

HALAMAN MOTTO

~Mungkin kita sampai, mungkin saja tidak, tugas kita hanyalah **berjalan**~
(The Jeblogs)

~Kegagalan adalah bumbu kehidupan, Belajar bangkit
dan rakit kembali~
(MCPR)

~ Asa besar tercipta di dada. Merangkul hati, saling menjaga.
Percaya akan tiba cahaya~
(Rebellion Rose)



HALAMAN PERSEMBAHAN

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala limpahan rahmat, karunia, dan ridha-Nya sehingga penulis diberikan kesehatan, kekuatan, serta kemudahan dalam menyelesaikan skripsi ini. Halaman persembahan ini merupakan ungkapan rasa syukur dan terima kasih yang tulus kepada seluruh pihak yang telah memberikan doa, kasih sayang, dukungan, motivasi, serta pengorbanan selama proses penyusunan skripsi. Dengan penuh kerendahan hati dan mengucapkan Bismillahirrahmanirrahim, penulis mempersembahkan skripsi ini kepada:

1. Allah SWT, atas segala limpahan rahmat, hidayah, nikmat kesehatan, kekuatan, serta kemudahan yang senantiasa diberikan sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
2. Ibu tercinta, Sri Kandarsih, yang selalu menjadi sumber kasih sayang, doa, semangat, dan kekuatan dalam setiap langkah kehidupan penulis. Terima kasih atas segala pengorbanan, perhatian, serta dukungan yang tiada henti demi masa depan penulis. Semoga Allah SWT senantiasa melimpahkan rahmat, keberkahan, kesehatan, dan kebahagiaan kepada Ibu.
3. Bapak tercinta, Kholil, yang telah membimbing penulis dengan penuh kesabaran serta mengajarkan nilai-nilai kehidupan, kerja keras, tanggung jawab, dan kemandirian. Terima kasih atas setiap pengorbanan, dukungan, dan kepercayaan yang selalu diberikan sehingga penulis dapat terus melangkah untuk meraih cita-cita. Semoga Allah SWT senantiasa melimpahkan rahmat, kesehatan dan keberkahan.
4. Kakak tersayang, Nanik Sulistyaningrum, yang senantiasa memberikan doa, dukungan, motivasi, serta semangat kepada penulis selama proses penyusunan skripsi ini. Terima kasih atas segala bentuk dukungan yang telah diberikan sehingga penulis dapat menyelesaikan studi dengan baik. Semoga Allah SWT senantiasa melimpahkan rahmat, kesehatan, keberkahan, dan kebahagiaan kepada Kakak beserta keluarga.
5. Seluruh civitas akademika Universitas Putra Bangsa, yang telah memberikan ilmu, pengalaman, serta lingkungan akademik yang mendukung proses pembelajaran penulis hingga terselesaikannya skripsi ini.
6. Teman-teman seperjuangan Program Studi Sains Data angkatan 2022, yang telah menjadi bagian dari perjalanan akademik penulis sejak semester pertama hingga akhir perkuliahan. Terima kasih atas kebersamaan, dukungan, semangat, serta pengalaman berharga yang telah dilalui bersama dan akan selalu menjadi kenangan yang tidak terlupakan.

ABSTRAK

Kesehatan mental merupakan isu yang semakin mendapat perhatian di Indonesia. Meningkatnya penggunaan media sosial, khususnya YouTube, mendorong masyarakat menyampaikan pendapat dan pengalaman mengenai kesehatan mental melalui kolom komentar. Komentar tersebut dapat dimanfaatkan untuk mengetahui kecenderungan sentimen masyarakat menggunakan teknik analisis sentimen. Penelitian ini bertujuan menganalisis sentimen komentar YouTube mengenai isu kesehatan mental di Indonesia menggunakan algoritma Naïve Bayes serta menentukan skenario pembagian data yang menghasilkan kinerja terbaik.

Penelitian menggunakan 900 komentar dari tiga video YouTube bertema kesehatan mental yang diperoleh melalui proses *scraping*. Data diseleksi dan diberi label secara manual oleh tiga *annotator* menggunakan metode *majority voting*. Tahap *preprocessing* meliputi *cleaning*, *case folding*, *tokenization*, dan *stopword removal*. Selanjutnya dilakukan pembobotan TF-IDF dan proses klasifikasi menggunakan algoritma Naïve Bayes. Pengujian dilakukan menggunakan tiga skenario pembagian data, yaitu 70:30, 80:20, dan 90:10. Evaluasi model dilakukan menggunakan *Confusion matrix* dengan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

Hasil penelitian menunjukkan bahwa skenario pembagian data 90:10 memberikan kinerja terbaik dengan *accuracy* sebesar 82,22%, *precision* sebesar 79,55%, *recall* sebesar 83,33%, dan *F1-score* sebesar 81,36%. Sementara itu, skenario 70:30 menghasilkan *accuracy* sebesar 76,67%, *precision* sebesar 78,90%, *recall* sebesar 68,25%, dan *F1-score* sebesar 73,17%, sedangkan skenario 80:20 memperoleh *accuracy* sebesar 77,78%, *precision* sebesar 77,56%, *recall* sebesar 73,81%, dan *F1-score* sebesar 75,61%. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes dengan pembobotan TF-IDF efektif digunakan untuk analisis sentimen komentar YouTube mengenai isu kesehatan mental.

Kata kunci: Analisis Sentimen, Kesehatan Mental, YouTube, Naïve Bayes

ABSTRACT

Mental health is an increasingly prominent issue in Indonesia. The increasing use of social media, particularly YouTube, encourages people to share their opinions and experiences regarding mental health through the comments section. These comments can be used to determine public sentiment trends using sentiment analysis techniques. This study aims to analyze the sentiment of YouTube comments on mental health issues in Indonesia using the Naïve Bayes algorithm and determine the best data sharing scenario.

The study used 900 comments from three YouTube videos on the topic of mental health obtained through a scraping process. The data was manually selected and labeled by three annotators using a majority voting method. The preprocessing stage included cleaning, case folding, tokenization, and stopword removal. Next, TF-IDF weighting and classification processes were performed using the Naïve Bayes algorithm. Testing was conducted using three data sharing scenarios: 70:30, 80:20, and 90:10. Model evaluation was conducted using a Confusion matrix with accuracy, precision, recall, and F1-score metrics.

The research results indicate that the 90:10 data split scenario yielded the best performance, with an accuracy of 82.22%, precision of 79.55%, recall of 83.33%, and an F1-score of 81.36%. Meanwhile, the 70:30 scenario resulted in an accuracy of 76.67%, precision of 78.90%, recall of 68.25%, and an F1-score of 73.17%, whereas the 80:20 scenario achieved an accuracy of 77.78%, precision of 77.56%, recall of 73.81%, and an F1-score of 75.61%. The findings demonstrate that the Naïve Bayes algorithm with TF-IDF weighting is effective for the sentiment analysis of YouTube comments regarding mental health issues.

Keywords: Sentiment Analysis, Mental Health, YouTube, Naïve Bayes

KATA PENGANTAR

Alhamdulillahirabbil'alamin, segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala limpahan rahmat, nikmat, serta karunia-Nya sehingga penulis diberikan kesehatan, kekuatan, dan kemudahan dalam menyelesaikan skripsi yang berjudul **"Sentimen Analisis Komentar YouTube Mengenai Isu Kesehatan Mental di Indonesia Menggunakan Algoritma Naïve Bayes"**. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains Data pada Program Studi Sains Data, Fakultas Sains dan Teknologi, Universitas Putra Bangsa.

Proses penyusunan skripsi ini memberikan banyak pengalaman dan pembelajaran bagi penulis. Berbagai tantangan dan hambatan yang dihadapi selama penelitian dapat dilalui berkat izin Allah SWT serta bantuan, doa, bimbingan, dan dukungan dari berbagai pihak. Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan sehingga masih terdapat kekurangan dalam penyusunannya. Oleh karena itu, kritik dan saran yang bersifat membangun sangat penulis harapkan sebagai bahan perbaikan pada penelitian selanjutnya. Penulis berharap skripsi ini dapat memberikan manfaat bagi pembaca serta menjadi salah satu referensi dalam pengembangan ilmu pengetahuan, khususnya di bidang analisis sentimen dan data mining.

Pada kesempatan ini, penulis menyampaikan rasa hormat dan terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr. Gunarso Wiwoho, S.E., M.M., selaku Rektor Universitas Putra Bangsa.
2. Bapak Yulianto, S.Kom., M.Kom., selaku Dekan Fakultas Sains dan Teknologi Universitas Putra Bangsa.
3. Bapak Sarjimin, S.Kom., M.Kom., selaku Ketua Program Studi Sains Data Universitas Putra Bangsa.
4. Ibu Wahyuni Windasari, S.Si., M.Sc., selaku dosen pembimbing yang dengan penuh kesabaran telah memberikan arahan, masukan, motivasi, serta meluangkan waktu dalam membimbing penulis hingga skripsi ini dapat terselesaikan.

5. Seluruh dosen dan tenaga kependidikan Fakultas Sains dan Teknologi Universitas Putra Bangsa, khususnya Program Studi Sains Data, yang telah memberikan ilmu, pengalaman, serta pelayanan selama penulis menempuh pendidikan.
6. Ibu Sri Kandarsih dan Bapak Kholil, yang senantiasa memberikan doa, perhatian, dukungan, motivasi, serta semangat kepada penulis selama menempuh pendidikan hingga menyelesaikan skripsi ini. Terima kasih atas segala kebaikan, kepedulian, dan bantuan yang telah diberikan. Semoga Allah SWT senantiasa melimpahkan rahmat, kesehatan, keberkahan, serta membalas segala kebaikan Bapak dan Ibu dengan pahala yang berlipat ganda.
7. Kakak penulis, Nanik Sulistyaningrum, yang selalu memberikan doa, motivasi, perhatian, dan dukungan kepada penulis selama proses penyusunan skripsi.
8. Teman-teman Program Studi Sains Data Angkatan 2022 yang telah menjadi bagian dari perjalanan akademik penulis serta memberikan kebersamaan, semangat, dan pengalaman yang berharga selama masa perkuliahan.
9. Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada Heralda Dwiningtiyas dan Laili Ufi Muzayanah yang telah berkenan menjadi annotator dalam penelitian ini. Terima kasih atas kesediaan, waktu, ketelitian, serta kontribusi yang diberikan dalam proses pelabelan data sehingga penelitian ini dapat diselesaikan dengan baik.
10. Seluruh pihak yang tidak dapat penulis sebutkan satu per satu yang telah memberikan bantuan, doa, dan dukungan sehingga skripsi ini dapat diselesaikan.

Semoga Allah SWT membalas segala kebaikan yang telah diberikan kepada penulis dengan balasan yang berlipat ganda. Penulis berharap skripsi ini dapat memberikan manfaat bagi pembaca, menjadi referensi bagi penelitian selanjutnya, serta memberikan kontribusi bagi perkembangan ilmu pengetahuan, khususnya dalam bidang analisis sentimen.

Kebumen, 30 Juli 2026

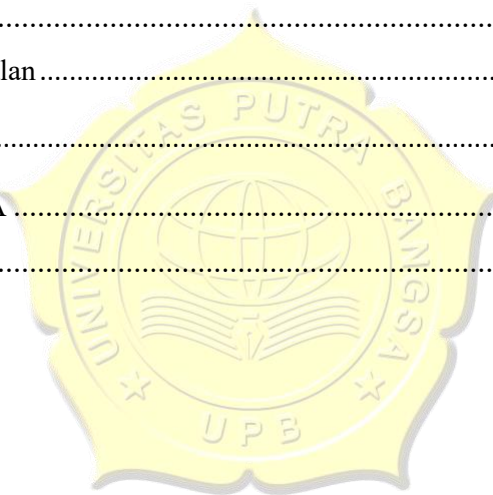
Andi Riawan

DAFTAR ISI

HALAMAN SAMPUL DEPAN.....	i
HALAMAN JUDUL.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN UJIAN.....	iv
HALAMAN BEBAS PLAGIARISME.....	v
HALAMAN MOTTO.....	vi
HALAMAN PERSEMBAHAN.....	vii
ABSTRAK.....	viii
ABSTRACT.....	ix
KATA PENGANTAR.....	x
DAFTAR ISI.....	xii
DAFTAR TABEL.....	xv
DAFTAR GAMBAR.....	xvi
DAFTAR LAMPIRAN.....	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	7
1.3 Batasan Masalah.....	7
1.4 Tujuan Penelitian.....	8
1.5 Manfaat Penelitian.....	8
BAB II KAJIAN PUSTAKA.....	10
2.1 Tinjauan Teori.....	10
2.1.1. Analisis Sentimen.....	10
2.1.2. <i>Machine learning</i>	11
2.1.3. Natural Language Processing.....	13
2.1.4. Naive Bayes.....	14

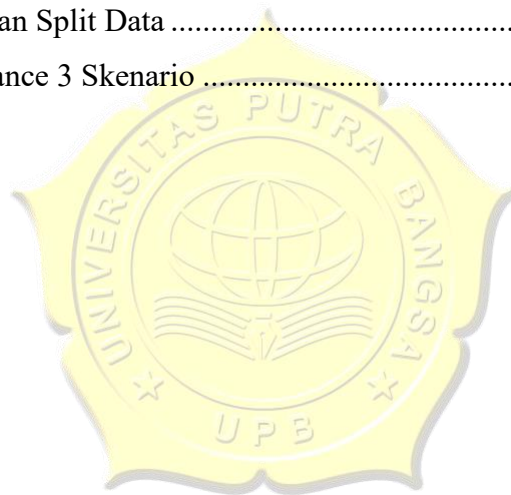
2.1.5.	Kesehatan Mental	17
2.1.6.	<i>Confusion matrix</i>	18
2.2	Tinjauan Penelitian Terdahulu.....	21
2.3	Kerangka Pemikiran.	24
2.4	Keaslian Penelitian	26
BAB III METODOLOGI PENELITIAN.....		28
3.1	Tahapan Penelitian.....	28
3.2	Objek dan Subjek Penelitian.....	30
3.3	Tools Pengujian	31
3.4	Metode Penelitian	31
3.4.1.	Pengambilan Data.....	31
3.4.2.	Labeling.....	35
3.4.3.	Tahapan Preprocessing.....	36
3.5	Tahapan Klasifikasi	41
2.5.1.	TF-IDF.....	41
2.5.2.	Skenario Uji Model	42
2.5.3.	Klasifikasi Naïve Bayes.....	43
2.5.4.	Evaluasi Model.....	43
BAB IV HASIL DAN PEMBAHASAN		45
4.1	Studi Literatur	45
4.2	Pengumpulan Data.....	45
4.3	Labeling	48
4.4	Preprocessing.....	50
4.4.1	Cleaning.....	50
4.4.2	Tokenize	51
4.4.3	Case Folding.....	52

4.4.4	Stopword.....	52
4.5	Analisis Deskriptif.....	53
4.5.1.	Distribusi Sentimen	53
4.5.2.	Wordcloud	55
4.6	TF-IDF	58
4.7	Skenario Pengujian Model.....	60
4.8	Klasifikasi Naive Bayes.....	62
4.9	Evaluasi Model	63
4.10	Pembahasan	66
BAB V	PENUTUP.....	71
5.1	Kesimpulan	71
5.2	Saran.....	72
DAFTAR	PUSTAKA.....	73
LAMPIRAN		78



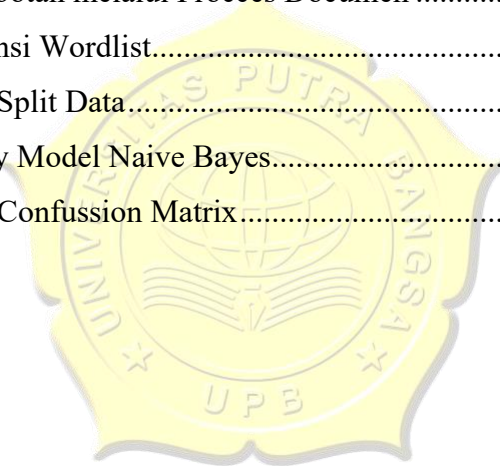
DAFTAR TABEL

Tabel II-1 Confussion Matrix.....	19
Tabel II-2 Penelitian terdahulu.....	22
Tabel III-1 Contoh Dataset.....	34
Tabel III-2 Tahap Preporcessing.....	39
Tabel IV-1 Data Komentar Youtube.....	48
Tabel IV-2 Labeling Annotator.....	49
Tabel IV-3 Hasil Cleaning	51
Tabel IV-4 Hasil Tokenize.....	52
Tabel IV-5 Hasil Case Folding	52
Tabel IV-6 Hasil Stopword	53
Tabel IV-7 Pembagian Split Data	61
Tabel IV-8 Performance 3 Skenario	65



DAFTAR GAMBAR

Gambar I-1 media sosial dengan rata-rata durasi sesi tertinggi	3
Gambar II-1 Kerangka Pemikiran	26
Gambar III-1 Tahapan Penelitian	28
Gambar IV-1 Scraping Video 1	46
Gambar IV-2 Scraping Video 2	46
Gambar IV-3 Scraping Video 3	47
Gambar IV-4 Proses Cleaning	50
Gambar IV-5 Distribusi Sentimen	54
Gambar IV-6 Worcloud komentar youtube	56
Gambar IV-7 Pembobotan melalui Procces Documen	58
Gambar IV-8 Frekuensi Wordlist.....	59
Gambar IV-9 Proses Split Data.....	61
Gambar IV- 10 Apply Model Naive Bayes.....	62
Gambar IV-11 Hasil Confussion Matrix.....	63



DAFTAR LAMPIRAN

Lampiran 1 Kartu Bimbingan	79
Lampiran 2 Kartu Seminar Proposal.....	80
Lampiran 3 Dataset dan Pelabelan 3 Annotator.....	81
Lampiran 4 Proses Pembersihan Data	148
Lampiran 5 Klasifikasi Naive Bayes.....	149

