

**ANALISIS SENTIMEN KOMENTAR YOUTUBE TERHADAP
TAYANGAN #TERBARU! TEMUAN PENYIMPANGAN AHLAK
DIPONPES AL-ZAITUN MENGGUNAKAN METODE NAÏVE BEYES**

SKRIPSI



Oleh:

MOCHAMMAD FEBRIALDI ANSYAH

2020503021

PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS IBRAHIMI

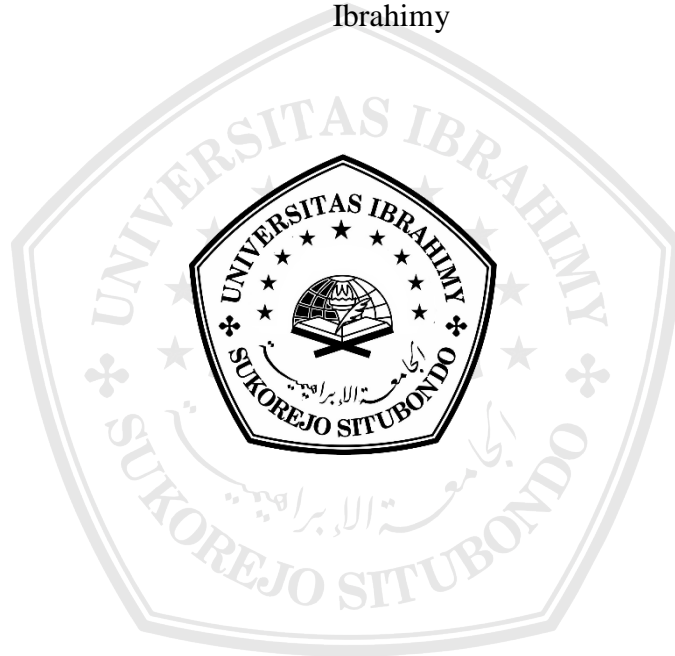
SITUBONDO

2024

**ANALISIS SENTIMEN KOMENTAR YOUTUBE TERHADAP TAYANGAN
#TERBARU! TEMUAN PENYIMPANGAN DAN MASALAH AHLAK DI
PONPES AL-ZAITUN MENGGUNAKAN METODE NAÏVE BEYES**

SKRIPSI

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



Oleh:

MOCHAMMAD FEBRIALDI ANSYAH

2020503021

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS IBRAHIMI
SITUBONDO
2024**

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : MOCHAMMAD FEBRIALDI ANSYAH
NPM/NIRM : 2020503021
Program Studi : Teknologi Informasi
Fakultas : Sains dan Teknologi

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

Situbondo, 17 Agustus 2024

Saya yang menyatakan,



MOCHAMMAD FEBRIALDI ANSYAH

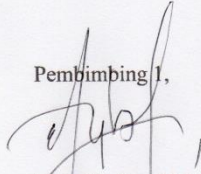


PERSETUJUAN PEMBIMBING

Nama : MOCHAMMAD FEBRIALDI ANSYAH
NPM : 2020503021
Judul : ANALISIS SENTIMEN KOMENTAR YOUTUBE
TERHADAP TAYANGAN #TERBARU! TEMUAN
PENYIMPANGAN DAN MASALAH AHLAK DI
PONPES AZ-LAITUN MENGGUNAKAN METODE
NAÏVE BEYES

Telah disetujui oleh:

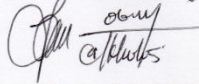
Pembimbing 1,



Abd Ghofur, M.Kom.

NIDN: 0711088303

Pembimbing 2,



Lukman Fakhri Lidimillah, M.Kom.

NIDN: 0715099001



PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN KOMENTAR YOUTUBE TERHADAP TAYANGAN
#TERBARU! TEMUAN PENYIMPANGAN DAN MASALAH AHLAK DI
PONPES AL-ZAITUN MENGGUNAKAN METODE NAÏVE BEYES**

MOCHAMMAD FEBRIALDI ANSYAH

2020503021

Telah dipertahankan di depan dewan penguji Sidang/Munaqasah Skripsi pada hari Kamis, tanggal 25 Juli 2024 sebagai salah satu syarat memperoleh gelar sarjana (S.Kom.) pada Fakultas Sains dan Teknologi Universitas Ibrahmy.

Tim Penguji,

Ketua Sidang,

Dr. Ach. Khumaidi, M.P
NIDN: 0722049001

Sekretaris Sidang,

Usman Hidayat, S.Kom

Penguji I,

Achmad Baijuri, M.Kom.
NIDN: 0722129201

Penguji II,

Farihin Lazim, M.Tr. T
NIDN: 0711099201

Mengetahui,
Dekan

Abd. Ghofur, M.Kom.
NIDN: 0711088303

KATA PENGANTAR

Segala Puji dan syukur alhamdulillah kita panjatkan kehadiran Allah SWT atas berkat-Nya lah penyusunan skripsi ini bisa dilaksanakan dengan baik. Tidak lupa pula solawat serta salam kepada nabi Muhammad SAW.

Penulis menyadari bahwa dalam proses penulisan skripsi ini banyak mengalami kendala, akan tetapi berkat bantuan, dan bimbingan dari berbagai pihak kendala-kendala yang dihadapi tersebut dapat diatasi.

Untuk itu penulis menyampaikan ucapan terima kasih kepada Bapak Ahmad Homaidi, M.Kom. dan Bapak Abd Ghofur, M.Kom. selaku pembimbing yang telah meluangkan waktu, tenaga dan pikiran untuk memberikan bimbingan, dan arahan yang sangat berharga kepada penulis selama menyusun skripsi ini.

Selanjutnya ucapan terima kasih penulis sampaikan pula kepada :

1. KHR. Achmad Azaim Ibrahimi selaku Pengasuh Pondok Pesantren Salafiyah Syafi'iyah Sukorejo.
2. KH. Ach. Fadlail, M.H. selaku Rektor Universitas Ibrahimi.
3. Bapak. Abdul Ghofur, M.Kom. selaku Dekan Fakultas Sains dan Teknologi.
4. Bapak. Firman Santoso, M.Kom. selaku Ka. Prodi Teknologi Informasi.

Penulis berharap semoga skripsi ini bisa bermanfaat, tidak lupa pula penulis berharap adanya saran dan kritik yang bersifat membangun demi kesempurnaan laporan ini.

Situbondo, 17 Agustus 2024

Penyusun,

MOCHAMMMAD FEBRIALDI ANSYAH

DAFTAR ISI

PENGESAHAN	i
KATA PENGANTAR.....	ii
DAFTAR ISI	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xi
DAFTAR SEGMENT PROGRAM.....	xii
DAFTAR RUMUS	xiii
BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2. Identifikasi Masalah.....	3
1.3. Rumusan Masalah.....	3
1.4. Batasan Masalah	3
1.5. Tujuan Penelitian	3
1.6. Manfaat Penelitian	4
1.7. Metode Penelitian	4
1.8. Sistematika Pembahasan.....	7
BAB II	9
TINJAUAN PUSTAKA	9
2.1. Penelitian Terdahulu	9
2.2. Landasan Teori	11
a. <i>Text mining</i>	11
b. Analisis	12
c. Sentimen.....	12
d. Analisis Sentimen.....	12
e. Komentar	13
f. <i>Youtube</i>	13
g. <i>Naïve Bayes</i>	13
h. <i>Bag of Word</i>	14
BAB III.....	15
PERANCANGAN SISTEM.....	15
3.1. Analisis Kebutuhan	15
3.1.1. Analisis Kebutuhan Sistem.....	15
3.1.2. Analisa Proses	16

3.1.3. Analisis Metode.....	18
3.2. Rancangan Sistem.....	19
3.2.1. Arsitektur Sistem.....	19
3.2.2. Arsitektur Perhitungan.....	20
BAB IV.....	28
IMPLEMENTASI SISTEM.....	28
4.1. Implementasi Sistem.....	28
4.1.1. Tahap <i>Preparation</i>	28
4.1.2. Tahap <i>Preprocessing</i>	29
4.1.3 Tahap implementasi.....	34
4.2. Hasil Uji Coba.....	34
a. Percobaan Pertama.....	38
b. Percobaan Kedua.....	38
c. Percobaan Ketiga.....	39
4.2. Analisis Hasil.....	39
a. Percobaan Pertama.....	39
b. Percobaan Kedua.....	40
c. Percobaan Ketiga.....	41
4.3. Evaluasi.....	42
BAB V.....	41
PENUTUP.....	41
5.1. Kesimpulan.....	41
5.2. Saran.....	41
DAFTAR PUSTAKA.....	41

DAFTAR GAMBAR

Gambar 1. 1 Web development live cycle (WDLC) model.....	6
Gambar 3. 1 Arsitektur Aplikasi.....	20
Gambar 3. 2 Arsitektur Sistem	21
Gambar 3. 3 Pengambilan Data Komentar	22
Gambar 3. 4 Flowchart Proses Tokenizing.....	24
Gambar 3. 5 Flowchart Proses Stopword	26
Gambar 3. 6 Flowchart Proses Labelling.....	27



DAFTAR TABEL

Tabel 3. 1 Spesifikasi perangkat yang di gunakan	16
Tabel 3. 2 Perangkat Yang di gunakan.....	16
Tabel 3. 3 Contoh Teks Normalizer	23
Tabel 3. 4 Contoh Tokenazing	24
Tabel 3. 5 Contoh Stopword	25
Tabel 3. 6 Contoh Lebellng.....	26
Tabel 4. 1 lowercassing	31
Tabel 4. 2 Tokenazing	32
Tabel 4. 3 Stopword	34
Tabel 4. 4 Labelling	34
Tabel 4. 5 Pengujian Pertama.....	38
Tabel 4. 6 Pengujian Kedua	39
Tabel 4. 7 Pengujian Ketiga	39
Tabel 4. 8 Data Hasil Klasifikasi.....	42



DAFTAR SEGMENT PROGRAM

Segmen 4. 1 Pengambilan Data	28
Segmen 4. 2 Teks Normalizer	29
Segmen 4. 3 Tokenazing	31
Segmen 4. 4 Stopword	32
Segmen 4. 5 Implementasi	34



DAFTAR RUMUS

Rumus 2. 1..... 13



ABSTRAK

Mochammad Febrialdi Ansyah, 2024. **Analisis Sentimen Komentar Youtube terhadap Tayangan #Terbaru! Temuan Penyimpangan dan Masalah Ahlak di Ponpes Al-Zaitun Menggunakan Metode Naïve Bayes.** Skripsi, Program Studi Teknologi Informasi, Universitas Ibrahimi. Pembimbing: (I) Abd. Ghofur M. Kom., (II). Lukman Fakhri Lidimillah M.Kom.

Cara mengetahui umpan balik dari konten di sosial media adalah dengan cara melihat komentar yang diutarakan oleh masyarakat. Hal ini bisa dipelajari melalui salah satu cabang ilmu teknologi komputer yaitu text mining. Dengan teks mining tersebut kita dapat menggali informasi dari data komentar. Analisis sentimen merupakan solusi dalam mengelompokkan respon masyarakat dalam bentuk komentar kedalam komentar positif, netral dan negatif. Dari channel youtube iNews menayangkan salah satu videonya tentang pondok pesantren Al-Zaytun. pada tayangan tersebut dapat diperoleh informasi dari data yang tidak terstruktur sehingga diperlukan suatu metode dalam mengelompokkan data komentar dalam komentar yang positif, netral dan negatif. Pada analisis sentimen ini akan menggunakan pengklasifikasian menggunakan metode naïve bayes. Naïve bayes merupakan metode yang sederhana namun tingkat akurasi yang baik di dalam proses klasifikasi. Pengujian tersebut mendapatkan nilai akurasi 64%, recall 93%, Precision 75% dengan ini dapat disimpulkan bahwa metode naïve bayes cocok digunakan dalam analisis sentiment pada komentar youtube.[1]

Kata kunci : Naïve bayes, Al-zaytun, Analisis sentiment.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Seiring perkembangan zaman teknologi informasi dan komunikasi berkembang dengan pesat, baik dari segi komputer, internet dan masih banyak lagi perkembangan yang dapat kita rasakan, hal ini sangat berdampak ke seluruh aspek di kehidupan manusia mulai dari pekerjaan, bisnis, sampai keseharian. Teknologi informasi dan komunikasi merupakan perkembangan yang paling cepat dan menyesuaikan dengan dengan kebutuhan manusia, sehingga tidak dapat dipungkiri rata-rata manusia baik dari anak-anak, remaja, sampai dewasa juga memanfaatkan teknologi informasi dan komunikasi. Hal ini sangat membantu pemindahan informasi antar media menjadi lebih mudah dan fleksibel.[2]

Faktanya penggunaan ponsel pintar telah melonjak pesat pada setiap tahunnya. Salah satu manfaat dari ponsel pintar ini adalah media social, tentu media social adalah salah satu aspek penting dalam pemanfaatan teknologi informasi dan komunikasi, pada tahun 2018 mencapai angka 171,17 juta jiwa dari total populasi 264,16 juta jiwa mencapai angka 64,8 persen dari penduduk Indonesia.[3] [4] YouTube adalah sebuah platform berbagi video online yang memungkinkan pengguna untuk mengunggah, menonton, dan berinteraksi dengan video-video dari berbagai jenis, mulai dari video musik, vlog pribadi, tutorial, sebagai media edukasi hingga video hiburan lainnya.[5]

Analisis sentiment merupakan proses mengolah data secara tekstual secara otomatis untuk mendapatkan informasi dari teks yang sedang diteliti. Hal ini bertujuan untuk mengelompokkan teks berupa pendapat atau opini terhadap suatu fenomena ataupun berupa masalah yang bersifat opini positif, netral dan negatif.[6][7] Selain itu analisis sentiment ini merupakan salah satu studi komputer tentang pendapat, perilaku dan perasaan manusia tentang keseluruhan.[8][9]

Selain sebagai media edukasi youtube juga sebagai salah satu media penyampaian berita terkini, sehingga masyarakat dapat menerima berita terkini dari salah satu chanel yang memang menyediakan konten-konten berita terkini. Palsanya dengan adanya youtube penyampaian sebuah berita lebih mudah dengan melihat pengguna youtube yang besar. Hal ini dapat dimanfaatkan sebagai fasilitator penyampaian berita oleh beberapa stasiun televisi. Pada tahun 2023 adalah tahun dimana masyarakat

dihebohkan dengan salah satu berita tentang pondok pesantren al-zaitun yang didalamnya terdapat penemuan penyimpangan dan masalah ahlak yang diajarkan kepada para santrinya tentu dengan video tersebut banyak komentar masyarakat yang tidak terarah baik dari komentar yang membangun dan juga ada komentar yang tidak membangun.[10][11]

Dengan berita tersebut masyarakat Indonesia dihebohkan dengan beredarnya berita tentang pondok pesantren Al-Zaytun yang diberitakan didalamnya terdapat ajaran yang melenceng dari syariat agama. Menanggapi kasus itu banyak media berita baik stasiun televisi, website berita sampai konten creator membuat ulasan tentang pondok pesantren Al-Zaytun tersebut didalam chenelnya salah satunya adalah chanel iNews.

Pada penelitian lain terkait analisis sentiment yang dilakukan oleh[12], yang membahas tentang analisis sentiment komentar di youtube tentang ceramah ustadz abdul somad menggunakan algoritma naïve bayes dengan data sebanyak 1000 komentar dari 10 video yang terbagi kedalam tiga bagian yaitu positif, netral dan negatif. Pada komentar ini terdapat (67%) atau 672 komentar positif, (27%) atau 267 komentar netral dan (6%) atau 61 komentar negatif. Pada penelitian lain yang dilakukan oleh[9], membahas tentang implementasi analisis sentiment masyarakat mengenai kenaikan harga Bbm pada komentar youtube dengan metode gaussian naïve bayes dengan total data evaluasi yang digunakan berjumlah 260 yang terdiri dari 172 sentimen negative(-1) atau 66,2% dan 88 sentimen positif (1) atau 33,8%. Kemudian pada penelitian yang dilakukan oleh[10] penggunaan data latih sebanyak 400 menghasilkan accuracy sebesar 89%. Menggunakan data latih sebanyak 700 mendapatkan akurasi sebesar 89% kemudian data latih sebanyak 900 menghasilkan akurasi sebesar 91%.

Metode yang akan digunakan dalam penklasifikasian komentar youtube pada peneleitian ini adalah metode naïve bayes. Model ini sangat populer sebab kesederhanaan yang memungkinkan semua atribut berkontribusi sama pada keputusan akhir. Teknik ini sangat menarik dan relevan pada banyak aspek meskipun teknik ini sederhana.[13] Selain itu pada proses komparasi algoritma naïve bayes dengan svm pada penelitian yang dilakukan oleh[8] menghasilkan naïve bayes lebih unggul dengan nilai akurasi 86% memiliki selisih sebesar 2,4% dengan metode svm.

Dari sekian video tentang pondok pesantren Al-Zaitun penulis menemukan konten yang memiliki banyak komentar yang mencapai angka 3 ribu komentar, dari sekian banyak komenar tentu akan kesulitan untuk mengelompokkan komentar kedalam kelompok positif, negatif dan juga netral agar dari komentar tersebut dapat menjadi bahan evaluasi dan saran kepada konten creator sehingga perlu kiranya penulis mengangkat Analisis Sentimen Komentar Pada Al-Zaitun didalam penerapan penelitian ini.

1.2. Identifikasi Masalah

Berdasarkan latar belakang diatas dapat diidentifikasi beberapa masalah, berikut adalah masalah yang diidentifikasi pada penelitian ini yaitu:

- a. Sulitnya pengklasifikasian data komentar kedalam 3 bagian yaitu: Positif. Netral dan negatif tanpa bantuan sistem.
- b. Pembuktian bahwa metode naïve beyes merupakan metode yang tepat dalam penelitian ini.

1.3. Rumusan Masalah

Berdasarkan uraian diatas baik dari latar belakang atau pada identifikasi masalah, maka rumusan masalahnya adalah sebagai berikut:

1. Bagaimana penerapan metode naïve bayes pada proses klasifikasi komentar youtube?
2. Bagaimana tingkat akurasi penerapan metode naïve bayes dalam mengklasifikasi komentar youtube?

1.4. Batasan Masalah

Agar tidak terjadi penyimpangan dalam proses penelitian ini maka akan ditetapkan Batasan masalah, berikut merupakan Batasan masalah dalam penelitian ini:

1. Data komentar di kelompokkan kedalam tiga bagian yaitu positif, netral dan negative.
2. Didalam penklasifikasian menggunakan metode naïve bayes.

1.5. Tujuan Penelitian

Berdasarkan rumusan masalah diatas maka dapat ditemukan solusi atau jawabanya, adapun tujuan penelitian adalah sebagai berikut:

1. Penerapan metode naïve bayes dalam analisis sentimen komentar youtube.
2. Mengetahui tingkat akurasi klasifikasi dalam penerapan metode naïve beyes.

1.6. Manfaat Penelitian

Manfaat penelitian merupakan hasil yang dapat dirasakan oleh semua pihak yang terkait didalam proses penelitian. Berikut adalah manfaat penelitian:

1. Bagi penulis

Manfaat penelitian bagi penulis yaitu memberikan pengalaman dan menambah wawasan tentang text mining dalam analisis sentiment komentar youtube.

2. Bagi pengguna youtube dan youtuber atau konten creator

bagi para viewer memberikan informasi dalam menilai sebuah konten berdasarkan ekstraksi data komentar youtube. Dan bagi konten creator membantu untuk dapat memudahkan dalam menyimpulkan ribuan komentar. Dan dapat dilihat kualitas dari konten yang dibuat.

1.7. Metode Penelitian

1.7.1. Jenis Penelitian

a. *Library Research*

Library research, juga dikenal sebagai penelitian kepustakaan, merupakan suatu metode penelitian yang dilakukan dengan mengumpulkan dan menganalisis data dari berbagai sumber literatur atau bahan pustaka yang relevan dengan topik penelitian. Tujuan utama dari *library research* adalah untuk memperoleh pemahaman yang mendalam tentang suatu topik melalui tinjauan literatur yang komprehensif. Dalam hal ini peneliti mengambil referensi dari paper dari beberapa jurnal, serta beberapa lirtasi lainnya terkait dengan penelitian.

b. *Eksperimen studi kasus*

adalah penelitian yang dilakukan dengan menerapkan beberapa tindakan untuk membuktikan suatu konsep. Metode ini membangun elemen-elemen dari dua metode yaitu eksperimen dan penelitian studi kasus.

1.7.2. Teknik Pengumpulan Data

a. Pengumpulan Data dan Analisis

Data yang digunakan pada penelitian ini merupakan data komentar youtube yang diambil dengan bantuan API youtube yang kemudian akan di pisah menjadi sentimen positif, netral dan negatif. Data komentar youtube tersebut diambil pada chanel Official iNews dalam tayangan youtube yang berjudul Terbaru! Temuan Penyimpangan dan Masalah Ahklak di Ponpes Al Zaytun #short.

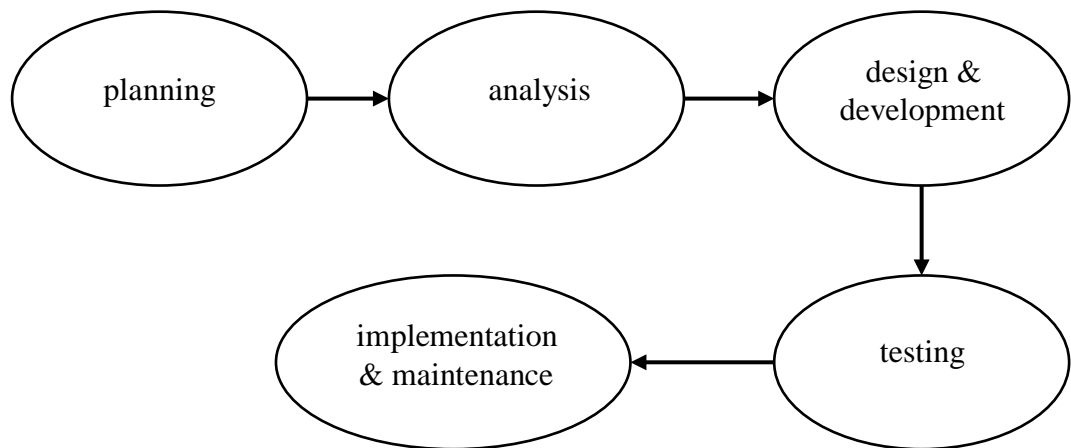
b. Studi Pustaka

Pengumpulan data dilakukan dengan mencari mengambil informasi dari buku, jurnal, skripsi yang berkaitan dengan permasalahan yang sedang diteliti.

1.7.3. Metode Pengembangan Sistem

Metode pengembangan sistem yang dipakai dalam penelitian ini disebut metode *Web Development Life Cycle* (WDLC). *Web Development Life Cycle* diperkenalkan oleh[21], metode ini adalah salah satu turunan dari *Software Development Life Cycle* (SDLC), serta berfokus kepada pengembangan dan memelihara aplikasi berbasis web. Struktur dasar yang dicakup oleh metode ini mampu memenuhi persyaratan untuk menyelesaikan sebuah aplikasi berbasis web. Metode ini juga bisa dimodifikasi dengan memasukkan pengulangan terhadap masing masing proses (Iteration) dan memasukkan metode prototyping.[22]

Seperti yang kita ketahui didalam fase ini terdapat beberapa tahapan yang kita lakukan, berikut adalah lima fase dalam *Web Development Life Cycle* (WDLC), *planning, analysis, design and development, testing, implementation and maintenance*. Yang akan di petakan dengan menggunakan gambar Sebagaimana gambar 1.1



Gambar 1. Web development live cycle (WDLC) model

a. Planning

Tahap pertama adalah perencanaan, tahap ini merupakan tahap yang sangat penting untuk dilakukan sebelum membuat sistem. Dalam tahap ini akan dilakukan identifikasi tujuan situs web, pemahaman tentang target pengguna, analisis kebutuhan bisnis, dan pengumpulan persyaratan. Pada tahap ini, juga dapat membangun rencana proyek, mengalokasikan sumber daya yang sesuai, dan merancang struktur situs website.

b. Analysis

Pada tahap ini, identifikasi terhadap kebutuhan pengguna dengan cara mengumpulkan sejumlah informasi, menganalisa fungsi dari sistem yang akan dibuat, serta hasil apa yang ingin didapatkan dari sistem. Lalu mulailah dipertimbangkan fitur apa saja yang nantinya akan dimuat didalam website yang akan dibangun.

c. Desain and Development

Setelah tahap analisis, selanjutnya akan dipersiapkan sebuah blueprint dari website. Selanjutnya dipersiapkan pula diagram dari objek logis dan fisik agar bisa dikembangkan dalam tahap perencanaan, maksudnya membuat model

data, model proses, dan model tampilan. Kemudian sistem akan dibuat dokumentasinya.

d. Testing

Tahap ini akan dilakukan pengujian untuk menunjukkan hasil kerja dari website yang telah dibangun. Disini akan terlihat apakah sistem yang telah dibuat sesuai dengan kebutuhan pengguna. Komponen yang diuji dapat mencakup fungsi, usability, konten, dan akurasi sistem.

e. Implementation and Maintenance

Disini, Website yang dibangun sudah bisa dioperasikan dan diletakkan dalam komputer pengguna untuk digunakan sebagaimana kebutuhannya. Pengguna juga bisa langsung berinteraksi dengan website tersebut. Maintenance diperlukan apabila dikemudian hari perlu ada perbaikan atau pemeliharaan, sehingga website bisa selalu diperbaharui, dan bisa digunakan dengan maksimal dan berfungsi dengan baik.

1.8. Sistematika Pembahasan

Sistematika skripsi kali ini, penulis membagi beberapa pokok masalah yang kemudian akan diuraikan kedalam 5 BAB.

1. BAB I PENDAHULUAN

Pada bab ini menjelaskan tentang latar belakang, identifikasi masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian, metode pengembangan sistem, dan sistematika pembahasan.

2. BAB II : TINJAUAN PUSTAKA

Pada bab ini menjelaskan tentang penjelasan referensi penelitian terdahulu yang dijadikan acuan serta teori-teori yang diangkat pada penelitian.

3. BAB III : PERANCANGAN DAN IMPLEMENTASI SISTEM

Pada bab ini menjelaskan tentang analisa terhadap kebutuhan sistem, algoritma yang dipakai dalam sistem, perancangan sistem, dan implementasi sistem.

4. BAB IV : HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan tentang hasil yang didapat dari sistem, temuan yang didapatkan serta keterbatasan dalam penelitian.

5. BAB VI : PENUTUP

Pada bab ini menjelaskan tentang simpulan menyeluruh dari hasil penelitian dan saran-saran untuk pengembangan lebih lanjut.



BAB II

TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

“Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes”

Pada penelitian ini menjelaskan terkait penerapan metode naïve bayes terhadap komentar youtube tentang ceramah ustad Abdus somad kemudian sedikit menyinggung tentang Sosial. Youtube merupakan salah satu media sosial yang terkenal di Indonesia dengan 93,8% pengguna. Youtube juga dimanfaatkan sebagai media Dakwah seperti yang dilakukan oleh Ustadz Abdul Somad. Beliau sering mengunggah video yang membahas berbagai jenis persoalan agama khususnya pada bidang hadist dan fiqih. Pengguna Youtube dapat memberikan feedback berupa like, dislike dan komentar terhadap video yang ditayangkandari opini masyarakat yang memiliki hak kebebasan sehingga sering kali sulit untuk penklasifikasian komentar tersebut. Penulis bertujuan untuk menganalisis sentiment masyarakat terhadap Ustadz Abdul Somad melalui komentar youtube menggunakan algoritma Naïve Bayes. Penelitian ini menggunakan 1000 komentar dari 10 video yang ada di Youtube mengenai Ustadz Abdul Somad. Naïve Bayes merupakan algoritma yang sederhana, namun memiliki akurasi yang tinggi dan dapat digunakan pada data yang sedikit. Berdasarkan hasil penelitian, didapatkan sebanyak 67% berkomentar positif, 27% berkomentar netral dan 6% berkomentar negatif. Berdasarkan pengujian didapatkan akurasi sebesar 87%, presisi 91% dan recall 97%. Berdasarkan pengujian tersebut dapat disimpulkan bahwa penelitian ini dapat digunakan untuk hasil sentimen dengan cepat di Youtube.[12]

“Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan Metode Gaussian Naïve bayes”

Pada penelitian ini penulis mengawali dengan menceritakan dari pengguna dari youtube yang semakin meningkat dan juga menjadi salah satu platform penyedia informasi, salah satunya yakni kenaikan harga minyak mentah dunia

hingga berada di atas US\$100 per barel. Berdasarkan permasalahan tersebut, penulis melakukan penelitian terkait analisis sentimen dari komentar pengguna Youtube mengenai kenaikan harga BBM menggunakan metode Gaussian naïve bayes. Percobaan dilakukan menggunakan 3053 dataset dengan pelabelan menggunakan lexicon dan split data 8:2. Penerapan vektorisasi kata menggunakan word embedding Fasttext dan Bag of word sebagai pembanding terhadap akurasi. Percobaan dilakukan dengan kombinasi perbedaan dimensi size pada proses pembuatan language model fasttext. Berdasarkan hasil penelitian yang telah dilakukan, didapatkan nilai akurasi tertinggi pada percobaan dengan dataset tanpa filtering stopwords dan model fasttext size 100 dengan akurasi sebesar 74%. Berdasarkan hasil evaluasi, sistem yang dibangun dapat mengklasifikasikan sentimen atau opini publik ke dalam sentimen positif dan sentiment negatif secara otomatis.[9]

“Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naïve Beyes dan Pembobotan TF-IDF Berdasarkan Komentar Pada Youtube”

Pada penelitian ini penulis menjelaskan tentang Kebijakan Kampus Merdeka yang menjadi salah satu kebijakan baru yang digagas oleh Menteri Pendidikan dan Kebudayaan Republik Indonesia. Kemudian dari situlah Kebijakan tersebut tengah ramai disorot publik khususnya pada platform Youtube berkaitan dengan video unggahan Mendikbud di kanalnya. Pada Youtube, opini masyarakat dapat membanjiri kolom komentar dalam sekejap karena kemunculannya sebagai platform pertama yang menawarkan fasilitas konten audio visual. Penelitian ini mencoba menganalisis opini masyarakat yang tertampung dalam kolom komentar Youtube ke dalam klasifikasi sentimen positif dan negatif. Klasifikasi dan panulis mengimplementasikan pada Google Colaboratory yang berbasis bahasa Python dan Jupyter Notebook dengan algoritme Naive Bayes Classifier dan klasifikasi. Hasil akurasi terbaik sebesar 97% yang didapat dengan menggunakan 900 data latih, 100 data uji, menerapkan pembobotan TF-IDF, dan 10-fold cross validation. Rata-rata akurasi yang didapat dari 10 iterasi pada k-fold cross validation yaitu sebesar 91.8% dengan nilai precision, recall, f-

measure sebesar 90.35%, 93.6%, 91.95%. sehingga dengan hasil tersebut, Naive Bayes Classifier cukup baik sebagai alternatif untuk analisis sentimen.[10]

2.2. Landasan Teori

a. *Text mining*

Text mining adalah suatu proses untuk mengekstraksi informasi yang berguna dari teks yang tidak terstruktur. Tujuan utama *text mining* adalah untuk mengubah teks yang didapat dari beragam sumber menjadi bentuk yang dapat dipahami oleh komputer dan kemudian menerapkan metode analisis dan pemodelan untuk mendapatkan informasi yang berguna dan berharga. [6] [9].

Text mining melibatkan sejumlah teknik dan algoritma yang digunakan untuk mengolah, menggali, dan menganalisis teks. Beberapa teknik umum yang digunakan dalam *text mining* termasuk:

1. *Preprocessing*: Tahap pra-pemrosesan yang melibatkan langkah-langkah seperti menghapus tanda baca, mengubah teks menjadi huruf kecil, menghilangkan kata-kata umum (*stopword* s), dan melakukan normalisasi seperti *Labelling* atau lemmatisasi.[13]
2. Pemberian bobot (*Weighting*): Memberikan bobot pada kata-kata dalam teks untuk menggambarkan tingkat pentingnya kata dalam dokumen. Salah satu metode yang umum digunakan adalah TF-IDF (*Term frequencyInverse document frequency*).[13]
3. Analisis sentimen: Mengidentifikasi dan mengekstraksi sentimen atau sikap yang terkandung dalam teks, seperti positif, negatif, atau netral.

Ini berguna dalam analisis opini atau evaluasi dalam teks, seperti ulasan pelanggan atau media social.

4. Pengelompokan (Clustering): Mengelompokkan dokumen-dokumen berdasarkan kesamaan mereka dalam hal topik, tema, atau karakteristik lainnya. Metode seperti k-means clustering sering digunakan dalam analisis teks.
5. Klasifikasi: Mengklasifikasikan teks ke dalam kategori atau label tertentu berdasarkan fitur-fitur atau pola-pola yang ada. Contohnya adalah klasifikasi email menjadi kotak masuk atau spam.

b. Analisis

Analisis adalah tahap dimana data-data dari komentar yang telah diproses dan diteliti secara detail dan sesuai dengan prosedur yang ditetapkan.[11] Kegiatan merupakan salah satu proses untuk mencari masalah dan data-data yang diperlukan didalam proses penelitian.

c. Sentimen

Merupakan suara masyarakat yang dituangkan kedalam komentar sebagai wujud untuk memberikan pendapat baik itu persetujuan atau penolakan terhadap konten youtube.[12] Sentiment ini merupakan sumber data yang akan diolah pada penelitian kali ini yang nantinya akan diproses menggunakan google collab, dan akan melewati beberapa tahapan didalamnya.

d. Analisis Sentimen

Merupakan solusi untuk mengelompokkan data komentar kedalam kelompok positif, netral dan negative.[5] Selain itu analisis sentiment juga

dapat didefinisikan sebagai proses analisis teks digital untuk menentukan nilai atau sikap dari suatu teks apakah bernilai positif, netral dan negatif.

e. Komentar

Komentar merupakan salah satu aspek penting yang ikut serta dalam mendatangkan kepuasan tersendiri.[14] Tanggapan atau ulasan terhadap sesuatu baik itu berupa pendapat, saran, atau ketidaksetujuan kepada sesuatu. Dalam penelitian ini komentar disesuaikan dengan studi kasus yang diambil yakni tanggapan pada suatu konten berupa video youtube.

f. Youtube

Youtube merupakan salah satu media sosial yang didalamnya berisi tentang konten-konten baik dari hiburan, music, berita terkini dan masih banyak lagi konten-konten di dalamnya. lebih-lebih pemanfaatan youtube pada bidang Pendidikan terutama ketika masa covid-19.[3]

g. Naïve Bayes

Metode ini dikemukakan oleh ilmuwan inggris dapat diartikan salah satu metode klasifikasi dengan menggunakan probabilitas dan data statistik. Model ini sangat populer sebab kesederhanaan yang memungkinkan semua atribut berkontribusi sama pada keputusan akhir.[9] Teknik ini sangat menarik dan relevan pada banyak aspek meskipun teknik ini sederhana. Adapun rumus Teorema Bayes adalah sebagai berikut:

$$P(Q) \frac{p(A)*p(A)}{p(Q)} \dots\dots\dots (2. 1)$$

Keterangan:

$P(A|Q)$: Probabilitas A terhadap kondisi Q

$P(Q|A)$: Probabilitas Q terhadap kondisi A

$P(A)$: Probabilitas A

$P(Q)$: Probabilitas Q

h. Bag of Word

Didalam proses pelabelan ini tentu agar lebih mudah kami menggunakan metode pendekatan Bag- of- Word (BoW) dimana pendekatan ini sering digunakan didalam penelitian-penelitian lain, model ini merupakan model manual dan bertujuan untuk mencocokkan dan menghitung setiap elemen dalam dokumen, dan mencari korelasi informasi di antara kata-kata. Sehingga proses pelabelan lebih akurat.[15]



BAB III

PERANCANGAN SISTEM

3.1. Analisis Kebutuhan

Sebelum merancang dan mengimplementasikan analisis sentiment ini, tentunya penting untuk melakukan analisis kebutuhan yang mendalam. Analisis kebutuhan ini akan membantu dalam mengidentifikasi persyaratan sistem dan memastikan bahwa sistem yang dibangun sesuai dengan kebutuhan yang ada. Berikut adalah beberapa aspek yang perlu diperhatikan dalam analisis kebutuhan:

3.1.1. Analisis Kebutuhan Sistem

Disini akan dianalisis apa saja kebutuhan yang harus dipersiapkan dalam melakukan analisis sentiment pada penelitian kali ini. Hasil analisis digunakan sebagai gambaran untuk mendukung proses implementasi. Hasil analisis ini juga digunakan sebagai dasar untuk membangun sistem.

Kebutuhan sistem disini akan dibagi menjadi dua bagian, perangkat lunak dan perangkat keras. Berikut merupakan rincian kebutuhan yang dibutuhkan:

a. Perangkat Keras

Perangkat keras adalah seluruh komponen atau unsur peralatan yang digunakan untuk menunjang pembangunan aplikasi. Spesifikasi minimum yang dibutuhkan untuk menyelesaikan pembuatan sistem analisis kemiripan skripsi adalah sebagai berikut:

Tabel 3. 1**Spesifikasi Perangkat Keras Yang Digunakan**

No	Kriteria	Spesifikasi
1.	Harddisk	250 GB SSD
2.	Processor	Intel(R) Core(TM) i5-3320M CPU @ 2.60GHz
3.	RAM	12 GB

Tabel diatas merupakan perangkat keras yang digunakan dalam melakukan analisis sentiment pada penelitian ini.

b. Perangkat Lunak

Perangkat lunak yang digunakan untuk membangun sistem ini adalah:

Tabel 3. 2**Perangkat Lunak Yang Digunakan**

No	Jenis Aplikasi	Nama Aplikasi
1.	Bahasa Pemrograman	Phyton
2.	Code Editor	Google Colab
3.	API Youtube	Google cloud
4.	Sistem operasi	Windows 10

3.1.2. Analisa Proses

Pada tahap ini dilakukan proses analisis terhadap apa yang sistem butuhkan. Proses ini penting untuk dilaksanakan karena bisa memastikan kesesuaian dengan tujuan dibuatnya sistem, menghindari kesalahan serta effort ke tempat yang salah, serta bisa meningkatkan efisiensi dan produktivitas.

Berdasarkan program kerja yang digunakan dalam bidang *text mining*, dapat diidentifikasi beberapa komponen yang harus diselesaikan. Komponen pertama business understanding atau memahami permasalahan apa yang akan diselesaikan, pada tahapan ini sangat penting sekali karena hal ini sangat berpengaruh

keberhasilan dari proses yang dilakukan, dalam penelitian ini kami mengambil topik pembahasan yang lumayan hangat yaitu video tentang pondok pesantren al-zaitun pada salah satu chanel yang memiliki pengikut yang lumayan banyak.

Setelah memahami yang diteliti maka kemudian kita melakukan crawling data atau pengambilan data komentar pada video yang telah ditentukan dengan menggunakan google colab dan juga dengan memiliki API youtube sehingga dapat mengambil data komentar yang dibutuhkan

Setelah melakukan pengambilan data atau crawling data maka kemudian dilanjutkan dengan labelling data, labelling data bertujuan untuk mengetahui bagai mana opini yang terkandung dalam text. Sehingga dengan proses ini kita dapat mengetahui opini tersebut masuk kedalam opini positif, netral dan negatif.

Setelah selesai maka dilaksanakanlah komponen *preprocessing* ke dalam dataset yang sudah disediakan. *Preprocessing* merupakan serangkaian tahapan untuk mengubah data teks mentah menjadi bentuk yang lebih terstruktur dan siap untuk diproses lebih lanjut. Tahapan dalam *preprocessing* meliputi *text normalizer*, *tokenizing*, *stopword removal* dan *labelling*.

Tahapan pertama dalam *preprocessing* adalah *text normalizer*. *Text normalizer* bertujuan untuk mengubah teks ke dalam bentuk standar, misalnya dengan mengkonversi huruf kapital menjadi huruf kecil dan menghilangkan karakter khusus. Dengan melakukan normalisasi teks, kita dapat menjadikan teks yang lebih bersih dalam dataset.

Tahap selanjutnya adalah *tokenizing*, yaitu proses membagi teks menjadi unit-unit yang lebih kecil yang disebut token. Token dapat berupa kata, frasa, atau karakter tergantung pada tujuan analisis. Pemisahan teks menjadi token

memungkinkan kita untuk menganalisis dan memanipulasi bagian-bagian individu dengan lebih mudah. Nantinya tokenisasi yang dipakai adalah perkata.

Setelah melakukan *tokenizing*, tahapan berikutnya adalah *stopword removal*. *Stopword* adalah kata-kata umum yang sering muncul dalam teks namun tidak memiliki nilai informasi yang signifikan, seperti kata penghubung atau kata bantu. Dalam tahap ini, *stopword* dihapus dari dataset untuk mengurangi dimensi dan meningkatkan efisiensi proses analisis nantinya.

Selanjutnya adalah *labelling*, yaitu memberikan label baik itu positif, negative dan netral. Dengan melalui tahapan *preprocessing* ini, dataset telah disiapkan dalam bentuk yang lebih terstruktur dan siap untuk diolah lebih lanjut. *Preprocessing* adalah langkah penting dalam pengolahan data teks yang membantu meningkatkan kualitas dan efisiensi analisis yang akan dilakukan.

3.1.3. Analisis Metode

Disini akan dianalisis apa saja metode atau langkah langkah terkait dengan penerapan metode *text mining* yang akan dipakai pada analisis sentiment komentar youtube. Metode yang akan dipakai adalah *preprocessing*, pembobotan, dan *similarity measurement*.

a. *Preprocessing*

Preprocessing meliputi langkah-langkah seperti *text normalizer*, tokenisasi, penghapusan *stopword*, dan *labelling*. Tujuannya adalah untuk membersihkan dan mengubah teks menjadi representasi yang lebih terstruktur agar dapat diolah lebih lanjut.

b. Pembobotan

Pembobotan atau *weighting* meliputi penggunaan metode TF-IDF (*term frequency-inverse document frequency*) untuk memberikan bobot pada

katakata dalam dokumen berdasarkan frekuensi dan relevansinya dalam korpus. Pembobotan ini membantu mengidentifikasi kata-kata yang lebih penting atau mempunyai karakteristik dalam setiap dokumen.

c. Metode Pendekatan Kata

Bag Of Word (BoW) merupakan cara atau metode pendekatan yang paling sederhana dalam mengextrak fitur dari teks, metode ini mengubah teks menjadi matriks kemunculan kata-kata dalam dokumen didalam kalimat.

d. Pemodelan

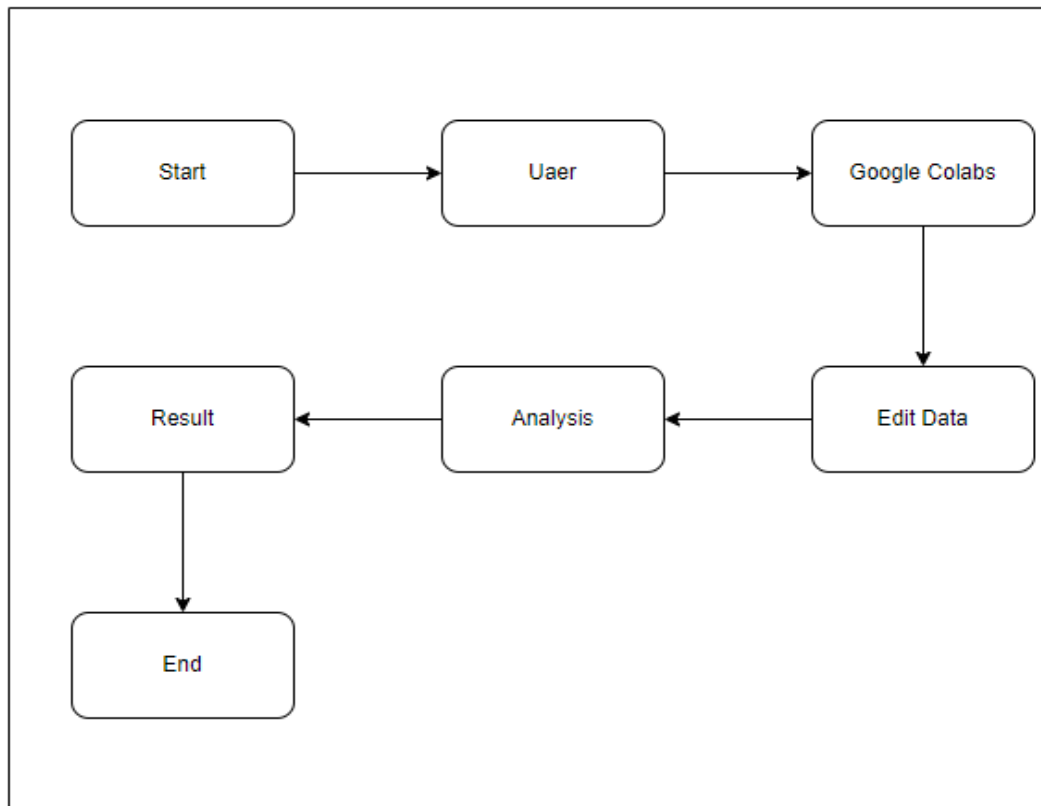
Metode yang digunakan pada proses ini adalah metode yang umum digunakan didalam analisis sentiment yaitu algoritma Naïve bayes. naïve bayes adalah Model ini sangat populer sebab kesederhanaan yang memungkinkan semua atribut berkontribusi sama pada keputusan akhir. Teknik ini sangat menarik dan relevan pada banyak aspek meskipun teknik ini sederhana.

3.2. Rancangan Sistem

Rancangan sistem analisis sentiment komentar youtube merupakan langkah penting dalam membangun sistem yang efektif dan efisien. Rancangan ini akan menggambarkan struktur, komponen, dan interaksi antar elemen pada sistem. Berikut adalah beberapa hal yang perlu diperhatikan dalam rancangan sistem analisis kemiripan:

3.2.1. Arsitektur Sistem

Berikut adalah gambaran tahapan analisis sentiment komentar youtube, yang menggabungkan formal dan representasi suatu system, yang diorganisasikan sedemikian rupa sehingga mendukung penalaran tentang struktur dan perilaku sistem:

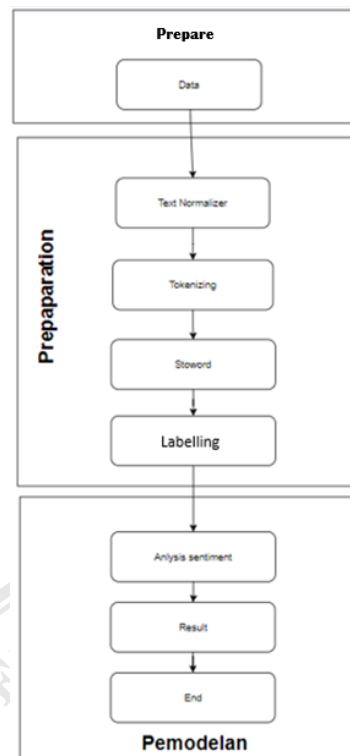


Gambar 3.1
Alur Penggunaan Sistem

Dari gambar diatas dapat dijkaskan bahwa sistem nantinya akan bisa dijangkau selama memiliki akses ke internet. Lewat website yang disediakan oleh google, yang nantinya didalam system tersebut kita bisa juga melakukan edit data yaitu tahapan-tahapan yang sudah dijelaskan diatas seperti contoh preprocessing, setelah data sudah di edit maka kita bisa melakukan proses analisis dengan metode yang sudah ditentukan, kemudain kita bisa melihat result atau hasil dari system tersebut.

3.2.2. Arsitektur Perhitungan

Arsitektur dari analisis sentiment komentar youtube ini terdiri dari tiga tahap utama yaitu tahap *preparation*, tahap *preprocessing* dan tahap Pemodelan. Arsitektur dari sistem ini dapat dilihat pada gambar 3.2.



Gambar 3. 2

Arsitektur Sistem

1. Preparation

Dalam mempersiapkan analisis sentimen terhadap komentar YouTube mengenai Pondok Pesantren Alzaitun menggunakan metode Naive Bayes, langkah awalnya adalah mengambil data melalui API YouTube dengan mendaftarkan aplikasi untuk mendapatkan API Key, kemudian menggunakan endpoint seperti "comments.item" untuk mengambil komentar-komentar yang relevan dengan kata kunci atau ID video terkait, menangani paginasi untuk memastikan semua komentar terambil, lalu membersihkan data dari duplikat dan spam serta melakukan normalisasi teks. Setelah itu, menerapkan metode Naive Bayes untuk klasifikasi sentimen, membuat visualisasi hasil analisis, dan menginterpretasikan temuan dalam sebuah laporan yang rinci.

Dataset diambil dari kometar youtube dengan menggunakan API

Youtube. Sebagaimana gambar berikut:

index	author	published_at	updated_at	like_count	text
0	@syarifudinpanjalu244	2024-07-31T14:30:33Z	2024-07-31T14:30:33Z	0	Jurnalis ya oon, tau aja belum sudah bilang sesat, tabayun dulu neng, biar jelas
1	@WaljirahJirah-xg4gn	2024-07-03T07:45:59Z	2024-07-03T07:45:59Z	0	Biasakan tabayyun dan klarifikasi, berbeda belum tentu salah
2	@rohana5711	2024-05-31T11:53:03Z	2024-05-31T11:53:03Z	0	Bakar bakar masjid rumah dan yang mengikuti ajaran itu
3	@hasantahitu8357	2024-05-13T06:28:45Z	2024-05-13T06:28:45Z	0	Mentri nya kerjanya apa. gak guna
4	@surayazukhruf2006	2024-03-16T20:58:19Z	2024-03-16T20:58:19Z	0	Bgtu mudahnya nyesat2in org. masyaAllah..
5	@user-pr9dq7zk7r	2024-03-18T11:42:09Z	2024-03-18T11:42:09Z	0	Bubarkan..... Ajaran gila
6	@kavianmura1754	2024-01-22T08:27:50Z	2024-01-22T08:27:50Z	0	Yuk bisa yuk bubarin
7	@FD.farell	2024-01-04T10:37:29Z	2024-01-04T10:37:29Z	0	Izin ambil soundtrack nya
8	@arisandi3467	2023-12-31T08:56:01Z	2023-12-31T08:56:01Z	0	shaf shalatnya renggang gitu
9	@hendrihamonangan0478	2023-12-27T09:27:21Z	2023-12-27T09:27:21Z	0	Yg menuduh itulah yg menyimpang, main ponis itu tdk sesuai adab juga ahlik dan kaidah islam, sudahkah benar islam anda?, jangan-jangan diri anda lebih menyimpang, mari introspeksi diri, tanggalkan sifat-sifat keakuan, kalau islam indonesia benar kok moralnya rusak, siapa yg salah ummat atau ulama atau ustadz nya, yg pasti ulama dan ustadz indonesia belum ahlinya Qur'an, yg terjadi bukan rahmatan lil alamin tapi malapetaka, tertinggal, terbelakang, miskin, seperti buih di tengah lautan (terombang-ambing), bukankah itu malapetaka?, 72 golongan masuk neraka kecuali yg satu,
10	@dharmoadjhie4438	2023-12-17T16:43:44Z	2023-12-17T16:43:44Z	0	Justru yg menyimpang nyimpang gini ,banyak banget pendukungnya lho Chonk . Makanya di NKRI ini warganya gampang banget berpaling ke ajaran yg baru muncul biarpun itu dilarang di negeri ini . Lihat saja contohnya , pesantren ini - pengikutnya begitu banyaknya ,padahal kan ajarannya menyimpang ,yg kedua ustad ternama yg menganut paham teroris yg menghalalkan membunuh sesama manusia yg berbeda sudut pandang nya . Ini pun tidak luput dari antrian panjang pengikutnya . Padahal ajarannya juga menghasut para pengikutnya agar memerangi kelompok yg dianggap kafir ,karena beda pahamnya termasuk kelompok muslim sendiri yg beda pandangannya . Itulah keistimewaan masarakat pada umumnya .
11	@Judin-ee3tc	2023-11-19T02:39:23Z	2023-11-19T02:39:23Z	0	Di ulang 2 ,model informasi apa ini. 🤔
12	@bobosan2630	2023-11-12T22:18:22Z	2023-11-12T22:18:22Z	0	Perlu undang2 anti zionis wahyudi
13	@user-ww9ot9fq7h	2023-11-01T09:47:56Z	2023-11-01T09:47:56Z	0	Tapi kasusnya hilang tertelang bumi
14	@nasirbae5883	2023-10-31T23:41:19Z	2023-10-31T23:41:19Z	0	Padahal syekh panji gumilang ga beraalah buktinya sekarang gimana kabar beritanya
15	@AufarHafzi	2023-10-31T23:15:46Z	2023-10-31T23:15:46Z	0	Ya Allah hancurkan kaum munafik yg masuk dalam barisan islam
16	@soekardikardi4754	2023-10-28T11:18:50Z	2023-10-28T11:18:50Z	0	Menurut saya agama itu adalah urusan individu, dengan Tuhan. Salah dan benar adalah Tuhan yg memutuskan, kenapa manusia sebagai pengadil ??? . Apa pengetahuan melebihi Tuhan ???. Apakah dia sendiri sudah yakin diakui Tuhan ???...
17	@user-uj2rq5tj8v	2023-10-18T08:49:08Z	2023-10-18T08:49:08Z	0	Memang benar, jika seseorang punya ilmu tapi tidak berakhlak itu sangat sinkronis karena akhlak itu lebih tinggi nilainya daripada ilmu. Sesuai dengan hadits SAWI. Yang artinya:" Sesungguhnya saya diutus hanya untuk menyempurnakan akhlak" mungkin hanya ini dari kami jika ada kesalahan baik dari tutur maupun kata kami ucapkan mohon maaf yang sebesar-besarnya 🙏🙏🙏
18	@user-uj2rq5tj8v	2023-10-18T08:43:32Z	2023-10-18T08:43:32Z	0	Astaghfirullahaladzim.....
19	@Adib714	2023-10-16T15:10:57Z	2023-10-16T15:10:57Z	0	Agama Islam Di buat main2
20	@arissurabaya3022	2023-10-12T01:03:36Z	2023-10-12T01:03:36Z	0	Terlalu bertele tele....sudah terbukti melenceng dari aqidah Islam dari dulu...kok masih saja membentuk team investigasi....kesuweeeeeen.....terkesan ada ketakutan utk membubarkan Al zaytun..... 🙏🙏🙏🙏

Gambar 3. 3

Pengambilan Data Komentar

2. Preprocessing

Tahap *preprocessing* adalah tahap untuk mengolah dataset yang sudah dipersiapkan ditahap sebelumnya. Tahap ini penting karena data akan dibersihkan supaya hasil yang didapatkan bisa lebih maksimal dan punya kinerja yang baik. Pada tahap *preprocessing* dilakukan *text normalizer*, *Tokenizing*, *stopword removal* dan *Labelling*.

Dari gambar diatas terlihat bahwa data tersebut merupakan komentar pada video yang berjudul #terbaru! Temuan penyimpangan ahlak diponpes al-zaitun menggunakan metode naïve beyes.

a. Text normalizer

Tahap ini mengubah text menjadi huruf kecil Dalam bahasa pemrograman Python, terdapat beberapa fungsi yang dapat digunakan untuk melakukan normalisasi teks, seperti mengubah teks menjadi huruf kecil dan menghilangkan tanda baca dan karakter khusus yang tidak diperlukan. Fungsi `lower()` digunakan untuk mengubah semua karakter dalam string menjadi huruf kecil. Fungsi `re.sub()` memanfaatkan regular expression (regex) untuk menghapus tanda baca dan tanda khusus yang tidak diperlukan. Sementara fungsi `strip()` digunakan untuk menghapus spasi dan karakter standar lainnya dari kedua sisi string. Terakhir, fungsi `replace()` dapat digunakan untuk menggantikan beberapa karakter dengan karakter lain dalam sebuah string. Dengan menggunakan kombinasi fungsi-fungsi tersebut, normalisasi teks dalam bahasa pemrograman Python dapat dilakukan dengan efisien. Adapun contohnya:

Tabel 3. 3

Contoh Teks Normalizer

Teks Awal	Hasil
Anak buah dajjal..semoga secepatnya dibubarkan 🙄🙄🙄	anak buah dajjal semoga secepatnya dibubarkan

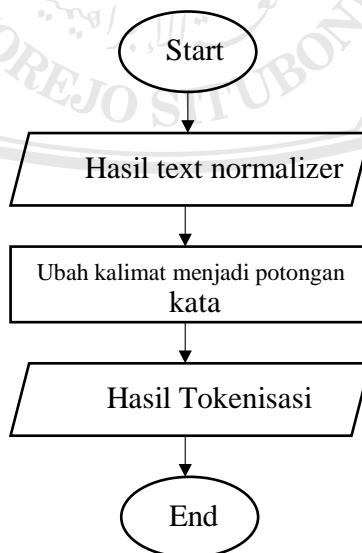
b. Tokenizing

Dalam bahasa pemrograman Python, tokenisasi adalah proses membagi kalimat atau teks menjadi unit-unit yang lebih kecil, seperti kata-kata. Salah satu cara untuk melakukan tokenisasi adalah dengan menggunakan metode ``split()``. Contohnya:

Tabel 3. 4
Contoh Tokenazing

Teks Awal	Hasil
anak buah dajjal semoga secepatnya dibubarkan	['anak' 'buah' 'dajjal' 'semoga' 'cepat' 'bubar']

teks dapat dibagi menjadi token berupa kata-kata dengan menggunakan spasi sebagai pemisah. Selain itu, kita juga dapat menggunakan modul ``nltk`` (Natural Language Toolkit) untuk melakukan tokenisasi yang lebih canggih dengan fungsi ``word_tokenize`` (Dengan cara ini, kita dapat melakukan tokenisasi teks dengan lebih canggih, termasuk pemisahan kata-kata dalam tanda baca tertentu. Seperti contoh pada gambar 3.4.



Gambar 3. 4
Flowchart Proses Tokenizing

c. *Stopword Removal*

Dalam pemrograman Python, stopword adalah kata-kata umum yang sering muncul dalam teks namun tidak memberikan kontribusi signifikan dalam analisis teks. Penghapusan stopword dilakukan untuk meningkatkan efisiensi dan fokus pada kata-kata yang lebih informatif. Untuk menerapkan penghapusan stopword dalam bahasa pemrograman Python, kita dapat menggunakan library Sastrawi. Adapun contohnya sebagai berikut:

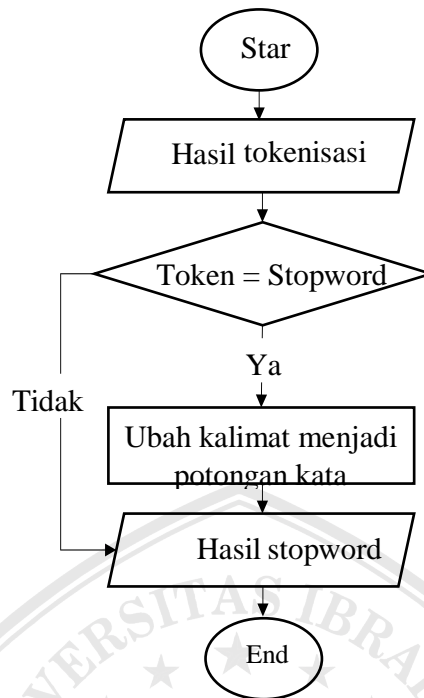
Tabel 3. 5

Contoh Stopword

Teks Awal	Hasil
['anak' 'buah' 'dajjal' 'semoga' 'cepat' 'bubar']	['dajjal' 'semoga' 'bubar']

Untuk memulai, Anda perlu menginstal library Sastrawi melalui pip dengan menjalankan perintah ``pip install Sastrawi`` dalam terminal atau command prompt. Setelah itu, Anda dapat mengimpor modul Sastrawi dan menggunakan ``StopWordRemoverFactory`` untuk membuat objek penghapus stopword. Dengan menggunakan metode ``remove()`` dari objek penghapus stopword, Anda dapat menghapus stopword dari teks yang diberikan. Dengan demikian, Sastrawi memungkinkan pengelolaan stopword dengan mudah dan efisien dalam pemrograman Python.

Untuk memanfaatkan fitur ini, Anda perlu menggunakan `StopWordRemoverFactory`, yang merupakan kelas bawaan dalam library Sastrawi. Kelas ini memungkinkan Anda untuk membuat objek penghapus stopword. Setelah objek ini dibuat, Anda dapat menggunakan metode `remove()` yang disediakan oleh objek tersebut untuk secara otomatis menghapus stopword dari teks yang diberikan, sehingga menghasilkan data yang lebih bersih dan lebih siap untuk dianalisis lebih lanjut



Gambar 3. 5

Flowchart Proses Stopword

Secara umum proses di *stopword removal* adalah dengan membandingkan antara tiap token yang sudah dipecah dengan setiap kata yang termasuk kedalam list *stopword*, apabila kata tersebut terdapat didalam list *stopword*, maka kata tersebut akan dihilangkan. Sedangkan apabila kata tersebut tidak ada didalam list tersebut maka kata akan dikembalikan.

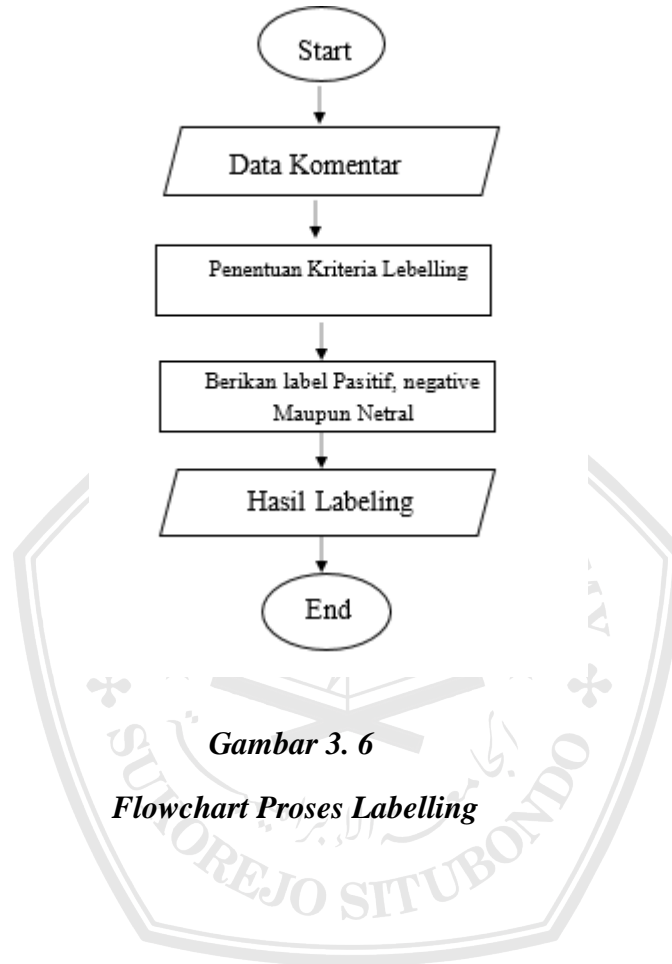
d. **Labelling**

Labeling adalah proses memberikan label atau penanda kepada data berdasarkan kriteria tertentu, seperti mengklasifikasikan teks atau komentar sebagai positif, negatif, atau netral, berikut merupakan contohnya:

Tabel 3. 6 Contoh Labelling

Teks Awal	Hasil
['dajjal' 'semoga' 'bubar']	Negatif

dalam analisis sentimen setelah data diambil, diproses, dan kriteria labeling ditentukan, adapun alurnya sebagai berikut seperti pada gambar 3.6.



Gambar 3. 6
Flowchart Proses Labelling

BAB IV

IMPLEMENTASI SISTEM

4.1. Implementasi Sistem

Implementasi sistem dibuat supaya bisa mengetahui langkah-langkah yang dilakukan saat menerapkan metode-metode yang telah dipilih dalam membangun *information retrieval system* supaya tujuan penelitian tercapai.

4.1.1. Tahap Preparation

Pada tahapan ini kita mengambil data melalui API YouTube dengan mendaftarkan aplikasi untuk mendapatkan API Key, kemudian kita mengambilnya dengan menggunakan *google collab* atau juga bisa menggunakan yang lain seperti *visual code* atau yeditor lainnya, berikut adalah *sourcecodenya*:

Segmen Program 4.1 Mengambil Komentar

```
import googleapiclient.discovery
import pandas as pd

api_service_name = "youtube"
api_version = "v3"
DEVELOPER_KEY = "AIzaSyC01aAXdjJyhqjhaU_w0QKs-18WQ80CPIk"

youtube = googleapiclient.discovery.build(
    api_service_name, api_version,
    developerKey=DEVELOPER_KEY)

request = youtube.commentThreads().list(
    part="snippet",
    videoId="w0EYAjQrOWk",
    maxResults=3000
)
response = request.execute()

comments = []

for item in response['items']:
```

Segmen Program 4.1 (Lanjutan)

```
comment = item['snippet']['topLevelComment']['snippet']
comments.append([
    comment['authorDisplayName'],
    comment['publishedAt'],
    comment['updatedAt'],
    comment['likeCount'],
    comment['textDisplay']
])

df = pd.DataFrame(comments, columns=['author',
    'published_at', 'updated_at', 'like_count', 'text'])

df.head(3000)
```

Dengan itu kita dapat mendapatkan data komentar pada video yang berjudul #terbaru! Temuan penyimpangan ahlak diponpes al-zaitun.

4.1.2. Tahap *Preprocessing*

Pada tahap preparation, ada 4 tahap yang akan dilakukan, yaitu tahap *text normalizer*, tahap *tokenizing*, *stopword removal*, dan *labelling*. *Preprocessing* dilakukan terhadap seluruh kata didalam dokumen, yang dalam hal ini

a) *Text normalizer*

Bisa juga disebut sebagai lowercasing yaitu mengubah kata yang menggunakan huruf kapital menjadi huruf kecil, emoticon dan menghapus karakter tanda baca yang tidak diperlukan. Adapun sourcodenya adalah sebagai berikut:

Segmen Program 4. 2 Teks Normalizer

```
# Step 1: Install necessary libraries
!pip install nltk unicode

# Step 2: Import libraries and download necessary
NLTK data
import re
import nltk
from unicode import unicode

# Download NLTK data (if not already downloaded)
nltk.download('punkt')
```

Segmen Program 4.2 (Lanjutan)

```
nlTK.download('stopwords')

# Step 3: Define text normalization functions
def remove_punctuation(text):
    """Remove punctuation from text."""
    return re.sub(r'^\w\s]', '', text)
def lowercase(text):
    """Convert text to lowercase."""
    return text.lower()

def remove_stopwords(text):
    """Remove stopwords from text."""
    stopwords =
set(nltk.corpus.stopwords.words('english'))
    words = nltk.word_tokenize(text)
    return ' '.join(word for word in words if word
not in stopwords)

def remove_numbers(text):
    """Remove numbers from text."""
    return re.sub(r'\d+', '', text)

def normalize_unicode(text):
    """Normalize unicode characters to ASCII."""
    return unidecode(text)

def normalize_text(text):
    """Normalize text by applying all normalization
functions."""
    text = normalize_unicode(text)
    text = lowercase(text)
    text = remove_punctuation(text)
    text = remove_numbers(text)
    text = remove_stopwords(text)
    return text

# Step 4: Normalize sample text
sample_text = "Hello, world! This is a sample text
with numbers 123 and some punctuations. Let's normalize
it."
normalized_text = normalize_text(sample_text)
print("Original text:", sample_text)
print("Normalized text:", normalized_text)
```

Berikut merupakan hasil dari proses teks normalizer:

Tabel 4. 1 lowercassing

Text	Lowercassing
Anak buah dajjal..semoga secepatnya dibubarkan 🙏🙏🙏 Klw menyimpang, bubarkan saja. Bukan pon pes, koskosan. menyimpang dari ajaran yang sebenarnya	anak buah dajjal semoga secepatnya dibubarkan klw menyimpang bubarkan saja bukan pon pes koskosan menyimpang dari ajaran yang sebenarnya
Bubarkan al zaiton yang paling sesat Wah..Hebat. Pesantren al zaitun menggajari islam dan Yahudi.. 😊	dari pada membuat gaduh bub bubarkan al zaiton yang paling sesat wah hebat Pesantren al zaitun menggajari islam dan Yahudi

b) Tokenazing

Merupakan proses pemisahan teks menjadi unit yang lebih kecil seperti kata atau frasa. Adapun sourcodenya adalah sebagai berikut:

Segmen Program 4. 3 Tokenazing

```
# Step 1: Install NLTK library
!pip install nltk

# Step 2: Import necessary libraries and download
NLTK data
import nltk

# Download NLTK data (if not already downloaded)
nltk.download('punkt')

# Step 3: Define functions for word and sentence
tokenization

def word_tokenize(text):
    """Tokenize text into words."""
    return nltk.word_tokenize(text)

def sentence_tokenize(text):
    """Tokenize text into sentences."""
    return nltk.sent_tokenize(text)

# Step 4: Tokenize sample text
sample_text = "Hello, world! This is a sample text.
It contains multiple sentences. Let's tokenize it."

# Word Tokenization
words = word_tokenize(sample_text)
print("Word tokens:", words)
```

Segmen Program 4.3 (Lanjutan)

```
# Sentence Tokenization
sentences = sentence_tokenize(sample_text)
print("Sentence tokens:", sentences)
```

maka nantinya akan menghasilkan Seperti contoh pada tabel 4.2.

Tabel 4. 2 Tokenazing

Text	Tokenizing
anak buah dajjal semoga secepatnya dibubarkan	['anak' 'buah' 'dajjal' 'semoga' 'cepat' 'bubar']
bukan pon pes koskosan	['bukan' 'ponpes' 'koskosan']
dari pada membuat gaduh bubarkan saja gak usah ribet	['membuat' 'gaduh' 'bubarkan' 'saja' 'usah' 'ribet']
bubarkan al zaiton yang paling sesat wah hebat Pesantren al zaitun menggajari islam dan Yahudi	['bubarkan' 'al zaiton' 'paling' 'sesat'] ['Hebat' 'pesantren' 'al zaitun' 'menggajari' 'islam' 'yahudi']

c) Stopword removal

Tahap stopwords berperan dalam menghilangkan kata-kata yang memiliki tingkat informasi yang rendah dari teks. Contoh hasil dari proses stopwords dapat dilihat dalam tabel dibawah ini. Langkah ini esensial dalam pra-pemrosesan teks karena membantu mengurangi noise dan meningkatkan relevansi informasi yang tersedia untuk analisis lebih lanjut. Dengan menghapus stopwords, fokus analisis lebih tertuju pada kata-kata yang membawa makna penting dalam teks, sehingga menghasilkan hasil yang lebih relevan dan akurat dalam proses analisis teks. Adapun sourcodenya adalah sebagai berikut:

Segmen Program 4. 4 Stopword

```
# Step 1: Install NLTK library
!pip install nltk
```

Segmen Program 4.4 (Lanjutan)

```
# Step 2: Import necessary libraries and
download NLTK dat

import nltk

from nltk.corpus import stopwords

# Download NLTK stopwords (if not already downloaded)
nltk.download('stopwords')

# Step 3: Define function for stopword removal
def remove_stopwords(text):
    """Remove stopwords from text."""
    stop_words = set(stopwords.words('english'))
    words = nltk.word_tokenize(text)
    filtered_words = [word for word in words if
word.lower() not in stop_words]
    return ' '.join(filtered_words)

# Step 4: Remove stopwords from sample text
sample_text = "This is a sample sentence, showing off
the stop words filtration."

filtered_text = remove_stopwords(sample_text)

print("Original text:", sample_text)

print("Filtered text:", filtered_text)
```

Pada Tabel 4.3, hasil dari proses penghapusan stopword ini disajikan untuk memberikan gambaran yang jelas tentang bagaimana teks mentah dapat dibersihkan dari elemen-elemen yang tidak relevan. Setelah stopword dihapus, teks menjadi lebih ringkas dan fokus pada kata-kata kunci yang memiliki makna lebih penting dalam konteks analisis.

Tabel 4. 3 Stopword

Text	Stopword
['anak' 'buah' 'dajjal' 'semoga' 'cepat' 'bubar'] ['menyimpang' 'bubar'] ['bukan' 'ponpes' 'koskosan'] ['menyimpang' 'ajaran' benar'] ['membuat' 'gaduh' 'bubarkan' 'saja' 'usah' 'ribet'] ['bubarkan' 'alzaiton' 'paling' 'sesat']	['dajjal' 'semoga' 'bubar'] ['menyimpang' 'bubar'] ['ponpes' 'koskosan'] ['menyimpang' 'ajaran' benar'] ['membuat' 'gaduh' 'bubarkan'] ['bubarkan' 'alzaiton' 'paling' 'sesat']

d) Labelling

Merupakan proses penentuan teks tersebut termasuk ke dalam teks positif, netral dan negatif. Seperti contoh pada tabel 3.6.

Tabel 4. 4 Labelling

Text	Label
['dajjal' 'semoga' 'bubar']	Negatif
['menyimpang' 'bubar']	Negatif
['ponpes' 'koskosan']	Netral
['menyimpang' 'ajaran' benar']	Negatif
['membuat' 'gaduh' 'bubarkan']	Negatif
['bubarkan' 'alzaiton' 'paling' 'sesat']	Negatif
['dajjal' 'semoga' 'bubar']	Positif

4.1.3 Tahap implementasi

Setelah melewati beberapa proses diatas pad abagian preprocessing maka data sudah siap untuk di ujikan menggunakan metode naïve bayes brikut adalah scrip dari implementasi asalisis sentiment komentar youtube:

Segmen Program 4. 5 Implementasi

```
# Impor library yang diperlukan
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.naive_bayes import MultinomialNB
```

Segmen Program 4.5 (Lanjutan)

```
from sklearn.metrics import classification_report,
accuracy_score

# Impor data komentar yang telah Anda kumpulkan

# Gantilah dengan data komentar Anda
comments = [
{"text": "Smga segra mndapt jln keluarnya dari Allah
utk mngatasi org org alzaitun. In Sha Allah",
"sentiment": "Positive"},
{"text": "Jangan berdamai dengan penjajah...alzaitun
suka...", "sentiment": "Negative"},
{"text": "Jangan berdamai dengan penjajah...alzaitun
suka...", "sentiment": "Negative"},

]

# Konversi data komentar menjadi DataFrame
df = pd.DataFrame(comments)

# Bagi dataset menjadi data latih dan data uji
X = df['text']
y = df['sentiment']
X_train, X_test, y_train, y_test = train_test_split(X,
y, test_size=0.2, random_state=42)

# Vektorisasi teks menggunakan CountVectorizer
vectorizer = CountVectorizer()
X_train_vec = vectorizer.fit_transform(X_train)
X_test_vec = vectorizer.transform(X_test)

# Inisialisasi model Naive Bayes
nb_classifier = MultinomialNB()

# Melatih model Naive Bayes
nb_classifier.fit(X_train_vec, y_train)

# Memprediksi sentimen pada data uji
y_pred = nb_classifier.predict(X_test_vec)
# Evaluasi performa model
accuracy = accuracy_score(y_test, y_pred)
classification_rep = classification_report(y_test,
y_pred)

print(f'Akurasi: {accuracy}')
print('Laporan Klasifikasi:\n', classification_rep)
```

4.2. Hasil Uji Coba

Pada bagian ini, data uji coba dibagi menjadi dua bagian utama: data testing dan data training. Percobaan dilakukan dengan tiga skenario berbeda: pertama, menggunakan 2000 data atau data keseluruhan dan nanti akan dibagi menjadi data training dan data testing. Kedua, menggunakan 1000 data komentar, dan ketiga, menggunakan 1000 data komentar juga. Pembagian ini bertujuan untuk menguji konsistensi dan akurasi sistem dalam mengolah data, serta untuk mengevaluasi performa model dalam konteks pengujian analisis sentiment. Evaluasi dilakukan berdasarkan respons sistem terhadap masing-masing skenario pembagian data.

Untuk mencari nilai pada pengujian ini maka kita akan menerapkan rumus berikut:

a. Precision

Precision adalah ukuran seberapa konsisten atau tepat hasil pengukuran atau observasi yang diulang dalam kondisi yang sama. Precision menunjukkan seberapa dekat nilai-nilai hasil pengukuran satu sama lain, tanpa memperhatikan apakah hasil tersebut mendekati nilai sebenarnya (accuracy).

$$\text{Precision} = \frac{TP}{TP+FP} \dots\dots\dots (4.1)$$

Keterangan:

TP: True Positif

FP: False Positif

b. Recall

Recall adalah ukuran dalam evaluasi model klasifikasi yang menunjukkan kemampuan model untuk menemukan semua contoh positif dalam data. Dalam konteks sederhana, recall mengukur seberapa banyak dari total data positif yang berhasil dikenali dengan benar oleh model.

$$\text{Recall} = \frac{TP}{TP+FN} \dots\dots\dots (4.2)$$

Keterangan:

TP: True Positif

FN: False Negatif

c. F1 Score

F1 Score adalah ukuran yang menggabungkan precision dan recall ke dalam satu metrik harmonis. Ini berguna ketika ada ketidakseimbangan antara kelas positif dan negatif, karena F1 Score mempertimbangkan keseimbangan antara kedua metrik tersebut.

$$F1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \dots\dots\dots (4.3)$$

d. Accuracy

Adalah metrik evaluasi yang mengukur seberapa banyak prediksi yang benar dari keseluruhan data yang diuji. Dalam konteks klasifikasi, akurasi menunjukkan proporsi prediksi yang tepat dibandingkan dengan semua prediksi.

Secara matematis, akurasi dihitung dengan rumus:

$$Accuracy = \frac{TP+TN}{Total Data} \dots\dots\dots (4.4)$$

e. Macro Average

Adalah Rata-rata dari precision, recall dan f1-score untuk semua kelas tanpa memperhitungkan jumlah data dalam setiap kelas berikut rumus dari macro:

$$Macro(precision, recall / f1) = \frac{N+M+O}{3} \dots\dots\dots (4.5)$$

Keterangan:

N: Nilai Kelas dari precision, recall dan f1

M: Nilai Kelas dari precision, recall dan f1

O: Nilai Kelas dari precision, recall dan f1

f. Weighted Average

Weighted average (rata-rata berbobot) adalah metode perhitungan rata-rata di mana setiap nilai memiliki bobot atau kepentingan yang berbeda. Dalam perhitungan ini, nilai dengan bobot lebih besar akan lebih berpengaruh terhadap hasil rata-rata.

$$Weighted Average (precision, recall / f1) = \frac{N \times P}{K} \dots\dots\dots (4.6)$$

Keterangan:

N : Nilai Support Masing-masing Kelas

P : Nilai Kelas dari precision, recall dan f1

K : Nilai Accuracy pada kelas support

Hasil dari pengujian yang telah dilakukan akan disajikan secara sistematis dalam bentuk tabel di bawah ini. Tabel ini berfungsi sebagai representasi data yang memuat informasi penting terkait parameter yang diujikan. Penyajian dalam bentuk tabel memudahkan pembaca dalam memahami hasil pengujian secara komprehensif, karena informasi yang disampaikan lebih terstruktur dan dapat dibandingkan dengan mudah.

4.2.1 Percobaan Pertama

Percobaan pertama dilakukan dengan menggunakan data keseluruhan untuk pembagian data training dan data testing. Dalam proses ini dibagi data tersebut menjadi 2 bagian yaitu 1000 komentar untuk data pelatihan dan 1000 komentar untuk data pengujian. Namun sebelum dilakukan pengujian kami melakukan proses klasifikasi secara keseluruhan yang menghasilkan nilai yang akan ditunjukkan pada tabel berikut:

Tabel 4. 5 Pengujian Pertama

	Precision	Recall	F1-score	Support
Negatif	0,63	0,90	0,74	222
Netral	0,48	0,22	0,30	93
Positif	0,41	0,16	0,24	73
Accuracy	-	-	0,60	338
Macro avg	0,51	0,43	0,42	388
Wighted avg	0,55	0,60	0,54	388

4.2.2 Percobaan Kedua

Percobaan kedua yaitu data training pada percobaan ini kami menggunakan 1000 data komentar yang akan di klasifikasi. Tabel merupakan hasil dari percobaan kedua.

Tabel 4. 6 Pengujian Kedua

	Precision	Recall	F1-score	Support
Negatif	0,56	0,92	0,69	100
Netral	0,41	0,22	0,28	51
Positif	0,86	0,12	0,22	48
Accuracy	-	-	0,55	199
Macro avg	0,61	0,42	0,40	199
Weighted avg	0,59	0,55	0,47	199

4.2.3 Percobaan Ketiga

Setelah mengetahui hasil dari percobaan data testing maka kita lanjut ke percobaan ketiga yaitu data testing pada percobaan ini kami menggunakan 1000 data komentar. Tabel merupakan hasil dari percobaan ketiga.

Tabel 4. 7 Pengujian Ketiga

	Precision	Recall	F1-score	Support
Negatif	0,64	0,93	0,75	113
Netral	0,61	0,27	0,38	62
Positif	0,40	0,14	0,21	28
Accuracy	-	-	0,62	203
Macro avg	0,55	0,45	0,45	203
Wighted avg	0,59	0,62	0,56	203

4.3. Analisis Hasil

Pada tahap ini akan dibahas hasil dari ketiga skenario yang dilakukan sebelumnya, hal ini perlu dilakukan supaya bisa memberikan gambaran sejauh mana faktor faktor yang mempengaruhi hasil klasifikasi analisis sentiment komentar youtube menggunakan metode naïve bayes.

4.3.1 Percobaan Pertama

Hasil percobaan pertama, model klasifikasi menunjukkan bahwa meskipun akurasi keseluruhan mencapai 64%, performa model dalam mengenali kelas dengan memuaskan. Dengan precision dan recall untuk kelas ini mencapai 0.63, untuk data negative, 0,48 untuk data netral dan 0,41 untuk data positif hal

ini mengindikasikan bahwa hasil data data tersebut menunjukkan bahwa keseimbangan antara precision dan recall sangat baik. Model ini menunjukkan performa baik dalam mengklasifikasikan kelas mayoritas, yaitu "Negative". Dengan precision mencapai 0.63 dan recall hampir menuju angka sempurna (0,90), model mampu mengidentifikasi sebagian besar sampel yang seharusnya diklasifikasikan sebagai "Negative". F1-score untuk kelas "Negative" adalah 0.74, mencerminkan keseimbangan yang baik antara precision dan recall.

Model ini menunjukkan bahwa peningkatan signifikan diperlukan untuk mengatasi masalah ketidakmampuan dalam mengenali kelas "Positive" dan "Netral". Karena dari sekian data komentar perbandingan komentar positive, netral dan negative lebih dominan kearah negative dibandingkan komentar yang mengarah ke positive dan netral. Langkah-langkah perbaikan mungkin termasuk penyesuaian model, penambahan data, atau teknik pemrosesan data yang lebih cermat untuk memperbaiki performa klasifikasi pada kelas minoritas tersebut.

4.3.2 Percobaan Kedua

Hasil model klasifikasi menunjukkan akurasi keseluruhan sebesar 86%, yang mengindikasikan bahwa dari sampel percobaan, sebanyak 86% di antaranya diprediksi dengan benar oleh model. Namun, perhatian utama tertuju pada kelas "Positive", di mana hasilnya menunjukkan hasil yang signifikan. Precision 0,86, recall 0,12, dan F1-score 0,22 untuk support 48, Di sisi lain, untuk kelas "Negative", model menunjukkan performa yang baik. Dengan precision mencapai 0.56 dari prediksi yang dilabeli sebagai "Negative" memang merupakan kasus negatif. Recall untuk kelas "Negative" adalah 0.56, menunjukkan bahwa model mampu menemukan sebagian besar dari semua sampel yang seharusnya diklasifikasikan sebagai "Negative". F1-score yang

mencapai 0.69 untuk kelas "Negative" menunjukkan bahwa keseimbangan antara precision dan recall dalam klasifikasi ini relatif baik. kinerja model dalam mengenali kelas "Netral" juga baik, dengan precision mencapai 0,41. Recall untuk kelas "Netral" adalah 0.22, F1-score untuk kelas "Netral" adalah 0.28, Ini menunjukkan bahwa meskipun model hampir selalu benar dalam memprediksi kelas "Netral" saat itu, model hanya mampu mengidentifikasi sebagian kecil dari semua sampel yang seharusnya diklasifikasikan sebagai "Netral".

4.3.3 Percobaan Ketiga

Hasil model klasifikasi menunjukkan bahwa akurasi keseluruhan model adalah 64%, yang berarti 64% dari total percobaan berhasil diprediksi dengan benar. Model menunjukkan kemampuan sempurna dalam mengenali kelas "Positive" dengan precision mencapai 0.40, namun recall dan F1-score untuk kelas ini hanya sebesar 0.14. Ini menunjukkan bahwa meskipun model hampir selalu benar dalam memprediksi kelas "Positive" saat itu, model hanya mampu mengidentifikasi sebagian kecil dari semua sampel yang seharusnya diklasifikasikan sebagai "Positive". Sama halnya dengan kelas "Netral" yang menghasilkan precision sebesar 0.61, recall 0.27 dan f-1 score sebesar 0.38 menunjukkan hasil yang lebih baik dibandingkan dengan kelas "Positif". Di sisi lain, kinerja model dalam mengenali kelas "Negative" lebih baik, dengan precision mencapai 0.64, yang artinya semua prediksi yang dilabeli sebagai "Negative" benar-benar merupakan kasus negatif. Recall untuk kelas "Negative" adalah 0.92, menunjukkan bahwa model berhasil menemukan sebagian besar dari semua sampel yang seharusnya diklasifikasikan sebagai

"Negative". F1-score untuk kelas "Negative" adalah 0.69, mencerminkan keseimbangan yang baik antara precision dan recall dalam klasifikasi ini.

Secara rata-rata (macro average), precision adalah 0.61 dan recall adalah 0.42, sedangkan secara tertimbang (weighted average), precision mencapai 0.59 dan recall mencapai 0.55. Evaluasi ini menunjukkan bahwa meskipun model sangat baik dalam mengenali kelas mayoritas "Negative", masih diperlukan peningkatan signifikan dalam kemampuan untuk mengenali dan memprediksi kelas minoritas "Positive" dan "Netral"

Berdasarkan hasil klasifikasi naïve bayes dengan tiga percobaan, model menunjukkan hasil yang rendah dalam klasifikasi komentar positif dan netral. Meskipun akurasi keseluruhan meningkat dari 0.51 data keseluruhan menjadi 0.61 untuk pengujian data uji dan 0,55 untuk data uji menunjukkan ketidakseimbangan dalam performa model tetap ada. Model ini sangat bias terhadap kelas negatif, menunjukkan perlunya peningkatan metode penanganan ketidakseimbangan kelas untuk meningkatkan pengenalan komentar positif.

4.3. Evaluasi

Berdasarkan hasil dari tiga percobaan menggunakan metode Naive Bayes untuk klasifikasi komentar, kesimpulannya adalah model menunjukkan akurasi yang bervariasi antara 64% hingga 86%.

Tabel 4. 8 Data Hasil Klasifikasi

	Precision	Recall	F1-score	Support
Percobaan Ke 1				
Negatif	0,63	0,90	0,74	222
Netral	0,48	0,22	0,30	93
Positif	0,41	0,16	0,24	73
Accuracy	-	-	0,60	338
Macro avg	0,51	0,43	0,42	388
Wighted avg	0,55	0,60	0,54	388

Tabel Lanjutan 4.8

Percobaan Ke 2				
Negatif	0,56	0,92	0,69	100
Netral	0,41	0,22	0,28	51
Positif	0,86	0,12	0,22	48
Accuracy	-	-	0,55	199
Macro avg	0,61	0,42	0,40	199
Weighted avg	0,59	0,55	0,47	199
Percobaan Ke 3				
Negatif	0,64	0,93	0,75	113
Netral	0,61	0,27	0,38	62
Positif	0,40	0,14	0,21	28
Accuracy	-	-	0,62	203
Macro avg	0,55	0,45	0,45	203
Wighted avg	0,59	0,62	0,56	203

Berdasarkan hasil klasifikasi menggunakan metode Naive Bayes, dapat disimpulkan bahwa meskipun model menunjukkan performa yang kuat dalam mengklasifikasikan kelas Negatif, dengan Precision dan Recall yang baik di semua percobaan, model masih mengalami kesulitan konsisten dalam mengidentifikasi kelas Netral dan Positif, dengan Precision dan Recall yang rendah; oleh karena itu, untuk meningkatkan kinerja model, diperlukan perbaikan melalui feature engineering, penyesuaian model, dan pengumpulan data tambahan untuk kelas-kelas tersebut.

Meskipun model sangat baik dalam mengenali kelas mayoritas "Negative" dengan akurasi tinggi dan keseimbangan yang baik antara precision dan recall, performa untuk kelas minoritas "Netral" dan "Positive" masih rendah. Hal ini terlihat dari nilai recall yang rendah untuk kedua kelas tersebut, menunjukkan bahwa model kesulitan dalam mengidentifikasi sebagian besar sampel yang seharusnya diklasifikasikan sebagai "Netral" dan "Positive". Evaluasi ini menyarankan perlunya penyesuaian lebih lanjut pada model, penambahan data,

atau penerapan teknik pemrosesan data yang lebih cermat untuk meningkatkan kemampuan dalam mengenali kelas minoritas tersebut.



BAB V

PENUTUP

5.1. Kesimpulan

Dengan tingkat akurasi yang baik dan tingkat recall yang tinggi serta nilai presisi yang dapat ditingkatkan, maka metode naïve bayes sangat cocok dalam analisis sentiment pada komentar youtube, Namun perlu diketahui bahwa kualitas nilai ini dapat dipengaruhi oleh kualitas data yang dianalisis. Sehingga evaluasi dan pengembangan lebih lanjut terhadap model tersebut masih diperlukan untuk meningkatkan kualitas secara keseluruhan, sehingga pada data testing menghasilkan nilai accuracy 64%, dengan nilai recall terbesar 93%, Precision terbesar 75% dengan support sebesar 203.

5.2. Saran

Beberapa saran bagi penelitian lebih lanjut adalah sebagai berikut:

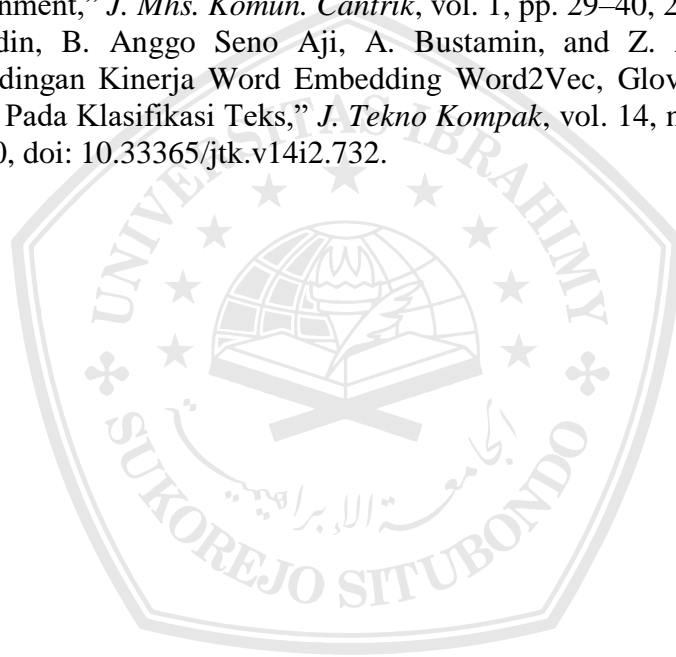
1. Menambahkan fitur atau metode untuk mengelola sinonim dan akronim, sehingga dapat meningkatkan kemampuan system.
2. Bagi peneliti yang ingin menggunakan metode ini pastikan data yang digunakan merupakan data yang representative. Dan semakin banyak variasi dalam pelatihan, maka semakin baik model dalam menggeneralisasi



DAFTAR PUSTAKA

- [1] M. F. Ansyah, P. Kedua, and P. Ketiga, "G-Tech : Jurnal Teknologi Terapan," vol. 8, no. 1, pp. 1–8, 2024.
- [2] I. A. Huda, "Perkembangan Teknologi Informasi Dan Komunikasi (Tik) Terhadap Kualitas Pembelajaran Di Sekolah Dasar," *J. Pendidik. dan Konseling*, vol. 2, no. 1, pp. 121–125, 2020, doi: 10.31004/jpdk.v1i2.622.
- [3] Z. S. K. L. Kohsasih, "Pemanfaatan Media Sosial dalam Proses Pembelajaran di SMPN 1 Mojo Kediri," *Indones. J. Islam. Educ. Stud.*, vol. 2, no. 1, pp. 50–60, 2019, doi: 10.33367/ijies.v2i1.850.
- [4] A. Liawati *et al.*, "ANALISIS SENTIMEN KOMENTAR POLITIK DI MEDIA SOSIAL X DENGAN PENDEKATAAN DEEP LEARNING," vol. 7, no. 6, pp. 3557–3563, 2023.
- [5] N. M. I. Priyanti and Nurhayati, "Penerapan Model Pembelajaran Problem Based Learning Berbantuan Media Youtube Untuk Meningkatkan Hasil Belajar," *J. Ilm. Mat. Realis. (JI-MR)*, vol. 4, no. 1, pp. 96–101, 2023.
- [6] P. Yoga Saputra, D. Hanifudin Subhi, and Z. Zain Afif Winatama, "Implementasi Sentimen Analisis Komentar Channel Video Pelayanan Pemerintah Di Youtube Menggunakan Algoritma Naïve Bayes," *J. Inform. Polinema*, vol. 5, no. 3, pp. 209–213, 2019, [Online]. Available: <http://jip.polinema.ac.id/ojs3/index.php/jip/article/view/259>
- [7] S. A. Anan Septia Machidhar, Maya Sekar Wangi, "EFEKTIVITAS PENGGUNAAN MEDIA SOSIAL YOUTUBE SEBAGAI MEDIA PENYEBARAN INFORMASI PADA CHANNEL YOUTUBE DISKOMINFO BOYOLALI," 2022.
- [8] A. S. Rahayu, A. Fauzi, and R. Rahmat, "Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, p. 349, 2022, doi: 10.30865/json.v4i2.5398.
- [9] S. Mujahidin, B. Prasetyo, and M. C. C. Utomo, "Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan Metode Gaussian naïve bayes," *Voteteknika (Vocational Tek. Elektron. dan Inform.)*, vol. 10, no. 3, p. 17, 2022, doi: 10.24036/voteteknika.v10i3.118299.
- [10] D. F. Zhafira, B. Rahayudi, U. Brawijaya, and P. Korespondensi, "ANALISIS SENTIMEN KEBIJAKAN KAMPUS MERDEKA MENGGUNAKAN NAIVE BAYES DAN PEMBOBOTAN TF-IDF SENTIMENT ANALYSIS OF KAMPUS MERDEKA POLICY USING NAIVE BAYES AND TF-IDF TERM WEIGHTING BASED ON YOUTUBE," vol. 2, no. 1, pp. 55–63, 2021.
- [11] P. Arsi, L. N. Hidayati, and A. Nurhakim, "Komparasi Model Klasifikasi Sentimen Issue Vaksin Covid-19 Berbasis Platform

- Instagram,” vol. 6, pp. 459–466, 2022, doi: 10.30865/mib.v6i1.3509.
- [12] H. A. R. Harpizon, R. Kurniawan, Iwan Iskandar, R. Salambue, E. Budianita, and F. Syafria, “Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes,” *JNKTI (Jurnal Nas. Komputasi dan Teknol. Informasi)*, vol. 5, no. 1, pp. 131–140, 2022, [Online]. Available: <http://repository.uin-suska.ac.id/59746/>
- [13] K. L. Kohsasih and Z. Situmorang, “Analisis Perbandingan Algoritma C4.5 dan Naïve Bayes Dalam Memprediksi Penyakit Cerebrovascular,” *J. Inform.*, vol. 9, no. 1, pp. 13–17, 2022, doi: 10.31294/inf.v9i1.11931.
- [14] A. N. Husna and P. Rianto, “Membaca Komentar di Media Sosial Sebagai Hiburan Reading Comments on Social Media for Entertainment,” *J. Mhs. Komun. Cantrik*, vol. 1, pp. 29–40, 2021.
- [15] A. Nurdin, B. Anggo Seno Aji, A. Bustamin, and Z. Abidin, “Perbandingan Kinerja Word Embedding Word2Vec, Glove, Dan Fasttext Pada Klasifikasi Teks,” *J. Tekno Kompak*, vol. 14, no. 2, p. 74, 2020, doi: 10.33365/jtk.v14i2.732.





**PANITIA TUGAS AKHIR DAN SKRIPSI
UNIVERSITAS IBRAHIMY
FAKULTAS SAINS DAN TEKNOLOGI**

JL. KHR. Syamsul Arifin No. 1-2 PO. Box 2 Phone (0852-3333-7581)
Fax. (0338) 453068 Situbondo 68374 website: www.ibrahimiy.ac.id e-mail : unib2018@ibrahimiy.ac.id
SUMBEREJO BANYUPUTIH SITUBONDO JAWA TIMUR

Nomor : 0828/405.174/071.095/M.3/VIII/2024
Prihal : Permohonan Izin Penelitian

22 Januari 2024

Kepada
Yth. Cabang Youtube

Di Tempat

Assalamu'alaikum warahmatullahi wabarakatuh.

Disampaikan dengan hormat, dalam rangka pelaksanaan penelitian guna penyusunan Skripsi. Judul "**Analisis sentimen komentar youtube terhadap tayangan #Terbaru! temuan penyimpangan dan masalah akhlak di ponpes al-zaytun menggunakan metode naive bayes**" dengan ini kami bermaksud mengajukan permohonan izin penelitian bagi mahasiswa kami :

Nama : MOCHAMMAD FEBRIALDI ANSYAH
NIM : 2020503021
Program Studi : S1 Teknologi Informasi
Waktu : 23 Januari 7 Februari 2024

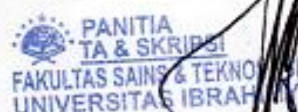
Untuk melakukan penelitian di lembaga yang Bapak/Ibu pimpin. Sebagai bahan pertimbangan bersama ini kami lampirkan :

1. Proposal Skripsi
2. Fotocopy Kartu Tanda Mahasiswa

Demikian permohonan kami, atas berkenannya kami sampaikan *Jazakumullahu khairan.*

Wassalamu'alaikum warahmatullahi wabarakatuh.

Ketua Panitia,



Dr. Ach. Khumaidi, M.P

index	author	published_at	updated_at	like_count	text
0	@syarifudinpanjalu244	2024-07-31T14:30:33Z	2024-07-31T14:30:33Z	0	Jurnal is ya oon, tau aja belum sudah bilang sesat, tabayun dulu neng, biar jelas
1	@WaljirahJirah-xg4gn	2024-07-03T07:45:59Z	2024-07-03T07:45:59Z	0	Biasakan tabayun dan klarifikasi, berbeda belum tentu salah
2	@rohana5711	2024-05-31T11:53:03Z	2024-05-31T11:53:03Z	0	Bakar bakar masjid rumah dan yang mengikuti ajaran itu
3	@hasantahitu8357	2024-05-13T06:28:45Z	2024-05-13T06:28:45Z	0	Mentri nya kerjanya apa .gak guna
4	@tsurayazukhruf2006	2024-03-16T20:58:19Z	2024-03-16T20:58:19Z	0	Bgtu mudahnya nyesat2in org, masyaAllah..
5	@user-pr9dq7zk7r	2024-03-16T11:42:09Z	2024-03-16T11:42:09Z	0	Bubarkan..... Ajaran gila
6	@kavianmura1754	2024-01-22T08:27:50Z	2024-01-22T08:27:50Z	0	Yuk bisa yuk bubarin
7	@FD.farell	2024-01-04T10:37:29Z	2024-01-04T10:37:29Z	0	Izin ambil soundtrack nya
8	@arisandi3467	2023-12-31T08:56:01Z	2023-12-31T08:56:01Z	0	shaf shalatnya renggang gitu
9	@hendrihamonangan6478	2023-12-27T09:27:21Z	2023-12-27T09:27:21Z	0	Yg menuduh itulah yg menyimpang, main ponis itu tdk sesuai adab juga ahlak dan kaidah islam, sudahkah benar islam anda?, jangan-jangan diri anda lebih menyimpang, mari introspeksi diri, tanggalkan sifat-sifat keakuan, kalau islam indonesia benar kok moralnya rusak, siapa yg salah ummat atau ulama atau ustadz nya, yg pasti ulama dan ustadz indonesia belum ahlinya Qur'an, yg terjadi bukan rahmatan lili alamin tapi malapetaka, tertinggal, terbelakang, miskin, seperti buih di tengah lautan (terombang-ambing), bukankah itu malapetaka?, 72 golongan masuk neraka kecuali yg satu,
10	@dharmoadjhie4438	2023-12-17T16:43:44Z	2023-12-17T16:43:44Z	0	Justru yg menyimpang nyimpang gini ,banyak banget pendukungnya lho Chonk . Makanya di NKRI ini warganya gampang banget berpaling ke ajaran yg baru muncul biarpun itu dilarang di negeri ini . Lihat saja contohnya , pesantren ini - pengikutnya begitu banyaknya ,padahal kan ajarannya menyimpang .,yg kedua ustad ternama yg menganut paham teroris yg menghalalkan membunuh sesama manusia yg berbeda sudut pandang nya . Ini pun tidak luput dari antrian panjang pengikutnya . Padahal ajarannya juga menghasut para pengikutnya agar memerangi kelompok yg dianggap kafir ,karena beda pemahannya .termasuk kelompok muslim sendiri yg beda pandangannya . Itulah keistimewaan masarakat pada umumnya .
11	@Judin-ee3tc	2023-11-19T02:39:23Z	2023-11-19T02:39:23Z	0	Di ulang 2 ,model informasi apa ini. 🤔
12	@bobosan2530	2023-11-12T22:18:22Z	2023-11-12T22:18:22Z	0	Perlu undang2 anti zionis wahyudi
13	@user-ww9ot9fq7h	2023-11-01T09:47:56Z	2023-11-01T09:47:56Z	0	Tapi kasusnya hilang tertelang bumi
14	@nasirbae5883	2023-10-31T23:41:19Z	2023-10-31T23:41:19Z	0	Padahal syekh panji gumilang ga beraalah buktinya sekarang gimana kabar beritanya
15	@AufarHafizi	2023-10-31T23:15:46Z	2023-10-31T23:15:46Z	0	Ya Allah hancurkan kaum munafik yg masuk dalam barisan islam
16	@soekardikardi4754	2023-10-28T11:18:50Z	2023-10-28T11:18:50Z	0	Menurut saya agama itu adalah urusan individu, dengan Tuhan. Salah dan benar adalah Tuhan yg memutuskan, kenapa manusia sebagai pengadil ??? . Apa pengetahuan melebihi Tuhan ???.. Apakah dia sendiri sudah yakin diakui Tuhan ???...
17	@user-uj2rq5tj8v	2023-10-18T08:49:08Z	2023-10-18T08:49:08Z	0	Memang benar, jika seseorang punya ilmu tapi tidak berakhlak itu sangat singkronis karena akhlak itu lebih tinggi nilainya daripada ilmu. Sesuai dengan hadits Nabi SAW. Yang artinya:" Sesungguhnya saya diutus hanya untuk menyempurnakan akhlak". mungkin hanya ini dari kami jika ada kesalahan baik dari tutur maupun kata kami ucapkan mohon maaf yang sebesar-besarnya 🙏🙏🙏
18	@user-uj2rq5tj8v	2023-10-18T08:43:32Z	2023-10-18T08:43:32Z	0	Astaghfirullahaladzim.....
19	@Adib714	2023-10-16T15:10:57Z	2023-10-16T15:10:57Z	0	Agama Islam Di buat main2
20	@arissurabaya3622	2023-10-12T01:03:36Z	2023-10-12T01:03:36Z	0	Terlalu bertele tele....sudah terbukti melenceng dari aqidah Islam dari dulu..kok masih saja membentuk team investigasi.....kesuweeeeee.....terkesan ada ketakutan utk membubarkan Al zaytun.....👉👉👉

Pembimbing I :

TANGGAL	CATATAN	PARAF
	Konsultasi bab 1	
	Konsultasi bab 1, 2 Revisi	
	Acc bab 1, 2, Revisi bab 3	
	Revisi bab 3, targetkan content proses pembelajaran.	
	Revisi bab 3 & 4 (targetkan pernyataan pengajaran)	
	Acc bab 3, Revisi bab 7	
	Acc bab 4	
	Acc bab 4 & 5 cek program	

Pembimbing II :

NO	TANGGAL	CATATAN	PARAF
		Bab I	
		Revisi bab I bab II	
		Acc bab 1, 2 revisi bab 3	
		Acc revisi bab 3	
		Acc bab 3 & aplikasi	
		Acc bab 4 & 5	

CURICULUM VITAE PENULIS



Mochammad Febrialdi A. is Me

Lahirnya dengan tangisan kebahagiaan orang tua yang bertempat di Bondowoso pada hari Ahad, 24 Februari 2002 silam.

Penulis ber agama islam, penganut ajaran ahlusunnah wal jamaah atau mudahnya warga nahdiyin

Hobinya seperti angin kadang suka sholawatan kadang suka musik-musikan.

Meski berstatus jomblo penulis sudah ada yang di perjuangin, yang dia lahir di kota yang sama pada 31 oktober 2001 dia di juluki huurun' ain

Penulis pernah menempuh pendidikan di TKPGRI Sumbertengah , SDN Sumber Tengah 1, SMP Ibrahimi 1 Sukorejo, SMK Ibrahimi 1 Sukorejo



Dsn. Pal 16 Ds. Sumber
Tengah, Kec. Binakal, Kab
Bondowoso



085258866858



Mochammadfebri127@gm
ail.com



Ig: all_fatihah