., & Helmud, E. (2021). Perbandingan Klasifikasi SMS Berbasis Support Vector Machine, Naive Bayes Classifier, Random Forest dan Bagging Classifier. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 10(3), 432–437. <https://doi.org/10.32736/sisfokom.v10i3.1302>
- Julianto, I. T. (2022). Analisis Sentimen Terhadap Sistem Informasi Akademik Institut Teknologi Garut. *Jurnal Algoritma*, 19(1), 449–456. <https://doi.org/10.33364/algoritma/v.19-1.1112>
- Nello, C., & Shawe, T. J. (n.d.). *An Introduction to Support Vector Machine and other kernel-based learning method*.
- Rohman, A. N., Handayani, R. D., Y. P., R. D., & Kusriani, K. (2021). Deteksi Emosi Media Sosial Menggunakan Term Frequency- Inverse Document Frequency. *CSRID (Computer Science Research and Its Development Journal)*, 11(3), 140. <https://doi.org/10.22303/csrid.11.3.2019.140-148>
- Rustiana, D., & Rahayu, N. (2017). ANALISIS SENTIMEN PASAR OTOMOTIF MOBIL: TWEET TWITTER MENGGUNAKAN NAÏVE BAYES. *Jurnal SIMETRIS*, 8.
- SINTA. (n.d.-a). *No Title*. Majalah Ilmiah Teknologi Elektro. Retrieved July 8, 2022, from

<https://sinta.kemdikbud.go.id/journals/profile/2539>

SINTA. (n.d.-b). *No Title*. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi). Retrieved July 8, 2022, from <https://sinta.kemdikbud.go.id/journals/profile/3999>

SINTA. (n.d.-c). *No Title*. CSRID (Computer Science Research and Its Development Journal). Retrieved July 8, 2022, from <https://sinta.kemdikbud.go.id/journals/profile/159>

Sya'bani, M. M., & Umilasari, R. (2018). Penerapan Metode Cosine Similarity dan Pembobotan TF/IDF pada Sistem Klasifikasi Sinopsis Buku di Perpustakaan Kejaksaan Negeri Jember. *JUSTINDO (Jurnal Sistem Dan Teknologi Informasi Indonesia)*, Vol 3, No 1 (2018): JUSTINDO, 31–42. <http://jurnal.unmuhjember.ac.id/index.php/JUSTINDO/article/view/2345>

Terkenal.co.id. (2022). *Video Rewind Indonesia 2021 Tuai Pro Kontra, Ini Alasannya*. PT. Terkenal Media Indonesia. <https://terkenal.co.id/read/news/6562/video-rewind-indonesia-2021-tuai-pro-kontra-ini-alasannya/>

Viny Christanti, M., Walla, & Sutrisno, T. (2020). Comments scraping application for review YouTube content. *IOP Conference Series: Materials Science and Engineering*, 852(1), 0–6. <https://doi.org/10.1088/1757-899X/852/1/012167>

Widayani, W., & Harliana, H. (2021). Analisis Support Vector Machine Untuk Pemberian Rekomendasi Penundaan Biaya Kuliah Mahasiswa. *Jurnal Sains Dan Informatika*, 7(1), 20–27. <https://doi.org/10.34128/jsi.v7i1.268>

Youtube Rewind Indonesia. (2022). Wikipedia:Enslikipedia Bebas. https://id.wikipedia.org/wiki/YouTube_Rewind_Indonesia?tableofcontents=0

Zuhad, F., & Wilantika, N. (2022). Perbandingan Penggunaan Kamus Normalisasi dalam Analisis Sentimen Berbahasa Indonesia. *Jlk*, 5(1), 13–23.

DAFTAR LAMPIRAN

A. Daftar Referensi Jurnal

Judul Artikel	Jenis	Publisher	Akreditasi (Indeks)	Keterangan
Text Mining pada Social Media untuk Mendeteksi Emosi Pengguna Menggunakan Metode Support Vector Machine dan K-Nearest Neighbour	Jurnal	Majalah Ilmiah Teknologi Elektro	SINTA 3	Penelitian terdahulu (Bab 2)
Deteksi Emosi Wicara pada Media On-Demand menggunakan SVM dan LTSM	Jurnal	Jurnal RESTI	SINTA 2	Penelitian terdahulu (Bab 2)
Deteksi Emosi Media Sosial Menggunakan Term Frequency-Inverse Document Frequency	Jurnal	Jurnal CSRID	SINTA 3	Penelitian terdahulu (Bab 2)
Support Vector Machines and other kernel-based learning methods	Buku	-		Tinjauan Pustaka (Bab 2)