

**ANALISIS SENTIMEN PEMBELAJARAN ONLINE
PADA MEDIA YOUTUBE MENGGUNAKAN NAIVE
BAYES**

Oleh

NAZMIA DJIBU

T3116001

SKRIPSI

**Untuk memenuhi salah satu syarat ujian guna
memperoleh gelar sarjana**



**PROGRAM SARJANA
TEKNIK INFORMATIKA
UNIVERSITAS ICHSAN GORONTALO
2023**

HALAMAN PERSETUJUAN

**ANALISIS SENTIMEN PEMBELAJARAN ONLINE
PADA MEDIA YOUTUBE MENGGUNAKAN NAÏVE
BAYES**

Oleh

NAZMIA DJIBU

T3116001

SKRIPSI

Untuk memenuhi salah satu syarat ujian guna memperoleh gelar
Sarjana Program Studi Teknik Informatika,
ini telah disetujui oleh tim pembimbing
Gorontalo, April 2023

Pembimbing I

Pembimbing II

Yasin Aril Mustofa, M.Kom

NIDN :0926088503

Sudirman S. Panna, M.kom

NIDN : 0924038205

PERNYATAAN SKRIPSI

Dengan ini saya menyatakan bahwa :

1. Karya tulis (skripsi) saya ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (Sarjana) baik di Universitas Ichsan Gorontalo maupun diperguruan tinggi lainnya.
2. Karya tulis (skripsi) saya adalah murni gagasan, rumusan, dan penelitian saya sendiri tanpa bantuan pihak lain, kecuali arahan dari tim pembimbing.
3. Dalam karya tulis (skripsi) saya ini tidak terdapat karya atau pendapat yg telah dipublikasikan orang lain, kecuali secara tertulis dicantumkan sebagai acuan/situasi dalam naskah dan dicantumkan pula dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan ketidakbeneran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya tulis ini, serta sanksi lainnya sesuai dengan norma-norma yang berlaku di Universitas Ichsan Gorontalo.

Gorontalo, April 2023

Yang membuat pernyataan

Penulis

KATA PENGANTAR

Alhamdulillah penulis dapat menyelesaikan skripsi ini dengan judul **“ANALISIS SENTIMEN PEMBELAJARAN ONLINE PADA MEDIA YOUTUBE MENGGUNAKAN NAIVE BAYES”**, untuk memenuhi salah satu syarat Ujian Akhir guna memperoleh gelar sarjana komputer pada Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo.

Penulis menyadari sepenuhnya bahwa skripsi ini tidak mungkin terwujud tanpa bantuan dan dorongan dari berbagai pihak, baik bantuan moril maupun materil. Untuk itu, dengan segala keikhlasan dan kerendahan hati, penulis mengucapkan banyak terima kasih dan penghargaan yang setinggi-tingginya kepada :

1. Dr. Yuriko Abdusamat, M.si, selaku ketua yayasan Pengembangan Ilmu Pengetahuan dan Teknologi (YPIPT) Ichsan Gorontalo
2. Dr, Abdul Gaffar La Tjokke, M.si, selaku Rektor Universitas Ichsan Gorontalo;
3. Irvan Abraham Salihi M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Ichsan Gorontal
4. Sudirman Melangi, M.kom, selaku Wakil Dekan I Bidang Akademik Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
5. Irma Surya Kumala Idris, M.Kom, selaku Wakil Dekan II Bidang Administrasi Umum dan Keuangan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo.
6. Sudirman S. Panna, M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo sekaligus selaku Pembimbing II yang telah memberikan bimbingan dan arahan kepada penulis untuk menyelesaikan skripsi ini
7. Yasin Aril Mustofa, M.Kom, selaku Pembimbing I, yang telah memberikan bimbingan dan arahan kepada penulis untuk menyelesaikan skripsi ini ;

8. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang telah mendidik dan mengajarkan berbagai disiplin ilmu kepada penulis;
9. Teristimewa kepada keluarga saya , Ayah saya Ma'mun Djibu, Ibu saya Maslin Nupu, Kakak saya Fatma, Djbu S.pd, dan Nur'ain Djibu, S.tr.Keb terimakasih atas segala kasih sayang jerih payah, doa restu serta telah mendukung penulis mencapai cita-cita untuk menjadi seorang sarjana;
10. Rekan – rekan seperjuangan Fikom angkatan 2016
11. Kepada semua pihak yang telah membantu dalam penyelesaian skripsi ini tak sempat penulis sebutkan satu-persatu.
12. Terakhir , saya ingin mengucapkan terima kasih kepada diri saya sendiri karena sudah sabar, tetap bertahan, dan selalu sehat dalam melalui proses panjang perkuliahan Teknik Informatika hingga penyelesaian skripsi ini.

Semoga Allah SWT melimpahkan balasan atas jasa-jasa mereka kepada kami. Penulis menyadari sepenuhnya bahwa apa yang telah dicapai ini masih jauh dari kesempurnaan dan masih banyak terdapat kekurangan. Oleh karena itu, penulis sangat mengharapkan adanya kritik dan saran yang konstruktif. Akhirnya penulis berharap semoga hasil yang telah dicapai ini dapat bermanfaat bagi kita semua, Aamiin.

Gorontalo, April 2023

Penulis

DAFTAR ISI

HALAMAN JUDUL	ii
PERSETUJUAN SKRIPSI.....	vi
PENGESAHAN SKRIPSI.....	iii
PERNYATAAN SKRIPSI.....	iv
ABSTRACT.....	iv
ABSTRAK.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2. Identifikasi Masalah.....	2
1.3. Rumusan Masalah.....	2
1.4. Tujuan Penelitian.....	3
1.5. Manfaat Penelitian.....	3
BAB II LANDASAN TEORI.....	4
2.1. Tinjauan Studi.....	4
2.2. Tinjauan Pustaka.....	6
2.2.1 Youtube.....	6
2.2.2 Analisis Sentimen.....	7
2.2.3 Term Frequency(TF).....	7
2.2.4 Klasifikasi.....	7
2.3. Kerangka Pikir.....	17
BAB III METODE PENELITIAN.....	18
3.1 Jenis, Metode, Subjek, Objek, Waktu, dan Lokasi Penelitian.....	18

3.2	Metode Penelitian	19
3.3	Tahap Pemodelan	19
3.3.1	Data Komentar	20
3.3.2	Tahap Crawling	20
3.3.3	Tahap Preprocessing	20
3.3.4	Tahap Labelisasi	20
3.3.5	Tahap Pemberian Bobot	20
3.3.6	Tahap Klasifikasi	21
3.3.7	Tahap Evaluasi	21
BAB IV HASIL PENELITIAN		22
4.1	Hasil Pengumpula Data	22
4.1.1	Penerapan Data Training	22
4.1.2	Mencar Probabilitas Pada Data Testing	24
BAB V HASIL DAN PEMBAHASAN		33
5.1	Klasifikasi Naïve Bayes Pada DataTesting	33
BAB VI KESIMPULAN DAN SARAN		39
6.1	Kesimpulan	39
6.2	Saran	39
DAFTAR PUSTAKA		
LAMPIRAN		

DAFTAR TABEL

Tabel 2.1 Tinjauan Studi.....	4
Tabel 2.2 Labelisasi Komentar.....	12
Tabel 2.3 Klasifikasi Sentimen.....	14
Tabel 4.1 Hasil Pengumpulan Data.....	15
Tabel 4.2 Pembagian Data Train Dan Data Test.....	23
Tabel 4.3 Komentar Data Training.....	24
Tabel 4.4 Case Folding.....	24
Tabel 4.5 Tokenization.....	25
Tabel 4.6 Stemming.....	25
Tabel 4.7 Pembobotan TF.....	26
Tabel 4.8 Hasil Pembobotan IDF.....	27
Tabel 4.9 Pembobotan TF-IDF.....	28
Tabel 4.10 Data Train Dan Data Test.....	32
Table 5.1 Daftar Data Testing.....	38
Table 5.2 Nilai Probabilitas	34
Table 5.3 Hasil Pembagian Data Training.....	35

DAFTAR GAMBAR

Gambar 2.3 Kerangka Pikir.....	17
Gambar 3.1 Pemodelan Sentimen Analisis.....	22
Gambar 5.5 Grafik Pembagian Data	37

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pembelajaran online yang dianggap sebagai strategi kemudia menjadi kontroversi karena perbedaan infrastruktu dan kualitas koneksi. Adanya komentar masyarakat karena perubahan mendadak dari pembelajaran tatap muka ke pembelajaran online di indonesia yang diambil dari salah satu channel youtube dan diolah menggunakan algoritma naïve bayes untuk menghasilkan analisis sentimen terhadap komentar masyarakat tentang pembelajaran online.

Penggunaan sosial media atau internet tidak hanya berdampak pada satu orang saja tapi bagi banyak orang. Youtube merupakan salah satusumber data yang isinya konten video yang penyediaannya banyak informasi yang sangat membantu. Informasi yang ada didalam channel youtube adalah berupa video, Opini serta pendapat tertulis dalam komentar yang berupa teks dan tidak terstruktur pada topik yang bersangkutan. Untuk itu diperlukan teknik Analisis yang tepat untuk mengolah sebuah data sehingga mendapatkan hasil yang akurat. Oleh karena itu, teknik yang cocok dan tepat untuk menganalisis suatu opini publik tersebut adalah dengan analisis sentimen.

Analisis Sentimen menerapkan Natural Language Processing (NLP) dan text mining dalam mengidentifikasi dan mengekstrak informasi tentang topik tertentu [1]. Seperti pada penelitian terdahulu yang berjudul Sentiment Analysis on Youtube, *A Brief Survey* menyimpulkan bahwa analisis sentimen pada Youtube dapat digunakan untuk memecahkan masalah seperti kamus sentimen yang terbatas, gaya bahasa yang tidak resmi yang digunakan oleh pengguna dan menetapkan label untuk nilai setiap kata. Analisis sentimen akan memberikan suatu pengambilan infomasi berkualitas tinggi dan komentar-komentar untuk diklasifikasikan. Teknik klasifikasian sentimen bisa dipakai dalam menentukan

polaritas pada setiap komentar tunggal dan kemudian disatukan menjadi rating sentimen.

Untuk mempermudah dalam mengklasifikasikan suatu komentar apakah bersifat positif, negatif atau netral maka penulis menggunakan text mining. Text mining atau pencarian pengetahuan yang berbasis data textual adalah sebuah proses untuk melakukan pencarian pengetahuan yang berfokus kepada data yang berbentuk dokumen atau teks, [2] . Data teks akan diproses menjadi data numerik agar dapat dilakukan proses lebih lanjut. Penelitian ini menjelaskan bahwa menggunakan bayes dan text mining bisa menambahkan akurasi ketepatan pencarian suatu informasi yang kita butuhkan.

Untuk menambahkan akurasi pencarian, penulis juga menggunakan metode bayes. Metode bayes adalah metode yang diciptakan oleh thomas bayes yang dapat digunakan untuk menghitung suatu peluang yang akan terjadi berdasarkan data sebelumnya. Text mining merupakan proses penambangan yang dilakukan oleh komputer dalam mendapatkan sesuatu yang baru, dan tidak diketahui sebelumnya, [3]. Memanfaatkan metode naive bayes untuk mengklasifikasikan data tweet calon gubernur DKI jakarta 2017, menggunakan metode naive bayes dengan nilai rata-rata akurasi mencapai 95% [4].

Berdasarkan latar belakang yang telah dipaparkan diatas, maka dalam penelitian ini penulis mengangkat judul “Analisis Sentimen Pembelajaran Online Pada Media Youtube Menggunakan Naïve Bayes”

1.2. Identifikasi Masalah

Berdasarkan penjelasan latar belakang diatas, dapat disimpulkan bahwa ada beberapa masalah di atas :

1. Sulitnya mengklasifikasikan sebuah informasi yang akurat tapi datanya tidak terstruktur.
2. Belum adanya akurasi dalam menentukan data sentiment terhadap komentar yang terdapat pada kelas positif dan netral ?

1.3. Rumusan Masalah

Berdasarkan penjelasan latar belakang diatas, maka rumusan masalah untuk penelitian ini adalah :

1. Bagaimana kinerja metode naïve bayes dalam mengklasifikasi tingkat sentimen pada kelas positif dan netral ?

1.4. Tujuan Penelitian

Tujuan penelitian merupakan sasaran atau jawaban yang ingin dicapai penulis dalam sebuah penelitian. Oleh sebab itu, tujuan penelitian ini adalah:

1. Untuk mengetahui kinerja metode naïve bayes dalam mengklasifikasikan tingkat sentiment pada kelas positif dan netral.

1.5. Manfaat Penelitian

1. Pengembangan Ilmu

Pada penelitian ini penulis berharap dapat memberikan persembahan dan masukan terhadap pengembangan ilmu pengetahuan dibidang text mining.

2. Praktisi

Dapat menjadi salah satu bahan kajian bagi semua unsur dalam bidang text mining.

3. Peneliti

Sebagai masukan bagi peneliti yang akan melakukan penelitian selanjutnya tentang text mining.

BAB II

LANDASAN TEORI

2.1. Tinjauan Studi

Ada beberapa penelitian yang menjadi landasan teori penulis dalam penelitian ini yaitu :

Tabel 2.1 Tabel Tinjauan Studi

Peneliti	JUDUL	Diskripsi Singkat
Balya , 2019	Analisis sentimen pengguna youtube di Indonesia pada review smartphone menggunakan naïve bayes	Penelitian ini menjelaskan bahwa Dengan melakukan scraping data, analisis sentimen dapat dilakukan secara otomatis dan cepat.
Fajar Ratnawati 2018	Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter	Metode ini diimplementasikan menggunakan python dan library NLTK. Dengan tujuan untuk dapat secara otomatis mengklasifikasikan tweet sebagai sentimen tweet yang positif dan negative. Pada pengujian ini digunakan sekitar 200 tweet positif dan 200 tweet negatif
Adhi Viky Sudiantoro Eri Zuliarso, 2018	Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naïve	Penelitian ini menjelaskan bahwa Naïve Bayes Classifier sangat efektif untuk digunakan sebagai proses klasifikasi yang dibutuhkan dalam sistem

	Bayes Classifier.	analisis sentiment ini dimana nilai yang di dapatkan dalam pengujian sampai 84%.
Samsir, ambiyar Unung Verawardina, firman edi, Ronal watrianthos, 2020	Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes	Penelitian ini menerapkan algoritma Naïve Bayes untuk klasifikasi sentimen. Analisis sentimen dalam study ini menggunakan data Twitter dengan kata kunci 'pembelajaran daring', 'kuliah online' yang difilter dengan kata kunci 'online' dan 'rumah' Penelitian pada periode tersebut menunjukkan 30% sentimen positif, 69% sentimen negatif, dan 1% netral.
Nuari Anisa Sivi	Analisis Sentimen Pada Media Sosial Menggunakan Metode Naive Bayes Classifier	penelitian ini menggunakan format ARFF yang dikumpulkan dari Twitter dan facebook dengan metode manual melihat dari setiap komentar dari media sosial Twitter dan facebook. klasifikasi yang dibangun memiliki kecenderungan untuk mengabaikan minority class. Data dibagi menjadi kata positif 100 kata, kata negatife 100 kata dan kata netral 100 kata.

2.2. Tinjauan Pustaka

2.2.1 Youtube

YouTube merupakan situs berbagi video yang didirikan pada tahun 2005 oleh Chad Hurley, Steve Chen dan Jawed Karim. Situs ini memberikan pengguna untuk mengunggah, menonton, berbagi dan juga menyatakan komentar pada video yang diunggah. Pada Saat mempresentasikan Brandcast youtube ke para pengiklan *Susan Wojcicki* selaku CEO YouTube juga menyatakan jumlah penonton bulanan mencapai sebesar 1,5 milyar pada pertengahan 2017.

2.2.2 Analisis Sentimen

Analisis sentimen adalah bidang yang menarik untuk dikembangkan pada dunia digital sebab saat ini masyarakat secara umum dapat memberikan pendapat juga hasil pemikiran mereka melalui dunia maya dengan bahasa teks dimana para pembaca terkadang memiliki kesalahpahaman. Analisis sentimen adalah proses menentukan sentimen dan mengelompokkan polaritas teks dalam dokumen atau kalimat sehingga kategori dapat ditentukan sebagai sentimen positif, negatif, atau netral [5]. Sentiment Analysis merupakan perpaduan dari teks mining dan data mining, dimana Sentiment Analysis sendiri di mengolah berbagai macam opini yang telah diberikan oleh masyarakat [6]. Dalam teks, sentimen dapat ditangkap pada berbagai tingkatan diantaranya pada tingkat dokumen, paragraf, kalimat, atau klausa [7]. Analisis Sentimen memiliki polaritas (positif atau negatif), sumber (orang atau kelompok orang yang memiliki sentimen), dan target (hal yang menjadi arah tujuan sentimen). Analisis sentimen adalah bidang yang sangat menarik untuk dikembangkan dalam dunia digital sebab pada saat ini masyarakat secara umum mengutarakan perasaan, pendapat dan juga hasil pemikiran mereka melalui dunia maya dengan bahasa teks dimana para pembaca terkadang memiliki kesalahpahaman dalam menterjemahkan sentimen yang terdapat didalamnya [8].

Sentimen positif memberikan suatu nilai yang baik, sentimen negatif memberikan nilai yang buruk pada konteks dalam bentuk teks, sedangkan

sentimen netral memberikan nilai yang tidak berpihak pada baik maupun buruk dan biasanya kalimat dengan sentimen ini tidak mengandung opini di dalamnya. Penelitian dibidang analisis sentimen dari suatu data merupakan hal yang penting dan dibutuhkan di *era big data* seperti saat ini. Analisis sentiment penelitian yang sedang tren di bidang Natural Language Processing (NLP) Analisis sentiment sering menggunakan informasi dari postingan – postingan atau berita yang ada di website media sosial. [9].

2.2.3 Term Frequency(TF)

Pada setiap kata mempunyai tingkat kepentingan yang berbeda dalam dokumen, maka pada penelitian ini sistem pembobotan atau pemberian nilai dilakukan sampai setiap kata tersebut diberikan sebuah indikator. Term frequency merupakan proses pembobotan nilai pada setiap term di dalam komentar yang sudah melewati tahap preprocessing. Term adalah suatu kata atau frase yang dapat dipakai untuk mengetahui konteks dari sebuah dokumen.

2.2.4 Text Mining

Text mining merupakan proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah data tak terstruktur yang nantinya akan diperoleh pola-pola data, tren dan ekstraksi pengetahuan yang potensial dari data teks. Text mining adalah proses untuk memperoleh informasi berkualitas tinggi dari teks. Informasi berkualitas tinggi biasanya didapatkan karena memperhatikan pola dan tren dengan cara mempelajari pola statistik [11].

Text Mining digunakan dalam teknik seperti kategorisasi, identitas ekstraksi dan analisis sentimen yang digunakan untuk mengidentifikasi wawasan dan pola trend dalam volume besar dari data terstruktur. Objek dari teks mining ini bisa berupadokumen yang tidak terstruktur atau semi terstruktur dan Prosesinformasi pada teks mining bisa menghasilkan analisis perasaan yang secara emosional mengidentifikasi pernyataan positif , negatif atau

netral. Text mining yang mempunyai sifat yang biasa dibidang mirip dengan data mining, tapi ada perbedaan seperti text mining lebih fokus ke data yang berbentuk teks dari pada dengan bentuk terstruktur.

2.2.5 Scraping

Scraping adalah proses untuk mengekstrak, mengukur, menganalisis dan memvisualisasikan data dari layar aplikasi menggunakan kode program. Scraping berarti latihan membaca pada data teks dari layar terminal komputer. Dalam bentuknya saat ini, scraping merupakan bagian dari pemrograman yang menghubungkan antara program aplikasi dan antarmuka pengguna. Hal ini dirancang untuk berinteraksi dengan perangkat dan sistem antarmuka sehingga program tampilan yang berupa teks maupun gambar yang tidak mengandung fungsi dalam bentuk logika masih dapat dimanfaatkan sebagai data dan kemudian dapat diolah menjadi dataset.

Pada penelitian ini datanya diambil dari alamat url YouTube. YouTube menjadi tempat untuk melakukan scraping data karena ada chanel YouTube yang membahas tentang pembelajaran online yang lagi terjadi di Indonesia dan tentu saja para pengguna YouTube yang aktif berinteraksi sehingga dapat dikumpulkan data komentar yang jumlahnya bisa digunakan sebagai acuan pada machine learning. Data komentar yang dihasilkan dari proses scraping berupa data asli yang ada pada elemen XML atau HTML pada halaman website dan data dari hasil scraping tersebut akan dijadikan sebagai dataset yang jumlahnya sekitar 2.000 komentar semi terstruktur, Adapun ekstensi file dataset yang digunakan dalam proses ini adalah csv.

2.2.6 Labelisasi

Labelisasi dibuat agar memberikan identitas pada setiap data komentar yang dilakukan secara manual sesuai dengan kebutuhan sistem sebagai data latih dimana ada tiga kategori sentimen yaitu sentimen positif, negatif, dan netral. Hasil dari labelisasi ialah data berbentuk csv yang sudah terlabel.

Tabel 2.2 Tabel Labelisasi Terhadap Komentar

Id User	Komentar	Sentimen
0	Baterainya cepat habis, jadi ribet harus bawa pb kemana-mana	Negatif
1	Dari semua teman-teman yang sudah pakai hp itu, mereka puas dengan spesifikasinya	Positif
2	Harga nya mahal tidak sesuai dengan fiturnya	Negatif
3	Smartphone sih bagus tapi desain nya masih perlu banyak perubahan	Netral
4	saya sudah beli smartphone nya dan harganya menjamin kualitas	Positif

2.2.7 Preprocessing

preprocessing dilakukan agar peneliti dapat mengolah data yang ada sehingga dapat menghindari gangguan terhadap data-data yang tidak konsisten, tujuannya agar hasil output dari klasifikasi memiliki tingkat keakuratan yang tinggi.

Preprocessing memiliki beberapa tahapan sebagai berikut :

- 1) *Tokenizing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Secara garis besar memecah sekumpulan karakter dalam suatu teks ke dalam satuan kata.

- 2) *Case Folding* merupakan tahap mengubah semua huruf yang terdapat pada komentar menjadi huruf kecil. Hanya huruf “a-z” yang dapat diterima.
- 3) *Filtering* adalah tahap mengambil kata-kata yang dianggap penting.

2.2.8 Klasifikasi

klasifikasi merupakan kata serapan dari bahasa belanda, *classificatie* yang sendirinya berasal dari bahasa prancis *classification*. Istilah ini menunjuk kepada sebuah metode untuk menyusun data secara sistematis atau menurut. Berikut merupakan contoh kalimat yang akan diklasifikasikan menggunakan Naïve Bayes yang dapat dilihat pada Tabel 2.3 di bawah ini :

Tabel 2.3 Contoh Kalimat Klasifikasi Sentimen

No	Komentar	Class Kata		Label
		Positif	Negatif	
1	handphone ini kamera bagus	Ada	Tidak Ada	Positif
2	harga mahal	Tidak Ada	Ada	Negatif
3	Handphone samsung di indonesia murah	Ada	Tidak Ada	Positif
4	Saya baru pertama kali pakai samsung	?	?	?

Klasifikasi atau Definisi dibagi menjadi :

- 1) Resolusi luas adalah yang terdiri dari sebagian besar paragraf. Definisi ini sangat diperlukan dalam sebuah konsep yang rumit yang susah untuk dipahami dalam kalimat dalam kalimat pendek.

- 2) Definisi pribadi adalah resolusi yang sering digunakan sebagai sebuah pedoman untuk melakukan kegiatan atau pekerjaan. Contohnya penelitian.
- 3) Definisi formal sering juga disebut resolusi terminologis, adalah merupakan resolusi berlandaskan logika formal yang terdiri dari tiga elemen. Bentuk dari resolusi ini perbedaan dan kelas.

2.2.9 Pembobotan TF-IDF

Term Frequency-Inverse Document Frequency(TF-IDF) adalah proses untuk ekstraksi fitur dengan cara memberikan bobot nilai pada masing-masing kata yang terdapat pada dokumen, atau bisa dikatakan mengubah kata menjadi numerik. Langkah untuk melakukan TF-IDF adalah sebagai berikut :

1. Menentukan Kata Kunci
 - a. Menentukan nilai TF bisa menggunakan biner, yaitu jika suatu kata terdapat dalam sebuah dokumen akan diberi nilai 1(satu), jika tidak akan diberi nilai 0
 - b. Menentukan nilai df dengan cara menghitung jumlah dokumen yang mendukung kata tertentu.
 - c. Hitung nilai frekuensi dokumen terbalik(IDF) yang berguna untuk menentukan pentingnya kata-kata dalam dokumen. IDF dirumuskan sebagai berikut.

$$IDF = \log D/df (1)$$

Keterangan :

Df = jumlah dokumen terhitung yang mengandung istilah

D = jumlah total dokumen

2.2.10 Confusion matrix

Constructin matrix biasanya menghasilkan suatu wewenang yang ganda, yaitu wewenang fusional dan wewenang horizontal. Kelebihan dari kontruksi ini adalah pada fleksibilitasnya kemampuan dalam suatu masalah yang khusus

ataupun persoalan teknis yang terbilang unik serta implementasi terhadap kegiatan struktur ini tidak mengganggu struktur yang ada.

Tabel 2.2 Confusion Matrix

Confusion matrix		Kelas Aktual	
		Positif	Negatif
Kelas Prediksi	Positif	TP	FP
	Negatif	FN	TN

Keterangan dari nilai matriks adalah :

- True Positif(TP) adalah sejumlah data yang diprediksi positif dan kenyataan positif.
- True Negatif(TN) adalah sejumlah data yang diprediksi negatif dan kenyataannya negatif.
- False Positif(FP) adalah sejumlah data yang diprediksi positif dan kenyataan negatif.
- False Negatif(FN) adalah sejumlah data yang diprediksi negatif dan kenyataan positif.

2.2.11 Naïve bayes

Naïve bayes adalah sebuah metode pengelompokkan yang bersumber dari *teorema bayes*. Metode pengelompokkan dengan cara menggunakan probabilitas atau kemungkinan yang terjadi dan klasifikasi pertama kali dikemukakan oleh ilmuwan dari inggris yang bernama *Thomas Bayes*, yang artinya memprediksi suatu peluang yang akan terjadi di masa depan dengan mempelajari kejadian di masa lalu sehingga dikenal dengan *Teorema Bayes*. Karakteristik dari naïve bayes adalah pendapat yang terbilang kuat terhadap independensi kepada masing-masing kejadian atau kondisi.

Menurut pendapat *Olson Delen* pada tahun 2008 yang mengatakan bahwa naïve bayes disetiap keputusan, menghitung peluang kemungkinan dengan

syaratnya maka keputusannya benar, memikirkan vector informasi objek. Keuntungan penggunaan metode naïve bayes ini membutuhkan data training atau data pelatihan yang jumlahnya kecil untuk perkiraan kriteria untuk sebuah operasi pengklasifikasian. Karena diasumsikan sebagai nilai independent, jadi hanya perubahan acak dari suatu nilai didalam sebuah kelas yang diperlukan untuk menentukan pengelompokkan, bukan dari keseluruhan matriks kovarians. Ada beberapa kegunaan dari Naïve bayes adalah :

- Pengelompokkan teks seperti teks akademis atau teks berita
- Naïve bayes adalah metode machine learning
- menggunakan kemungkinan terjadi
- Membuat suatu analisis medis
- Menyaring atau mendeteksi spam
- Bahasa pemrogramannya atau codenya bisa dibilang sederhana

Persamaan metode naïve bayes sebagai berikut :

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

Tabel 2.3 Frekuensi Tabel

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	overcast	4	0
	Rainy	2	3

$$P(x \setminus c) = P(\text{sunny} | \text{Yes}) = 3/9 = 0.33$$

Likelihood Table		↑ Play Golf		
		Yes	No	
Outlook	Sunny	3/9	2/5	5/14
	Overcast	4/9	0/5	4/14
	Rainy	2/9	3/5	5/14
		9/14	5/14	

$$P(x/c) = P(\text{Sunny} | \text{Yes}) = P(x) =$$

$$P(\text{Sunny}) = 5/14 = 0.36$$

$$= 0.33 \times 0.64 + 0.36 = 0.60$$

$$P(c) = P(\text{Yes}) = 9/14 = 0.64$$

Likelihood		Class Prior Probability
$P(x \setminus c) = \frac{p(x \setminus c) p(c)}{P(x)}$		
Posterior probability		Predictor Prior Probability

Keterangannya:

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

$P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

$P(H)$: Probabilitas hipotesis H (prior probabilitas)

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

$P(X)$: Probabilitas X

Rumus yang dijelaskan diatas tentang peluang masuknya contoh karakteristik yang spesifik didalam kelas C atau posterior adalah probabilitas munculnya kelas C (sebelum masuk contoh tersebut, yang disebut prior), yang dikali dengan probabilitas karakteristik contoh pada kelas C (yang disebut likelihood) , dan dibagi dengan probabilitas karakteristik contoh secara global(disebut evidence) . Jadi, rumus yang di atas bisa juga ditulis seperti ini.

$$posterior = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

Nilai evidence selalu sama untuk disetiap kelas pada satu contoh. Nilai posteriornanti akan membandingkan dengan nilai posterior lainnya buat menentukan kelas suatu contoh yang akan diklasifikasikan. Penguraian lebih lanjut dari rumus naïve bayes tersebut yang dilakukan dengan menjabarkan $(C|F_1, \dots, F_n)$ yang menggunakan aturan perkalian seperti berikut :

$$\begin{aligned} P(C|F_1, \dots, F_n) &= P(C)P(F_1, \dots, F_n|C) = P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) = \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2) = P(C)P(F_1|C)P(F_2|C, F_1) \\ &P(F_3|C, F_1, F_2)P(F_4, \dots, F_n|C, F_1, F_2, F_3) = P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, \\ &F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \end{aligned}$$

Dapat dilihat dari hasil penjabarannya yang mengakibatkan semakin banyak dan semakin rumit faktor syarat yang bisa mempengaruhi nilai yang bisa terjadi dan hampir sulit terjadi untuk dapat melihat satu persatu, sehingga menyebabkan perhitungannya cukup sulit dilakukan. Karena itu digunakanlah satu pendapat independensi yang terbilang tinggi, sehingga masing-masing menunjukkan saling bebas independen satu sama lain. Di sinilah digunakan asumsi independensi yang sangat tinggi (naif), bahwa masing-masing petunjuk (F_1, F_2, \dots, F_n) saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i) \cdot P(F_j)}{P(F_j)} = P(F_i) \text{ Untuk } i \neq j, \text{ sehingga}$$

$$P(F_i | C, F_j) = P(F_i | C)$$

Persamaan diatas merupakan model dari naïve bayes dan selanjutnya didalam operasi klasifikasi. Untuk klasifikasinya dengan data yang berkelanjutan digunakan oleh rumus Densitas Gauss

$$P(X_i = x_i | Y = y_i) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}}$$

Keterangannya :

P = Peluang

X_i = Atribut Ke i

X_i = Nilai atribut ke i

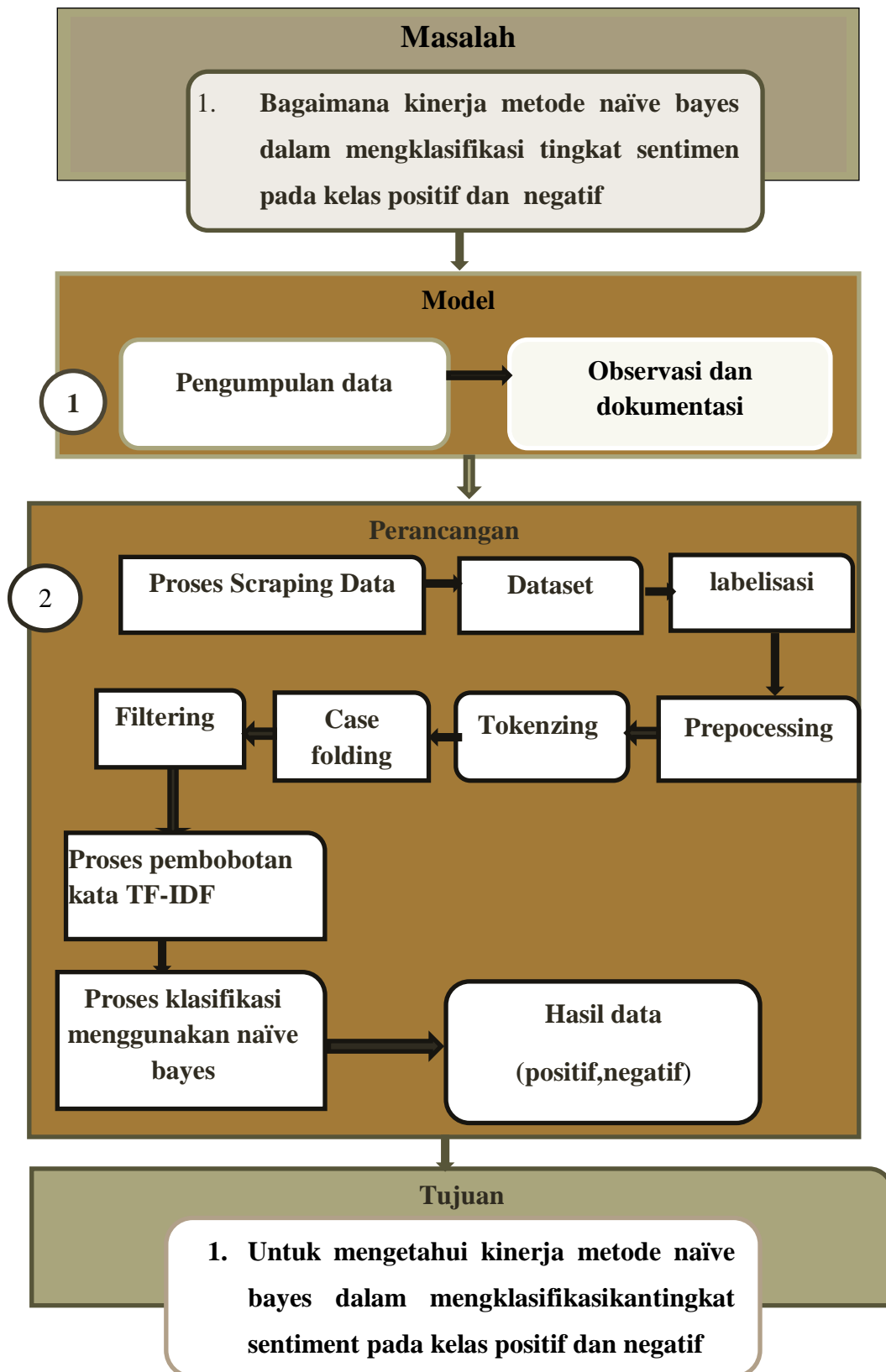
Y = Kelas yang Dicari

Y_i = Sub kelas Y yang dicari

μ = Mean, yang mengatakan rata-rata dari semua atribut

σ = Deviasi standar, mengatakan varian dari semua atribut mean

2.3. Kerangka Pikir



BAB III

METODE PENELITIAN

3.1 Jenis, Metode, Subjek, Objek, Waktu, dan Lokasi Penelitian

Berdasarkan penjelasan pada latar belakang dan kerangka pikir maka yang menjadi objek pada penelitian ini adalah **Analisis Sentimen Pembelajaran Online Pada Media Youtube Menggunakan Naïve Bayes**

3.2 Metode Penelitian

Pada penelitian ini metode yang digunakan adalah metode eksperimen. Metode eksperimen yaitu merupakan suatu percobaan untuk membuktikan suatu pertanyaan atau hipotesis tertentu. Metode eksperimen merupakan suatu cara penyajian pembelajaran yang melibatkan peserta didik secara langsung untuk membuktikan sebuah teori dari materi dari pembelajaran yang didapatkannya.

a) Sumber Data

Data primer adalah data yang diperoleh dari objek penelitian secara langsung atau berupa data mentah. Data yang di gunakan adalah data komentar dari akun youtube Tridaya Online tentang pembahasan matematika kelas 10 SMA.

Data sekunder merupakan data yang sudah tersedia atau diperoleh melalui studi pustaka, yang merupakan upaya pengumpulan data dan teori melalui buku, surat kabar serta informasi penunjang seperti dokumen, agenda, hasil penelitian dan jurnal yang berkaitan dengan penelitian ini.

b) Pengumpulan Data

1) Observasi

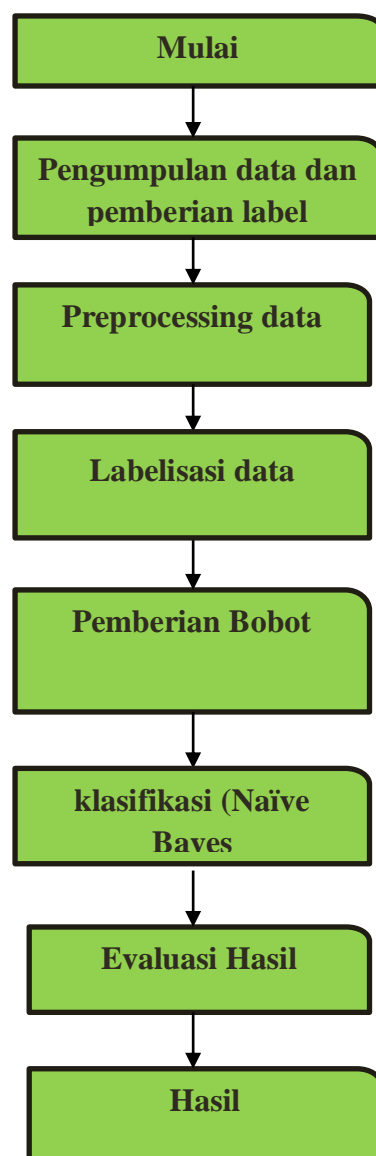
mendapatkan data-data dan fakta dari pengamatan langsung di media sosial youtube dan mengambil komentar yang ditinggalkan oleh

pengunjung kemudian menentukannya dengan menggunakan naïve bayes.

2) Studi pustaka

dilakukan dengan cara membaca, mengutip dari jurnal yang berhubungan dengan masalah yang akan dibahas dalam karya ilmiah ini. Hal ini dimaksudkan agar penulis memiliki landasan teori yang kuat.

3.3 Tahap Pemodelan



Gambar 3.1 Model yang Diusulkan

3.3.1 Data Komentar

Penelitian ini menggunakan data berupa komentar yang terdapat dalam media sosial youtube. Komentar yang diambil adalah komentar yang menggunakan Bahasa Indonesia. Jumlah komentar yang diambil sebanyak 3000 komentar dari video channel tentang pembahasan matematika kelas 10 SMA dichannel youtube Tridaya Online. Data ini di ambil dari komentar yang ditinggalkan oleh pengunjung dichannel youtube, kemudian dicrawling dengan menggunakan aplikasi python

3.3.2 Tahap Crawling

Dataset penelitian ini merupakan teks komentar dalam bahasa Indonesia yang mengandung opini. Data tersebut diperoleh dari komentar yang di crawling dari kolo, komentar pada video channel tentang pembahasan “Matematika kelas 10 SMA Dichannel Youtube Tridaya Online”. Komentar diunduh dari komentar video yang ditinggalkan oleh pengunjung sebanyak 3000 komentar.

3.3.3 Tahap Preprocessing

- 1) Tahap Cleaning
- 2) Tahap Case Folding
- 3) Tahap Tokenzing
- 4) Tahap Stemming

3.3.4 Tahap Labelisasi

Proses pemberian label dilakukan untuk menentukan komentar tersebut termasuk ke dalam kelas positif , negatif atau netral.

3.3.5 Tahap Pemberian Bobot

tahap selanjutnya dengan memberikan bobot pada tiap-tiap kata (term). Pembobotan ini dilakukan agar mendapatkan nilai dari kata yang berhasil diekstrak

3.3.6 Tahap Klasifikasi

Tahap ini adalah tahap dimana proses yang akan menentukan akurasi jumlah komentar pengguna youtube yang bersifat positif, negatif dan netral. Dengan melakukan perhitungan ini kemungkinan yang ada bisa menambah akurat dalam menentukan data yang dicari. Metode klasifikasi yang digunakan dalam penelitian ini adalah naive bayes. Naive bayes merupakan metode berdasarkan probabilitas dan Teorema Bayesian.

3.3.7 Tahap Evaluasi

Melakukan evaluasi dari eksperimen yang telah dilakukan. Evaluasi dilakukan dengan menggunakan Confusion Matrix yaitu true positive rate, true negative rate, false positive rate dan false negative sebagai indikator. TP rate adalah persentase dari kelas positif yang berhasil diklasifikasi sebagai kelas positif, sedangkan TN rate adalah persentase dari kelas negatif yang berhasil diklasifikasi sebagai kelas negatif. Kelas negatif yang diklasifikasi sebagai kelas positif. FN rate adalah kelas positif yang diklasifikasi sebagai kelas negatif

3.3.8 Tahap Implementasi Dan Pengujian Metode

Melakukan evaluasi dari eksperimen yang telah dilakukan. Evaluasi dilakukan dengan menggunakan Confusion Matrix yaitu true positive rate, true negative rate, false positive rate dan false negative sebagai indikator. TP rate adalah persentase dari kelas positif yang berhasil diklasifikasi sebagai kelas positif, sedangkan TN rate adalah persentase dari kelas negatif yang berhasil diklasifikasi sebagai kelas negatif. Kelas negatif yang diklasifikasi sebagai kelas positif. FN rate adalah kelas positif yang diklasifikasi sebagai kelas negatif.

BAB IV

HASIL PENELITIAN

4.1 Hasil Pengumpula Data

Penelitian ini menggunakan data berupa komentar yang terdapat dalam media sosial youtube. Jumlah komentar yang diambil sebanyak 3000 komentar dari video channel tentang pembahasan “Matematika Kelas 10 SMA Dichannel Youtube Tridaya Online”, Data ini di ambil dari komentar yang ditinggalkan oleh pengunjung. Data ini dibagi menjadi 2 (dua) data yaitu data training dan data testing, dimana data training sebanyak 2432 dan data Testing sebanyak 609.

Data komentar pembelajaran online ini diambil secara online lewat google collab dan phyton kemudian disalin ke file excel untuk dimasukkan ke dalam database. Data file excel ini diseleksi dan dilabeli terlebih dahulu untuk menghapus data-data yang kosong dan tidak sesuai. Data komentar para pada database dilakukan secara preprocessing dan di simpan dalam database bersih.

Tabel 4.1 Hasil pengumpulan data

No	User-id	Komentar	Kelas
1	@ugwUKFUzNL	Mantap kk	0
2	@j4hugjn7	Terima kasih kak	1
3	@zy8Ryou	Siapa yg kesini gara* susah tidur	1
4	@Ugz9rFMx9F7HHJ4Aa	Komen nya pada bilang mudah,tpi kok gw ga paham2 yah	-1
5	@HpYdCdimN4AaABAg	Lama-Kelamaan jadi susah	-1
6	@UgxxgfX_oD3PKCn4AaABAg	sgt membantu	1

7	@Ugx3WCxDcuPqQMVVSkV4 Aa	puyeng	0
8	@UgzaYV__IkJa04MG6uZ4AaA BAg	Wah makasih kak penjelasan nya sangat mudah dimengerti	0
...
3040	@UGJ6pr0qy	Makasih kak ditunggu kelanjutannya	1

Keterangan

Tanggapan positif : 1

Tanggapan negatif : -1

Tanggapan netral : 0

4.1.1 Penerapan Metode pada Data Training

Perhitungan penyelesaian menggunakan naïve bayes Classifier dalam klasifikasi komentar.

Tabel 4.3 Pembagian Data training Dan Testing

No	Kelas	Data Training	Data Testing
1	Positif	191	609
2	Negatif	47	609
3	Netral	2278	609
Total		2.516	2.516

Tabel 4.2 Komentar Data training

No	Id	Komentar	Kelas
1	@UgzaYV__IkJa04MG6uZ4AaABAg	Wah makasih kak penjelasan nya sangat mudah dimengerti	1
2	@UGJ6pr0qy	Makasih kak ditunggu kelanjutannya	0

1. Tahap Preprocessing

- ❖ Case folding : merubah huruf besar menjadi huruf kecil dan menghilangkan seluruh tanda baca.

Tabel 4.3 Case Folding

No	Komentar	Hasil Case Folding
1	Wah Makasih Kak Penjelasan nya Sangat Mudah dimengerti	wah makasih kak penjelasannya sangat mudah dimengerti
2	Makasih Kak Ditunggu Kelanjutannya	makasih kak ditunggu kelanjutannya

- ❖ Tokenization : pemecahan Teks Menjadi kumpulan kata supaya muda untuk pembobotan tiap kata

Tabel 4.4 Tokenization

No	Komentar	Hasil Tokenization
1	Makasih Kak Ditunggu Kelanjutannya	“Makasih Kak Ditunggu Kelanjutannya”
2	Wah makasih kak penjelasan nya sangat mudah dimengerti	“Wah makasih kak penjelasan nya sangat mudah dimengerti”

❖ Stemming : kata dirubah menjadi kata dasar

Tabel 4.5 Stemming

Komentar	Hasil Stemming
Makasih	“Makasih
Kak	Kak
Ditunggu	Ditunggu
Kelanjutannya	Kelanjutannya
Wah	Wah
makasih	Makasih
kak	kak
penjelasannya	penjelasannya
sangat	sangat
mudah	mudah
dimengerti	dimengerti”

❖ Langkah pertama menghitung TF dengan cara menghitung tiap bobot term pada teks

Tabel 4.7 Pembobotan TF

No	Teks	TF Positif	TF Negatif	TF Netral
1	Makasih	1	1	1
2	Kak	1	1	1
3	Ditunggu	1	0	0
4	Kelanjutan	1	0	0
5	Wah	0	1	0
6	Makasih	1	1	1
7	Kak	1	1	1
8	Penjelasan	0	1	0
9	Sangat	0	1	0
10	Mudah	0	1	0
11	Dimengerti	0	1	0
12	Terima	1	1	1
13	Kasih	1	1	1
14	Kak	1	1	1
Jumlah Term		9	11	7

Didapatkan jumlah term Count Positif = 9

Count Negatif = 11

Total jumlah teks = 14

❖ Langkah ketiga menghitung IDF menggunakan rumus :

IDF : $\log(D/F)$

D = Total Dokumen

DF = Banyak dokumen yang mengandung kata yang dicari

Tabel 4.8 Hasil Pembobotan IDF

No	Term	TF			DF	IDF
		D1	D2	D3		
1	Makasih	1	1	1	3	0.214285714
2	Kak	1	1	1	3	0.214285714
3	Ditunggu	1	0	0	1	0.071428571
4	Kelanjutan	1	0	0	1	0.071428571
5	Wah	0	1	0	1	0.071428571
6	Makasih	1	1	1	3	0.214285714
7	Kak	1	1	1	3	0.214285714
8	Penjelasan	0	1	0	1	0.071428571
9	Sangat	0	1	0	1	0.071428571
10	Mudah	0	1	0	1	0.071428571
11	Dimengerti	0	1	0	1	0.071428571
12	Terima	1	1	1	3	0.214285714
13	Kasih	1	1	1	3	0.214285714
14	Kak	1	1	1	3	0.214285714s

1) Menghitung TF-IDF

Pada tabel berikutnya adalah hasil pembobotan berdasarkan data ditabel sebelumnya. Bobo akhir term mengalihkan keduanya (**TF X IDF**)

Tabel 4.8 Pembobotan TF-IDF

Teks	Tf			Df	Idf	Tf x IDF		
	D1	D2	D3			D1	D2	D3
Makasih	1	1	1	3	0.214285714	0.214285714	0.214285714	0.214285714
Kak	1	1	1	3	0.214285714	0.214285714	0.214285714	0.214285714
Ditunggu	1	0	0	1	0.071428571	0.214285714	0	0
Kelanjutan	1	0	0	1	0.071428571	0.071428571	0	0

Wah	0	1	0	1	0.071428571	0	0.071428571	0
Makasih	1	1	1	3	0.214285714	0.214285714	0.214285714	0.214285714
Kak	1	1	1	3	0.214285714	0.214285714	0.214285714	0.214285714
Penjelasan	0	1	0	1	0.071428571	0	0.071428571	0
Sangat	0	1	0	1	0.071428571	0	0.071428571	0
Mudah	0	1	0	1	0.071428571	0	0.071428571	0
Dimengerti	0	1	0	1	0.071428571	0	0.071428571	0
Terima	1	1	1	3	0.214285714	0.214285714	0.214285714	0.214285714
Kasih	1	1	1	3	0.214285714	0.214285714	0.214285714	0.214285714
Kak	1	1	1	3	0.214285714	0.214285714	0.214285714	0.214285714

4.1.2 Proses Mencari Probabilitas Kata Pada Data Testing

Mencari probabilitas kata dilakukan untuk mendapatkan term dengan nilai yang lebih penting dan dianggap relevan agar dijadikan kata kunci. Proses pembobotan menggunakan algoritma Naive Bayes Classifier dalam proses perhitungan [12].

$$nComentpositif = 609 \quad nComentnegatif = 609$$

$$nComentnetral = 609$$

$$P(w|positif/negatif) = \frac{count(trainingset,sentiment+1)}{count(sentimen)+|v|}$$

1) Probabilitas kata”makasih”

$$P(makasih|positif) = \frac{1+1}{609+14} = 3.210272873$$

$$P(makasih|negatif) = \frac{1+1}{609+14} = 3.210272873$$

$$P(makasih|netral) = \frac{1+1}{609+14} = 3.210272873$$

2) Probabilitas kata “aman”

$$P(aman|positif) = \frac{1+1}{609+14} = 3.210272873$$

$$P(\text{aman}|\text{negatif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{aman}|\text{netral}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

3) Probabilitas kata “tunggu”

$$P(\text{tunggu}|\text{positif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{tunggu}|\text{negatif}) = \frac{0 + 1}{609 + 14} = 1.605136437$$

$$P(\text{tunggu}|\text{netral}) = \frac{0 + 1}{609 + 14} = 1.605136437$$

4) Probabilitas kata “kelanjutan”

$$P(\text{kelanjutan}|\text{positif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{kelanjutan}|\text{negatif}) = \frac{0 + 1}{609 + 14} = 1.605136437$$

$$P(\text{kelanjutan}|\text{netral}) = \frac{0 + 1}{609 + 14} = 1.605136437$$

5) Probabilitas kata “wah”

$$P(\text{wah}|\text{positif}) = \frac{0 + 1}{609 + 14} = 1.605136437$$

$$P(\text{wah}|\text{negatif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{wah}|\text{netral}) = \frac{0 + 1}{609 + 14} = 1.605136437$$

6) Probabilitas kata “makasih”

$$P(\text{makasih}|\text{positif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{makasih}|\text{negatif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{makasih}|\text{netral}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

7) Probabilitas kata “kak”

$$P(\text{kak}|\text{positif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{kak}|\text{negatif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{kak}|\text{netral}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

8) Probabilitas kata “penjelasan”

$$P(\text{penjelasan}|\text{positif}) = \frac{0 + 1}{609 + 14} = 1.605136437$$

$$P(\text{penjelasan}|\text{negatif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{penjelasan}|\text{netral}) = \frac{0 + 1}{609 + 14} = 3.210272873$$

9) Probabilitas kata “sangat”

$$P(\text{sangat}|\text{positif}) = \frac{0 + 1}{609 + 14} = 1.605136437$$

$$P(\text{sangat}|\text{negatif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{sangat}|\text{netral}) = \frac{0 + 1}{609 + 14} = 1.605136437$$

10) Probabilitas kata “mudah”

$$P(\text{mudah}|\text{positif}) = \frac{0 + 1}{609 + 14} = 1.605136437$$

$$P(\text{mudah}|\text{negatif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{mudah}|\text{netral}) = \frac{0 + 1}{609 + 14} = 1.605136437$$

11) Probabilitas kata “mengerti”

$$P(\text{mengerti}|\text{positif}) = \frac{0 + 1}{609 + 14} = 1.605136437$$

$$P(\text{mengerti}|\text{negatif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{mengerti}|\text{netral}) = \frac{0 + 1}{609 + 14} = 1.605136437$$

12) Probabilitas kata ”terima “

$$P(\text{terima}|\text{positif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{terima}|\text{negatif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{terima}|\text{netral}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

13) Probabilitas kata ”kasih”

$$P(\text{kasih}|\text{positif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{kasih}|\text{negatif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{kasih}|\text{netral}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

14) Probabilitas kata “ kak”

$$P(\text{kak}|\text{positif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{kak}|\text{negatif}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

$$P(\text{kak}|\text{netral}) = \frac{1 + 1}{609 + 14} = 3.210272873$$

Tabel 4.9 Tabel data Testing

No	Cid	Tanggapan	Value
1	@Uti50MU	valid sih	1
2	@aABag	Guru mtk paling mantap semoga sukses selalu	1

Tabel 4.10 Data Test Setelah Preprocessing

No	id	Komentar	Value
1	@Uti50MU	valid sih	1
2	@aABag	Guru mtk mantap semoga sukses selalu	1

Tabel 4.11 Data Train dan Data Test

No	Id	Hasil Preprocessing	Value
1	@__IkJa04MG6uZ4AaABAg	wah makasih kak penjelasannya sangat mudah dimengerti	1
2	@UGJ6pr0qy	Makasih kak ditunggu kelanjutannya	0
Data Testing			
1	@Uti50MU	valid sih	?
2	@aABag	Guru mtk paling mantap semoga sukses selalu	?

BAB V

HASIL DAN PEMBAHASAN

5.1 Proses Klasifikasi Naïve Bayes Pada Data Testing

Prose klasifikasi pada data testing adalah dengan mengalihkan semua nilai peluang. Nilai yang lebih tinggi merupakan kelas dari data tersebut. Perhitungan probabilitas Coment terdapat 2516 coment yang dibagi dalam data testing 609 kelas positif, 609 nkelas negatif dan 609 kelas netral .

$$P(\text{positif}) = \frac{609}{2516} = 0.2 \quad P(\text{negatif}) = \frac{609}{2516} = 0.2$$

$$P(\text{netral}) = \frac{609}{2516} = 0.2$$

Tabel 5.1 Daftar Testing kata

No	Kata	Prob positif D1	Prob negatif D2	Prob netral D3
1	Makasih	3.179650238	3.179650238	3.179650238
2	Kak	3.179650238	3.179650238	3.179650238
3	Ditunggu	3.179650238	1.589825119	1.589825119
4	Kelanjutan	3.179650238	1.589825119	1.589825119
5	Wah	1.589825119	3.179650238	1.589825119
6	Makasih	3.179650238	3.179650238	3.179650238
7	Kak	3.179650238	3.179650238	3.179650238
8	Penjelasan	1.589825119	3.179650238	1.589825119
9	Sangat	1.589825119	3.179650238	1.589825119
10	Mudah	1.589825119	3.179650238	1.589825119
11	Dimengerti	1.589825119	3.179650238	1.589825119
12	Terima	3.179650238	3.179650238	3.179650238
13	Kasih	3.179650238	3.179650238	3.179650238
14	Kak	3.179650238	3.179650238	3.179650238

Untuk mendapatkan nilai probabilitas nilai tertinggi langkah pertama yakni menghitung $(P(W_k|V_j)P(V_j))$. $P(V_j)$ didapat probabilitas dokumen:

Probabilitas Kata :

$$(makasih|Positif) : 3.179650238 * 0.2 = 635.9300476$$

$$(makasih|Negatif) : 3.179650238 * 0.2 = 635.9300476$$

$$(makasih|Netral) : 3.179650238 * 0.2 = 635.9300476$$

$$(kak|Positif) : 3.179650238 * 0.2 = 635.9300476$$

$$(kak|Negatif) : 3.179650238 * 0.2 = 635.9300476$$

$$(kak|Netral) : 3.179650238 * 0.2 = 635.9300476$$

$$(tunggu|Positif) : 3.179650238 * 0.2 = 635.9300476$$

$$(tunggu|Negatif) : 1.589825119 * 0.2 = 317.9650238$$

$$(tunggu|Netral) : 1.589825119 * 0.2 = 317.9650238$$

$$(kelanjutan|Positif) : 3.179650238 * 0.2 = 635.9300476$$

$$(kelanjutan|Negatif) : 1.589825119 * 0.2 = 317.9650238$$

$$(kelanjutan|Netral) : 1.589825119 * 0.2 = 317.9650238$$

$$(wah|Positif) : 1.589825119 * 0.2 = 317.9650238$$

$$(wah|Negatif) : 3.179650238 * 0.2 = 635.9300476$$

$$(wah|Netral) : 1.589825119 * 0.2 = 317.9650238$$

$$(makasih|Positif) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{makasih}|\text{Negatif}) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{makasih}|\text{Netral}) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{kak}|\text{Positif}) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{kak}|\text{Negatif}) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{kak}|\text{Netral}) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{penjelasan}|\text{Positif}) : 1.589825119 * 0.2 = 317.9650238$$

$$(\text{penjelasan}|\text{Negatif}) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{penjelasan}|\text{Netral}) : 1.589825119 * 0.2 = 317.9650238$$

$$(\text{sangat}|\text{Positif}) : 1.589825119 * 0.2 = 317.9650238$$

$$(\text{sangat}|\text{Negatif}) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{sangat}|\text{Netral}) : 1.589825119 * 0.2 = 317.9650238$$

$$(\text{mudah}|\text{Positif}) : 1.589825119 * 0.2 = 317.9650238$$

$$(\text{mudah}|\text{Negatif}) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{mudah}|\text{Netral}) : 1.589825119 * 0.2 = 317.9650238$$

$$(\text{mengerti}|\text{Positif}) : 1.589825119 * 0.2 = 317.9650238$$

$$(\text{mengerti}|\text{Negatif}) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{mengerti}|\text{Netral}) : 1.589825119 * 0.2 = 317.9650238$$

$$(\text{terima}|\text{Positif}) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{terima}|\text{Negatif}) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{terima}|\text{Netral}) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{kasih}|\text{Positif}) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{kasih}|\text{Negatif}) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{kasih}|\text{Netral}) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{kak}|\text{Positif}) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{kak}|\text{Negatif}) : 3.179650238 * 0.2 = 635.9300476$$

$$(\text{kak}|\text{Netral}) : 3.179650238 * 0.2 = 635.9300476$$

Selanjutnya menghitung nilai probabilitas tertinggi dari masing-masing kategori

❖ Probabilitas positif tertinggi :

$$635.9300476 * 635.9300476 * 635.9300476 * 635.9300476 * 3.179650238 * 635.9300476 * 635.9300476 * 317.9650238 * 317.9650238 * 317.9650238 * 635.9300476 * 635.9300476 * 3.179650238 * 635.9300476 = \mathbf{5.528163296}$$

❖ Probabilitas negatif tertinggi :

$$635.9300476 * 635.9300476 * 317.9650238 * 317.9650238 * 635.9300476 * 635.9300476 * 635.9300476 * 635.9300476 * 635.9300476 * 635.9300476 * 635.9300476 * 635.9300476 * 635.9300476 = \mathbf{6.951040806}$$

❖ Probabilitas netral tertinggi

$$635.9300476 * 635.9300476 * 317.9650238 * 317.9650238 * 317.9650238 * 635.9300476 * 635.9300476 * 317.9650238 * 317.9650238 * 317.9650238 * 317.9650238 * 635.9300476 * 635.9300476 * 635.9300476 = \mathbf{13.82040824}$$

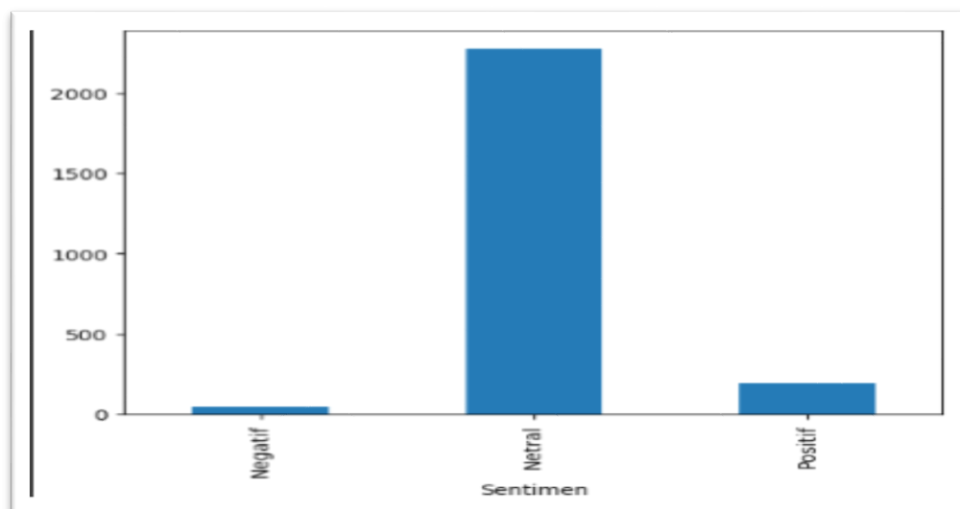
Tabel 5.2 Nilai Probabilitas

Positif tertinggi	Negatif tertinggi	Netral tertinggi
5.528163296	6.951040806	13.82040824

Dari perhitungan algoritma naive bayes yang dilakukan, lalu dicari perbandingan nilai probabilitas tertinggi dari setiap Class sehingga Class yang sudah di testing dapat di kategorikan dalam dokumen Class opini yang sesuai dengan isi teksnya. Dapat dilihat hasil dari keseluruhan proses perhitungan probabilitas tertinggi dengan naive bayes. Dalam tabel terdapat probabilitas netral memiliki nilai tertinggi. Sehingga dapat dipastikan Class yang dipilih merupakan dokumen Class netral.

Tabel 5.3 Pembagian data training

No	Class	Data training
1	Positif	191
2	Negatif	47
3	Netral	2278

**Tabel 5.4 Grafik pembagian data**

❖ Menghitung akurasi data

Akurasi adalah penghitungan dari perbandingan antara jumlah data dokumen yang relevan dan jumlah keseluruhan dokumen dalam database [13].

$$\text{Akurasi} = \frac{\text{jumlah dokumen}}{\text{jumlah keseluruhan dokumen}}$$

$$\text{Akurasi} = \frac{14}{2516}$$

$$\text{Akurasi} = 55.77\%$$

BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

1. Berdasarkan hasil klasifikasi menunjukan bahwa sentimen analisis pembelajaran online pada media youtube lebih dominan netral. Dimana total data 2516, dengan kelas netral memiliki data 2278 , kelas positif memiliki 171 data dan kelas negatif 47 . Setelah dilakukan klasifikasi dengan algoritma naive bayes dan pembobotan TF-IDF pada data maka didapatkan hasil akurasi yang tidak seimbang.

6.2 Saran

Berdasarkan kesimpulan skripsi di atas, peneliti memberikan saran untuk yang akan meneliti selanjutnya,

1. Dapat dikembangkan dengan metode lain, selain naive bayes agar dapat meningkatkan hasil yang lebih baik lagi.
2. Bisa mendapatkan hasil data yang lebih akurat dan hasil akhir yang seimbang.

DAFTAR PUSTAKA

- [1] P. Kowalchuk, "Implementing a Drilling Reporting Data Mining Tool Using Natural Language Processing Sentiment Analysis Techniques," in SPE Middle East Oil and Gas Show and Conference, 2019.
- [2] Hermanto and Astrid Noviriandini, "Analisis Sentimen Terhadap Belajar Online Pada Masa COVID-19 Menggunakan Algoritma Support Vector Machine Berbasis Particle Sarm Optimization." *Jurn. Infor. Kaputama (JIK)*, Vol. 5 No. 1, Januari 2021.
- [3] Fransiska Vina Sari and Arief Wibowo, "Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi." *Jurn. Simetris*. Vol. 10 . No. 2. Nov. 2019.
- [4] Buntoro, 2017.
- [5] Samsir, Ambiyar Unung Verawardani, Firman edi and Ronal Watriantos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes, " *J. Media Inf. Budidarma*, "Vol. 5, No. 1, Jan 2021, Page 157-163.
- [6] Abu Salam , Junta Zeniarja and Rima Septiyan Uswatun Khasanah , "Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearst Neighbor," *Prosiding SIN*. 2018, 978-602-8557-20-7.
- [7] Kadek Ary Budi Permana, Made Sudarma and Wayan Gede Ariastina, "Analisis Rating Sentimen pada Video di Media Sosial Youtube Menggunakan STRUCT-SVM," *Majalah. Ilmiah. Tekno.Elektro*, Vol. 18, No. 1, Januari - April 2019.
- [8] Balya, "Analisis Sentimen Pengguna Youtube Di Indonesia Pada Reiew Smartphone Menggunakan Naïve Bayes, 2019.

- [9] Sumarni Adi, Murni Wulandari, Anis Kemala Mardiana and Ahmad Muzakki,"Topik Dan Tren Analisis Sentimen Pada Media Online,"Seminar .Nas.Teknol. Inform. Multi, Univ. Amikom. Yogya, 10 Februari 2018. Issn: 2302-3805.
- [10] Adhi Viky Sudiantoro and Eri Zuliarso,"Analisis Sentimen Twitter Menggunakan Text Mining Dengan Algoritma Naïve Bayes Clasifier," Infor.Vol.10, No.2, Okt. 2018 : 69-73 E-Issn:2714-8769| -Issn: 2085-3343.
- [11] Agatha Deolika, Kusrini and Emha Taufiq Luthfi , " Analisa Pembobotan Kata Pada Klasifikasi Text Mining," Journ .Tekno. Inform. Vol.3, No.2. Des.2019

LISTING PROGRAM

```
!pip install textblob
!pip install sastrawi
!pip install emoji
!pip install PySastrawi
!pip install textblob
!pip install preprocessor
pip install emoji
pip install tweet-preprocessor
import os
import pandas as pd
import re
import string
from textblob import TextBlob
import preprocessor as p
# from preprocessor.api import clean, tokenize, parse
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
import datetime
from datetime import timedelta
import numpy as np
import emoji
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.feature_extraction.text import CountVectorizer
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
```

```

# from Sastrawi.StopWordRemover.StopwordRemoverFactory import StopwordR
emoverFactory
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRe
moverFactory
import pandas as pd
def load_data():
data = pd.read_excel('/content/data_youtub(1).xlsx')
return data
youtube_def = load_data()
youtube_def = pd.DataFrame(youtube_def[['time','author','text']])
youtube_def.head(3041)

#Preprocessing

#Cleaning

#hapus mantion
def remove_pattern(text, pattern_regex):
r = re.findall(pattern_regex, text)
for i in r:
text = re.sub(i,"",text)
return text

#hapus simbol
def remove_simbol(text):
text = ' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t]) |(\w+:\w+\S+) "," ",text)
.split())
return text
youtube_def['remove_http'] = youtube_def['clean_pertama'].apply(lambda x: rem
ove_simbol(x))
youtube_def.head()
def remov(text):
#remove stock market tickers like $GE
text = re.sub(r"\$\w*", "",text)

```

```

#remove old style retweet text "RT" (khusus data dari tweeter)
text = re.sub(r'^RT[\s]+', '', text)

#remove hastags
#only removing the hash # sign from the word
text = re.sub(r'#', '', text)

#remove angka
text = re.sub('[0-9]+', '', text)

return text

youtube_def['remove_hashtag'] = youtube_def['remove_http'].apply(lambda x:remov
v(x))
youtube_def.head()

#Case Folding
#Tokenizing
#Filtering
#Stemming

import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
stopwords_indonesia = stopwords.words('indonesian')
from nltk.tokenize import TweetTokenizer
from Sastrawi.StopWordRemover.StopWordRemoverFactor import StopWordRe
moverFactory, StopWordRemover, ArrayDictionary
stop_factory=StopWordRemoverFactory(get_stop_words())
more_stopwords=["yg", "dg", "dgn", "ny", "d", "klo", "kak", "gara", "kalo", "amp", "bi
ar", "bikin", "bilang", "gak", "ga", "krn", "nya", "nih", "sih", "si", "tau", "tdk", "tuh", "utk"
, "ya", "jd", "jgn", "sdh", "aja", "n", "t", "wa", "kka", "kk", "nyg", "hehe", "pen", "u", "nan",
"loh", "rt", "&", "yah", "jgn", "ga", "ok", "banget", "bgt", "rp", "ngk", "bgus", "cmn",
"kyk", "gw", "bapa", "gua", "mas", "log", "itung", "tpi", "kereeennn", "gan", "wow", "ini

```

```

h", "abrar", "kuli", "mudahhh", "erti", "thn", "sy", "wkww", "tksh", "ko", "jh", "las", "ap
", "gtu", "skrng", "udh", "ngerti", "thanks", "lgi", "mtk", "tp", "sya", "hrs", "smk"]

data = stop_factory + more_stopwords dictionary = ArrayDictionary(data)
str = StopWordRemover(dictionary)

print(data)

#import sastrawi

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

factory = StemmerFactory()

stemmer = factory.create_stemmer()

#numpy Emoticons happy

emoticons_happy = set([':', ':)', ':)', ':o)', ':3', ':c)', ':>', '=]', '#)', '=)', ':}', ':^', ':-D', ':D', '8-
D', '8D', 'xD', 'X-D', 'XD', '=-D', '=D', '=-3', '=3', ':-))', ':'-)'', ':')', ':*', ':^*', '>:p', ': p', ':p', 'X-
P', 'x-p', 'xp', 'XP', ':-p', ':p', '=p', ':-b', ':b', '>:)', '>:)', '>:-))])

#set Emoticons sad

emoticons_sad = set([
':L', ':-/', '>:/', ':S', '>:', ':@', ':-(', ':(, ':-||', '=L', '<:', ':-[', ':-<', '=\\', '=/', '>:', ':(, '(>.', '<:', ':-
(", ':", '(`', ':\\", ':-c', ':c', ':(, '>:', '\\', ':(')

#all emoticons (happy + sad)

emoticons = emoticons_happy.union(emoticons_sad)

def clean_youtube(text):

    #tokenize tweets

    tokenizer = TweetTokenizer(preserve_case=False, strip_handles=True, reduce_le
n=True) tweet_token = tokenizer.tokenize(text)

    tweet_clean = []

    for word in tweet_token:

        if(word not in data and

        word not in emoticons and

        word not in string.punctuation):

            stem_word = stemmer.stem(word)

            tweet_clean.append(stem_word)

    return tweet_clean

```

```

youtube_def['bersih'] =youtube_def['remove_hastag'.
apply(lambda x: clean_youtube(x))
youtube_def.head()
#remove punct
def remove_punct(text):
text = " ".join([char for char in text if char not in string.punctuation])
return text
youtube_def['bersih'] = youtube_def['bersih'].apply(lambda x: remove_punct(x))
youtube_def.head()
youtube_def.head()
#remove komentar kosong
youtube_def = youtube_def[youtube_def['bersih']!=""]
youtube_def.head()
#reset index
youtube_def = youtube_def.reset_index(drop=True)
youtube_def.head()
youtube_def.drop_duplicates(subset="remove_hastag",keep="first",inplace=True)
youtube_def.head()
#remove kolom
youtube_def.drop(youtube_def.columns[[0,1,2,3,4,5]],axis=1,inplace=True)
youtube_def.head(10)
#simpan data bersih
youtube_def.to_csv("data_bersih.csv",encoding='utf8',index=False)
!pip install googletrans==3.1.0a0
#import library yang dibutuhkan
import pandas as pd
import googletrans
from googletrans import Translator
#membaca dan menyimpan file csv sebagai dataframe
df = pd.read_csv('data_bersih.csv') df.head(10)
translator = Translator()

```

```

translations = {}
for column in df.columns:
    #unique elements of the column
    unique_elements = df[column].unique()
    for element in unique_elements:
        #Adding all the translations to a dictionary (translations)
        translations[element] = translator.translate(element).text
translations
df.replace(translations,inplace=True)
df.head(10)
df.to_csv('translate_data.csv',encoding='utf8',index=False)
import pandas as pd
def load_data_translate():
    data = pd.read_csv('translate_data.csv')
    return data
youtube_df = load_data_translate()
youtube_df.head(10)
!pip install vaderSentiment
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
analyser = SentimentIntensityAnalyzer()
scores = [analyser.polarity_scores(x) for x in youtube_def['bersih']]
print(scores)
youtube_def['compound_Score'] = [x['compound'] for x in scores]
youtube_def.head(10)
youtube_def.loc[youtube_def['compound_Score'] < 0, 'Sentimen'] = 'Negatif'
youtube_def.loc[youtube_def['compound_Score'] == 0, 'Sentimen'] = 'Netral'
youtube_def.loc[youtube_def['compound_Score'] > 0, 'Sentimen'] = 'Positif'
youtube_def.head(10)
youtube_def.to_csv('hasil_pelabelan.csv',encoding='utf8',index=False)
youtube_def['Sentimen'].value_counts()

```

```

# most frequent words counter (code adapter from https://www.kaggle.com/rodolfo
oluna/fake-news-detector)
from nltk import tokenize
import seaborn as sns
from matplotlib import pyplot as plt
token_space = tokenize.WhitespaceTokenizer()
def counter(saran, columnSaran, quantity):
all_words = ' '.join([saran for saran in saran[columnSaran].astype('str')])
token_phrase = token_space.tokenize(all_words)
frequency = nltk.FreqDist(token_phrase)
df_frequency = pd.DataFrame({"Word": list(frequency.keys()),
"Frequency": list(frequency.values())})
df_frequency = df_frequency.nlargest(columns = "Frequency", n = quantity)
plt.figure(figsize=(12,8))
ax = sns.barplot(data = df_frequency, x = "Word", y = "Frequency", palette='dee')
# ax = sns.barplot(data = df_frequency, x = "Word", y = "Frequency", color = 'blue')
ax.set(ylabel = "Count")
plt.xticks(rotation='vertical')
plt.show()
print(youtube_def.groupby(['Sentimen'])['bersih'].count())
youtube_def.groupby(['Sentimen'])['bersih'].count().plot(kind="bar")
plt.show()
import pandas as pd
import numpy as np
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
from openpyxl import workbook
!pip install openpyxl==3.0.0
youtube_def.head()
scores = [analyser.polarity_scores(x) for x in youtube_def['bersih']]
youtube_def['score_komentar'] = [x['compound'] for x in scores]

```

```

youtube_def.head()
youtube_def.loc[youtube_def['score_komentar'] < 0, 'Sentimen_pred'] = 'Negatif'
youtube_def.loc[youtube_def['score_komentar'] == 0, 'Sentimen_pred'] = 'Netral'
youtube_def.loc[youtube_def['score_komentar'] > 0, 'Sentimen_pred'] = 'Positif'
youtube_def.head()
import pandas as pd
import numpy as np
path = "/content/data_bersih.csv"
youtube_def.head()
#pembagian data training dan testing
import pandas as pd
import numpy as np
path = "/content/data_youtube (1).xlsx"
dataset=pd.read_excel(path)
dataset
dataset.head()
import matplotlib.pyplot as plt
plt.scatter(dataset['cid'],dataset['text'])
dataset.dtypes
X=dataset[['cid', 'text']]
y=dataset[['author']]
print(X)
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(X,y, test_size =0.2)
len(X_train)
len(X_test)
X_train
X_test
Y_train
Y_test

```


Dataset

No	Id	Komentar	Kelas
1	UgwUKFUzNLSFRwL1hPB4AaABAg	Mantap kka.	netral
2	UgwJ4hUqJn7JvMq4sex4AaABAg	Terima kasih kak	netral
3	Ugzx8Ry0Uafmsedxxat4AaABAg	Siapa yg kesini gara* susah tidur	netral
4	UgytwRhiJirpkq8Y6ap4AaABAg	Terimakasih membantu banget nih ☺	netral
5	UgwzDOlrbmhaRvhmEKx4AaABAg	wa nya ngk ada yang balas	negatif
6	UgymmdEakgZ3CFdmPI94AaABA	Terimakasih ilmunya pak☺☺	positif
7	UgzD-8oyXPIKbrl0Oot4AaABA	Terimakasih ilmunya pak☺☺	positif
8	UgxQuS9UkCSKaDoFgYl4AaABAg	Pembahasaanya bgus bgt kak, cmn otaku aja yg minim :'	negatif
9	Ugyl9b9BFqQJoKYTqRZ4AaABAg	seru banget ngejelasinya, soalnya kakanya bobrok banget. lebih paham disini sih gak kaku kyk di sekolahhh T_T	positif
10	UgzhCp-bjpxdfJkT2ZV4AaABAg	terima kasih pak, penjelasannya sangat mudah dipahami	netral
11	Ugz9rFMx9F-7sd8nHHJ4AaABAg	Komen nya pada bilang mudah,tpi kok gw ga paham2 yah☹	negatif
12	Ugzy_DfWMP08GTTXUl4AaABAg	jujur aku ngebleng☹	negatif
13	UgyweN7vTCjqIQOY5up4AaABA	Pas bapa ngomong "gampang ya adik adik ya?", wow pak alis sayah kerang kerung inih	negatif

14	Ugx7h0vI78vXGu2h9-Z4AaABAg	kereeennnnn	netral
15	UgyheuxncGglwwGD-St4AaABAg	Keren gan videonya	netral
16	Ugy-eyJ7TkEmXYp5sz94AaABAg	Andai Kakak yang ajarin di kls sya	netral
17	UgygqW4YrGIQKVai5Tx4AaABA	Mudahhhh dimengerti sumpah	positif
18	UgxID2_JsMPghpRt58t4AaABAg	selama ini saya bingung diajar guru saya,karna guru saya sistem nya yg cpt tnggp lngsng next ² aja ga mikir kita yg tablo lngsng ditinggal gtu aja,tp skrng udh ngerti si wkwwk,thanks kak udh mau ajarin ga kayak guru saya wkwwk	netral
19	UgxjQP_nQTLjqGf0t554AaABAg	Kak contohnya dimudahin lgi kak biar mudah diingat	negatif
20	Ugwho3kRu3iap4LAIA14AaABAg	Tksh sangat jelas buat ngajar anak saya	positif
21	Ugy92wZaEagqaP09vY54AaABAg	Aku kesini Krn mau ngajar.. ²	netral
22	UgyKM4RtqPJMcuCbJth4AaABAg	guru mtk paling mantap sedunia semoga suksen terus pak ²²²²	positif
23	UgwohDdpLVAM_cfGWhF4AaABAg	Mudah di pahami	netral
24	UgykOeaf2E-Wq7MOJUx4AaABAg	Andai guru MTK sya seperti kakak ²²	Netral
25	Ugw4phI-HmWNUuGzWop4AaABAg	Emang pro beliau ini ²	Netral

26	UgxV-3p60v6vBwVTNkZ4AaABAg	makasih kakak, membantu banget buat aku yang lagi sekolah farmasi soalnya harus lancar di matematika ☺☺	positif
27	UgzjPQVNhIhtCfiX18R4AaABAg	Thanks Kak sangat membantu	netral
28	UgyV_oI1AAsLTpflCPh4AaABAg	Kaka ada tempat bimbanya ga	netral
29	UgwS6t-Ku11B29Cpzop4AaABAg	ga sanggup logariitmaaa☹☹☹☹	negatif
30	Ugx9grBUBXVW1U2zzt4AaABAg	Penjelasannya kayak guru matematika waktu SMP	netral
31	Ugxb8ZcFB2_CKuWrg54AaABAg	Ini beneran keluar di pelajaran kelas 10 kan kak?	netral
32	Ugzlh7N8rjwzpyhFcyJ4AaABAg	Maksih ka sangat membantu☺	netral
33	UgxYhrNIO4MS6LxRKKd4AaABAg	Kaka makasih banget Yaa aku jadi sering nonton chanel nya Kaka walaupun kadang suka aku ulang ulang trus biar ngertiii	netral
34	UgyEs3B8jM6z-DjfAuF4AaABAg	Kukira susah ternya mudah ☺	positif
35	UgyTe-y1azTI0ngxmUx4AaABAg	Matematika mudah dan menyenangkan bukan?	netral
...
3041	Ugyy7hcwjSSiB2Fup5V4AaABAg	Mudah banget di pahami penjelasan nya kak:)	positif

Daftar Riwayat Hidup



Nama	: Nazmia Djibu
Nim	:T3116001
Tempat ,Tanggal Lahir	:Gorontalo,15 Desember 1995
Alamat	:Jln. Kenangan
Agama	:Islam
Kewarganegaraan	:WNI
Email	: nazmiadjibu12@gmail.com

Riwayat pendidikan dan pekerjaan :

1. Tahun 2008 , menyelesaikan pendidikan di Sekolah Dasar Negeri Bolobungkang
2. Tahun 2011, menyelesaikan pendidikan di Sekolah Menengah Pertam Negeri 1 Lobu
3. Tahun 2014, menyelesaikan pendidikan di Sekolah Menengah Atas Negeri 1 Lobu
4. Tahun 2016, diterima menjadi mahasiswa di Perguruan Tinggi Swasta Universitas Ichsan Gorontalo.

