

IMPLEMENTASI *BIDIRECTIONAL ENCODER REPRESENTATION FROM TRANSFORMERS* (BERT) UNTUK ANALISIS SENTIMEN KOMENTAR YOUTUBE CERAMAH ISLAM

SKRIPSI

Oleh :

SARAH ARELIA RAHMAH
NIM. 210605110049



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

**IMPLEMENTASI *BIDIRECTIONAL ENCODER REPRESENTATION*
FROM TRANSFORMERS (BERT) UNTUK ANALISIS SENTIMEN
KOMENTAR YOUTUBE CERAMAH ISLAM**

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
SARAH ARELIA RAHMAH
NIM. 210605110049

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025**

HALAMAN PERSETUJUAN

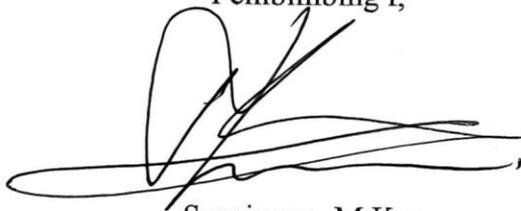
**IMPLEMENTASI *BIDIRECTIONAL ENCODER REPRESENTATION*
FROM TRANSFORMERS (BERT) UNTUK ANALISIS SENTIMEN
KOMENTAR YOUTUBE CERAMAH ISLAM**

SKRIPSI

Oleh :
SARAH ARELIA RAHMAH
NIM. 210605110049

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 2 Mei 2025

Pembimbing I,



Supriyono, M.Kom
NIP. 19841010 201903 1 012

Pembimbing II,



Dr. Zainal Abidin, M.Kom
NIP. 19760613 200501 1 004

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Fachrul Kurniawan, M.MT., IPU
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

**IMPLEMENTASI *BIDIRECTIONAL ENCODER REPRESENTATION*
FROM TRANSFORMERS (BERT) UNTUK ANALISIS SENTIMEN
KOMENTAR YOUTUBE CERAMAH ISLAM**

SKRIPSI

**Oleh :
SARAH ARELIA RAHMAH
NIM. 210605110049**

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 26 Mei 2025

Susunan Dewan Penguji

Ketua Penguji	: <u>Fatchurrochman, M.Kom</u> NIP. 19700731 200501 1 002	()
Anggota Penguji I	: <u>Nur Fitriyah Ayu Tunjung Sari, M.Cs</u> NIP. 19911226 202012 2 001	()
Anggota Penguji II	: <u>Supriyono, M.Kom</u> NIP. 19841010 201903 1 012	()
Anggota Penguji III	: <u>Dr. Zainal Abidin, M.Kom</u> NIP. 19760613 200501 1 004	()

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Fachrul Kurniawan, M.MT., IPU
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Sarah Arelia Rahmah
NIM : 210605110049
Fakultas / Program Studi : Sains dan Teknologi / Teknik Informatika
Judul Skripsi : Implementasi *Bidirectional Encoder Representation From Transformers* (BERT) Untuk Analisis Sentimen Komentar Youtube Ceramah Islam

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 2 Mei 2025

Yang membuat pernyataan,



Sarah Arelia Rahmah

NIM. 210605110049

MOTTO

"Tidak ada keberhasilan tanpa perjuangan, pengorbanan, dan air mata."

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur kepada Allah SWT atas segala rahmat dan kemudahan-Nya, sehingga skripsi ini dapat terselesaikan dengan baik. Karya ini penulis persembahkan kepada:

Ibunda tercinta, Eli dan Ayahanda tercinta, Arief,
Yang selalu mengalirkan kasih sayang, usaha terbaik, do'a-do'a tulus, dukungan baik materi maupun non materi, dan nasehat yang tiada henti.

Adik tercinta,
Yang menjadi salah satu motivasi dan dorongan untuk terus maju hingga skripsi ini terselesaikan.

Segenap keluarga besar,
Yang selalu mengiringi perjalanan penulis dengan do'a dan dukungan.

Aster Teknik Informatika angkatan 2021,
Yang telah berjuang bersama dari awal hingga akhir.

teruntuk diri sendiri,
Terima kasih sudah berusaha dan bertahan sejauh ini.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Alhamdulillah *rabbilalamin*, segala puji dan syukur senantiasa penulis panjatkan ke hadirat Allah subhanahu wa ta'ala atas berkat Rahmat, serta hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Implementasi *Bidirectional Encoder Representation From Transformers* (BERT) Untuk Analisis Sentimen Komentar Youtube Ceramah Islam” dengan baik dan lancar. Shalawat serta salam tetap turunkan kepada Nabi Muhammad SAW, yang telah membawa kita dari zaman kebodohan menuju zaman kebenaran yakni Islam dan zaman yang penuh dengan ilmu pengetahuan sebagaimana yang di rasakan pada saat ini. Dan semoga kita semua mendapat syafaatnya di hari akhir kelak, Aamiin.

Penulis mengucapkan rasa terima kasih yang begitu besar kepada seluruh pihak yang memberikan dukungan dan motivasi kepada penulis sehingga dapat menyelesaikan skripsi ini. Ucapan terima kasih ini penulis disampaikan kepada:

1. Prof. Dr. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Hj. Sri Harini, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Ir. Fachrul Kurniawan, M.MT., IPU., selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.

4. Supriyono, M.Kom., selaku dosen pembimbing I yang telah dengan sabar memberikan bimbingan, arahan, dan dukungan yang sangat berarti selama proses penyusunan skripsi ini
5. Dr. Zainal Abidin, M.Kom., selaku dosen pembimbing II penulis yang senantiasa memberikan bimbingan, arahan dan masukan untuk penulis dalam mengerjakan dan menyelesaikan skripsi ini.
6. Fathurrochman, M.Kom selaku penguji I dan Nur Fitriyah Ayu Tanjung Sari, M.Cs., selaku penguji II yang telah berkenan menguji, memberikan masukan, kritik, saran, serta memberikan masukan yang membangun sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
7. Syahiduz Zaman, M.Kom., selaku dosen wali yang senantiasa membantu serta turut memberikan ilmu dan saran untuk penulis selama menjalani masa studi pada Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
8. Nia Faricha S, Si selaku admin Program Studi Teknik Informatika yang selalu sabar memberikan informasi, membantu, dan memberikan arahan selama perkuliahan dan proses penulisan skripsi ini.
9. Segenap dosen, laboran, dan jajaran staff Program Studi Teknik Informatika yang telah memberikan ilmu, pengetahuan, dan dukungan selama penulis menjalani studi.
10. Kedua orang tua penulis, Ayahanda Arief dan Ibunda Elli, yang selalu memberikan dukungan, kasih sayang, dan semangat tanpa henti. Bimbingan dan pengorbanan dari Ayahanda dan Ibunda menjadi salah satu alasan

utama yang menguatkan penulis untuk melanjutkan perjalanan ini hingga hari ini.

11. Adik Penulis Balqish Arelia yang telah menjadi salah satu sumber motivasi dan semangat. Kehadiran adik membuat penulis lebih dewasa dan terus termotivasi untuk menyelesaikan skripsi ini.
12. Keluarga Besar Penulis Nenek, Kakek, Om, Tante, Pakdhe, Budhe, dan Bulek, yang selalu memberikan dukungan dan Doa yang membuat penulis bisa menyelesaikan skripsi ini.
13. Orang teristimewa, Muhammad Raidditya Arrafi atas waktu, perhatian, dan kesediaanmu untuk selalu ada selama proses perkuliahan.
14. Sahabat sekamar Salma Chesha Putri Pramudya Wardani yang telah menjadi teman berbagi cerita, lelah, dan tawa sepanjang perjalanan ini. Terima kasih atas kebersamaan yang menyenangkan.
15. Sahabat Penulis Mutiara, Aisya, Fitri, Frisca, Hajratun, Nurjihan, Najah, Rofidatus, Nurdiva yang telah kebersamaian penulis memberi semangat dan warna selama perkuliahan.
16. Teman-teman Penulis KKM 163 dan IKAPMI Malang atas dukungan dan membuat cerita selama masa perkuliahan.
17. Seluruh warga Teknik Informatika khususnya angkatan 2021 “Aster” yang telah memberikan kehangatan, motivasi, semangat, dan dukungan kepada penulis.
18. Seluruh pihak yang telah terlibat, baik secara langsung maupun tidak langsung dari awal perkuliahan hingga akhir penulisan skripsi ini.

19. Diri Sendiri yang sampai hari ini masih bertahan walaupun pasti ada saat-saat sulit yang membuat ingin menyerah tapi kamu berhasil melewati kesulitan itu, teruslah maju gapai semua cita-citamu ini bukan akhir perjalananmu. Hiduplah lebih lama lagi ya.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih jauh dari kata sempurna. Maka dari itu penulis menerima saran, kritik dan masukan yang bersifat membangun dari berbagai pihak untuk meningkatkan kualitas dan pengembangan lebih lanjut dari penelitian ini. Penulis percaya bahwa proses pembelajaran tidak pernah berhenti, dan setiap masukan yang diberikan akan menjadi langkah berharga dalam perjalanan akademik dan pengembangan ilmu pengetahuan. Penulis berharap semoga skripsi ini dapat memberikan manfaat nyata bagi pembaca, peneliti selanjutnya, maupun pihak-pihak terkait. Dan semoga penelitian ini berkontribusi bagi pengembangan ilmu pengetahuan dan dapat bermanfaat untuk di masa yang akan datang.

Wassalamu 'alaikum warahmatullahi wabarakatuh.

Malang, 26 Mei 2025

Penulis

DAFTAR ISI

HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	xii
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xv
ABSTRAK	xvi
ABSTRACT	xvii
المخلص	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
BAB II STUDI PUSTAKA	6
2.1 Penelitian Terkait	6
2.2 Analisis Sentimen	7
2.3 <i>Bidirectional Encoder Representation from Transformers</i>	9
2.4 Ceramah	10
BAB III DESAIN DAN IMPLEMENTASI	10
3.1 Desain Penelitian	11
3.2 Implementasi.....	13
3.2.1 Pengumpulan Data.....	14
3.2.2 <i>Labeling Data</i>	14
3.2.3 <i>Preprocessing Data</i>	15
3.2.4 Menetapkan Nilai	16
3.2.5 Membersihkan Data.....	17
3.2.6 Konversi Label	18
3.2.7 <i>Bert Tokenizer</i>	19
3.2.8 Pembagian Data.....	24
3.2.9 Menyiapkan Data dan Model	25
3.2.10 Pengaturan Pelatihan	28
3.2.11 Pelatihan Model.....	29
3.2.12 Pengujian Model.....	30
3.2.13 Evaluasi Model.....	32
BAB IV HASIL DAN PEMBAHASAN	35
4.1 Pengumpulan Data	35
4.2 <i>Prepossessing Data</i>	37

4.3	Pembagian Data	42
4.4	Pelatihan model BERT	44
4.5	Evaluasi Model	50
4.6	Pembahasan.....	51
4.7	Integrasi Sains dan Islam	53
	BAB V KESIMPULAN DAN SARAN	58
5.1	Kesimpulan	58
5.2	Saran	59
	DAFTAR PUSTAKA	
	LAMPIRAN	

DAFTAR GAMBAR

Gambar 2.1 Arsitektur BERT	9
Gambar 3.1 Desain Penelitian.....	11
Gambar 3.2 Tokenisasi.....	20
Gambar 3.3 Penambahan Token [CLS] dan [SEP].....	21
Gambar 3.4 Mengubah Token ID	22
Gambar 3.5 Penambahan <i>Token Padding</i>	22
Gambar 3.6 <i>Sentence Embedding</i>	23
Gambar 3.7 <i>Positional Embedding</i>	24
Gambar 4.1 Grafik Komentar	37
Gambar 4.2 Training <i>Loss Split 90</i>	45
Gambar 4.3 Grafik Akurasi <i>Split 90</i>	46
Gambar 4.4 Grafik <i>Training Loss Split 70</i>	47
Gambar 4.5 Grafik Akurasi Split 70	48
Gambar 4.6 Grafik <i>Training Loss Split 50</i>	49
Gambar 4.7 Grafik Akurasi <i>Split 50</i>	50

DAFTAR TABEL

Tabel 2.1 Penelitian Terkait	6
Tabel 3.1 <i>Preprocessing</i> Data	15
Tabel 3.2 <i>Case Folding</i>	15
Tabel 3.3 Normalisasi	15
Tabel 3.4 Tokenisasi	16
Tabel 3.5 <i>Stopword removal</i>	16
Tabel 3.6 Evaluasi Model	34
Tabel 4.1 Data Berlabel.....	36
Tabel 4.2 Hasil <i>Case Folding</i>	39
Tabel 4.3 Hasil Normalisasi.....	40
Tabel 4.4 Hasil Tokenisasi.....	41
Tabel 4.5 Hasil <i>Stopword Removal</i>	42
Tabel 4.6 Pembagian Data	43
Tabel 4.7 Hasil Evaluasi	51

ABSTRAK

Rahmah, Sarah Arelia. 2025. **Implementasi *Bidirectional Encoder Representation From Transformers (BERT)* untuk Analisis Sentimen Komentar Youtube Ceramah Islam**. Skripsi. Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Supriyono, M.Kom (II) Dr. Zainal Abidin, M.Kom.

Kata Kunci: Analisis Sentimen, BERT, Ceramah Islam, Komentar, Pemrosesan Bahasa Alami

Penelitian ini bertujuan untuk menerapkan model *Bidirectional Encoder Representations from Transformers (BERT)* untuk analisis sentimen komentar pada video ceramah Islam di YouTube. Analisis sentimen ini bertujuan untuk mengidentifikasi sentimen positif, negatif, dan netral dari komentar-komentar yang terdapat pada video ceramah. Hasil analisis ini dapat memberikan wawasan terkait penerimaan *audiens* terhadap materi yang disampaikan. Proses penelitian dimulai dengan pengumpulan data komentar menggunakan *YouTube Data API*, yang kemudian diproses melalui tahapan *preprocessing*, seperti *case folding*, normalisasi, tokenisasi, dan *stopword removal*. Setelah itu, model BERT diterapkan untuk mengklasifikasikan komentar berdasarkan sentimennya. Evaluasi dilakukan dengan menggunakan *confusion matrix* yang terdiri dari akurasi, presisi, *recall*, dan *f1-score*. Hasil penelitian menunjukkan bahwa model BERT mencapai akurasi sebesar 83% dalam mengklasifikasikan sentimen komentar, dengan kinerja terbaik pada kategori sentimen negatif. Meskipun model menunjukkan hasil yang baik, terdapat beberapa kesalahan dalam mengklasifikasikan sentimen positif dan netral. Penelitian ini juga memberikan kontribusi dalam penerapan teknologi analisis sentimen dalam konteks ceramah Islam dan dapat digunakan untuk meningkatkan kualitas dakwah di era digital.

ABSTRACT

Rahmah, Sarah Arelia. 2025. **Implementation of Bidirectional Encoder Representation From Transformers (BERT) to Sentiment Analysis of Youtube Comments on Islamic Lectures**. Thesis. Department of Informatics Engineering, Faculty of Science and Technology, State Islamic University Maulana Malik Ibrahim Malang. Advisor: (I) Supriyono, M.Kom (II) Dr. Zainal Abidin, M.Kom.

This study aims to apply the Bidirectional Encoder Representations from Transformers (BERT) model in analysing the sentiment of comments on Islamic lecture videos on YouTube. This sentiment analysis aims to identify positive, negative, and neutral sentiment from the comments on the lecture videos. The results of this analysis can provide insights into the audience's acceptance of the material presented. The research process began with the collection of comment data using the YouTube Data API, which was then processed through preprocessing stages, such as *case folding*, normalisation, tokenisation, and *stopword removal*. After that, the BERT model was applied to classify comments based on their sentiment. The evaluation was carried out using a confusion matrix and other evaluation metrics, such as Accuracy, precision, recall, and F1-Score. The results of the study show that the BERT model achieved 83% Accuracy in classifying comment sentiment, with the best performance in the negative sentiment category. Although the model showed good results, there were some errors in classifying positive and neutral sentiment. This study also contributes to the application of sentiment analysis technology in the context of Islamic lectures and can be used to improve the quality of da'wah in the digital age.

Key words: BERT, Comment, Islamic Lectures, Natural Language Processing, Sentiment Analysis

الملخص

رحمة، سارة أريليا. 2025. تنفيذ تمثيل التشفير ثنائي الاتجاه من المحولات) بيرت (في تحليل المشاعر لتعليقات اليوتيوب على المحاضرات الإسلامية. الأطروحة. قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم مالانج
د. زين العابدين زين العابدين، م. كوم (II) سوبريونو، م. كوم (I): الإسلامية الحكومية. المشرف

الكلمات المفتاحية: محاضرات إسلامية، تعليقات، معالجة اللغة الطبيعية، BERT

هدف هذا البحث إلى تطبيق نموذج تمثيل الترميز ثنائي الاتجاه من المحولات (BERT) لتحليل المشاعر في التعليقات على مقاطع الفيديو لخطاب الإسلامية على يوتيوب. يهدف تحليل المشاعر هذا إلى تحديد المشاعر الإيجابية والسلبية والمحايدة من التعليقات الموجودة على مقاطع الفيديو. يمكن أن توفر نتائج هذا التحليل رؤى تتعلق بقبول الجمهور للمواد المقدمة. تبدأ عملية البحث بجمع بيانات التعليقات باستخدام واجهة برمجة تطبيقات بيانات يوتيوب، والتي يتم معالجتها بعد ذلك من خلال مراحل المعالجة الأولية، مثل تقليل الحالة، والتطبيع، والتجزئة، وإزالة الكلمات الشائعة. بعد ذلك، تم تطبيق نموذج بيرت لتصنيف التعليقات بناءً على مشاعرها. تم إجراء التقييم باستخدام المصفوفة الالتباسية التي تتكون من الدقة، والموثوقية، والاسترجاع، ودرجة ف1. أشارت نتائج البحث إلى أن نموذج بيرت حقق دقة تبلغ 83٪ في تصنيف مشاعر التعليقات، مع أفضل أداء في فئة المشاعر السلبية. على الرغم من أن النموذج أظهر نتائج جيدة، إلا أن هناك بعض الأخطاء في تصنيف المشاعر الإيجابية والمحايدة. يساهم هذا البحث أيضًا في تطبيق تكنولوجيا تحليل المشاعر في سياق الخطب الإسلامية ويمكن استخدامه لتحسين جودة الدعوة في العصر الرقمي.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan komunikasi yang pesat telah mengubah cara manusia berinteraksi dan berbagi informasi. Salah satu *platform* yang sangat populer dalam berbagi informasi adalah Youtube. Youtube, sebagai salah satu situs berbagi video terbesar di dunia, tidak hanya menyediakan berbagai jenis hiburan, tetapi juga digunakan sebagai media untuk menyebarkan pengetahuan, diskusi, dan ceramah. Setiap harinya. Jutaan video diunggah dan dikomentari oleh pengguna dari berbagai latar belakang. Hal ini menciptakan volume besar pada data teks dalam bentuk komentar yang mengandung berbagai opini, pandangan dan sentimen (Kumari & Arockiaraj, 2023).

Komentar-komentar di Youtube dapat mencerminkan opini pengguna terhadap konten yang disajikan, termasuk pada video ceramah (Rani, 2023). Video ceramah khususnya yang menyangkut topik keagamaan atau sosial, seringkali mendapat respon yang beragam dari audiens (Lestari, 2024). Komentar tersebut bisa bersifat positif, netral atau negatif. Analisis sentimen dari komentar-komentar tersebut dapat memberikan wawasan mendalam tentang bagaimana opini Masyarakat terhadap suatu ceramah yang diterima (Suharsono & Nurahman, 2024).

Dalam Al-Qur'an, pentingnya pengelompokan manusia dapat dilihat dalam ayat berikut :

يَا أَيُّهَا النَّاسُ إِنَّا خَلَقْنَاكُمْ مِنْ ذَكَرٍ وَأُنْثَىٰ وَجَعَلْنَاكُمْ شُعُوبًا وَقَبَائِلَ لِتَعَارَفُوا ۗ إِنَّ أَكْرَمَكُمْ عِنْدَ اللَّهِ
 أَتَقْوَىٰ ۗ إِنَّ اللَّهَ عَلِيمٌ خَبِيرٌ

“Wahai manusia! Sungguh, Kami telah menciptakan kamu dari seorang laki-laki dan seorang perempuan, kemudian Kami jadikan kamu berbangsa-bangsa dan bersuku-suku agar kamu saling mengenal. Sesungguhnya yang paling mulia di antara kamu di sisi Allah Swt ialah orang yang paling bertakwa. Sungguh, Allah Swt. Maha Mengetahui, Mahateliti.” (Q.S. Al-Hujurat:13).

Pada ayat ini menekankan bahwa pengelompokan manusia bertujuan untuk saling mengenal dan memahami. Dalam konteks ini, analisis sentimen pada komentar youtube dapat mengidentifikasi pandangan beragam dari masyarakat terhadap ceramah yang disampaikan, sehingga menjadi sarana untuk saling memahami lebih baik.

Namun, analisis sentimen dari komentar-komentar ini tidak selalu mudah. Komentar yang ditulis cenderung tidak terstruktur, mengandung bahasa informal, slang atau bahasa daerah yang kurang dipahami. Tantangan ini menuntut metode yang lebih canggih dalam memproses bahasa alami agar bisa memahami konteks dan makna dari setiap komentar dengan baik. Dalam beberapa tahun terakhir, model BERT telah menjadi teknologi canggih karena kemampuannya untuk menangkap konteks dua arah dalam kalimat, sehingga memungkinkan pemahaman yang lebih mendalam terhadap teks.

Penggunaan metode BERT untuk menganalisis sentimen memberikan peluang besar dalam meningkatkan akurasi sentimen pada teks yang kompleks seperti komentar Youtube. Model ini diharapkan dapat menjadi solusi yang efektif dalam menganalisis sentimen dari komentar pada video ceramah. Akan tetapi,

penerapan model ini memerlukan pendekatan yang sistematis dan pengujian yang mendalam untuk memastikan hasil yang optimal.

Untuk memastikan bahwa implementasi model BERT sesuai dengan yang diharapkan, diperlukan pendekatan pengujian yang efektif. Salah satu metode yang dapat digunakan adalah *confusion matrix*. *Confusion matrix* adalah alat evaluasi yang memungkinkan pengukuran kinerja model dalam mengklasifikasikan data, dengan memberikan gambaran yang jelas mengenai prediksi dan kenyataan dalam klasifikasi sentimen.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk menerapkan BERT dalam analisis sentimen komentar pada video ceramah serta menguji kinerja model tersebut menggunakan *confusion matrix*. Dengan pendekatan ini diharapkan hasil yang diperoleh akan lebih akurat dan dapat diimplementasikan dalam skala yang lebih luas di masa mendatang.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, maka penelitian ini maka penelitian ini difokuskan untuk menjawab pertanyaan-pertanyaan berikut.

- a. Bagaimana implementasi BERT dapat digunakan untuk analisis sentimen komentar pada video ceramah islam di Youtube?
- b. Bagaimana menguji kinerja model dalam mengklasifikasikan komentar menggunakan *confusion matrix*?

1.3 Batasan Masalah

Penelitian ini memiliki batasan-batasan yang perlu diperjelas untuk memastikan fokus dan ruang lingkup yang jelas dalam analisis yang dilakukan, antara lain:

- a. Penelitian ini hanya menggunakan data berupa komentar pada video ceramah islam di Youtube menggunakan bahasa Indonesia.
- b. Komentar yang dianalisis dalam penelitian ini hanya terdiri dari komentar yang memiliki sentimen positif dan negatif.

1.4 Tujuan Penelitian

Berdasarkan permasalahan yang diidentifikasi pada latar belakang, penelitian ini bertujuan untuk :

- a. Menerapkan model BERT untuk analisis sentimen komentar YouTube pada video ceramah Islam dan mengklasifikasikan komentar menjadi positif, netral, atau negatif.
- b. Menggunakan *confusion matrix* untuk menguji kinerja model dalam mengklasifikasikan komentar.

1.5 Manfaat Penelitian

Diharapkan hasil dari penelitian ini mampu memberikan manfaat baik dari segi teoritis maupun praktis, sebagai berikut :

- a. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan analisis sentimen dan pemrosesan bahasa alami,

khususnya dalam penerapan model BERT pada komentar YouTube ceramah Islam berbahasa Indonesia.

- b. Penelitian ini diharapkan dapat digunakan oleh pengembang aplikasi atau Perusahaan, atau individu untuk memahami opini publik secara lebih terstruktur melalui analisis sentimen yang efektif, khususnya dalam konteks ceramah Islam di YouTube.

BAB II

STUDI PUSTAKA

2.1 Penelitian Terkait

Pada penelitian ini, analisis sentimen menjadi salah satu metode yang diimplementasikan untuk memahami opini publik terhadap ceramah islam di platform Youtube. Dalam studi Pustaka ini, beberapa penelitian terkait telah dikaji untuk memberikan Gambaran mengenai perkembangan teknologi dalam bidang analisis sentimen dan penerapan metode pemrosesan bahasa alami, seperti BERT. Penelitian-penelitian ini tidak hanya mencakup topik analisis sentimen, tetapi juga pemanfaatan media digital dalam penyebaran dakwah islam. Tabel 2.1 merangkum 5 penelitian terkait yang relevan dengan topik ini, dengan fokus pada metodologi, hasil utama, kebaruan serta kelemahan dari penelitian-penelitian tersebut.

Tabel 2.1 Penelitian Terkait

No	Peneliti	Judul	Metode	Hasil
1	(Areshey & Mathkour, 2023)	<i>Transfer Learning for Sentimen Classification Using Bidirectional Encoder Representations from Transformers (BERT) Model</i>	BERT, KNN, Naves Bayes, SVM	BERT dengan <i>sequence length</i> 384 dan <i>batch size</i> 32 memiliki akurasi tertinggi 97.3%
2	(Putra & Gata, 2020)	Analisis Sentimen Komentar Youtube Terhadap Ceramah Ning Umi Laila Sindir Rhoma Irama Menggunakan Algoritma LSTM	LSTM	Akurasi analisis sentimen sebesar 96%
3	(Harpizon et al., 2022)	Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes	Naïve Bayes	Akurasi sebesar 87%, presisi 91%, dan <i>recall</i> 97%
4	(Mualfah et al., 2023)	Analisis Sentimen Komentar YouTube TvOne Tentang Ustadz Abdul Somad Dideportasi	<i>Support Vector Machine (SVM)</i>	Akurasi 95.02%, presisi 95.18%, <i>recall</i> 95.02%, dan <i>F1-Score</i> 95.01%

No	Peneliti	Judul	Metode	Hasil
		Dari Singapura Menggunakan Algoritma SVM		menggunakan 29,122 komentar dari YouTube
5	(Vidya Chandradev et al., 2023)	Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT	<i>SmallBERT (Fine-Tuned BERT)</i>	Akurasi 91.40%, presisi 90.51%, <i>recall</i> 90.51%, dan <i>F1-Score</i> 90.51% menggunakan dataset 515k review hotel

2.2 Analisis Sentimen

Analisis sentimen adalah sebuah teknik dalam *Natural Language Processing* (NLP) yang digunakan untuk memahami dan mengklasifikasikan emosi, opini, atau sikap dalam teks yang tidak terstruktur. Teks tersebut bisa berasal dari berbagai sumber, seperti media sosial, ulasan produk, artikel berita, atau komentar online. Tujuan dari analisis sentimen adalah untuk mengidentifikasi apakah teks tersebut mengandung sentimen positif, netral, atau negatif (Singgalen, 2022). Teknik ini sering diterapkan menggunakan algoritma pembelajaran mesin yang mampu mengenali pola bahasa tertentu untuk menentukan sentimen yang terkandung. Dengan analisis setimen, kita dapat memahami bagaimana produk, layanan, atau topik tertentu diterima oleh publik, sehingga memungkinkan mereka untuk membuat keputusan berdasarkan kata yang dihasilkan dari opini Masyarakat (Arham et al., 2022).

Penelitian yang dilakukan oleh Giovani menjadi contoh yang relevan tentang penerapan analisis sentimen dalam dunia nyata. Dalam studinya, mereka menggunakan teknik ini untuk menilai Tingkat kepuasan pengguna aplikasi Ruangguru dengan menganalisis komentar pada ulasan yang diposting di media sosial. Analisis ini menghasikan data yang memberikan gambaran umum tentang

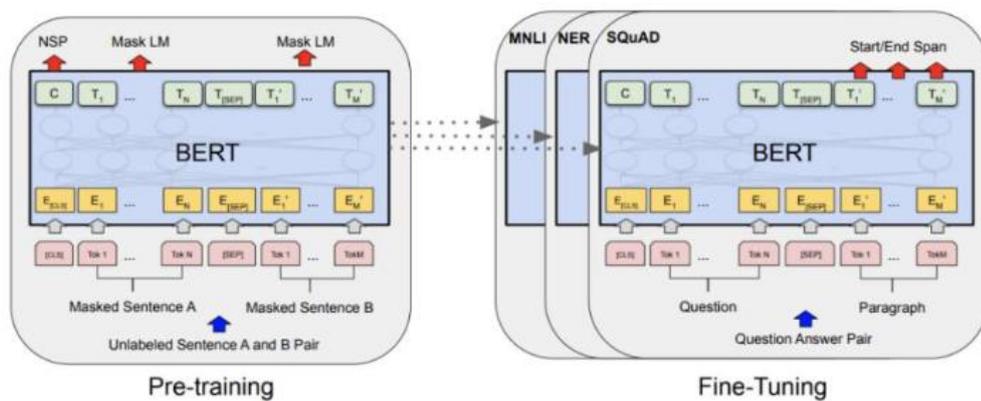
bagaimana aplikasi tersebut diterima oleh penggunanya, mengidentifikasi aspek-aspek yang disukai dan yang memerlukan perbaikan. Ini menunjukkan bahwa analisis sentimen dapat menjadi alat evaluasi yang sangat baik untuk menentukan strategi suatu produk (Giovani et al., 2020).

Safira dan Hasan membahas bagaimana teknik ini dapat secara otomatis mengklasifikasikan teks seperti tweet atau postingan di media sosial ke dalam kategori sentimen. Mereka menekankan bahwa penggunaan teknologi ini sangat efisien dalam memproses volume data yang besar, memungkinkan perusahaan untuk dengan cepat memahami sentimen publik tanpa memerlukan analisis manual yang memakan waktu. Hal ini sangat relevan dalam lingkungan digital yang terus berkembang, di mana opini publik dapat berubah dengan cepat dan membutuhkan respons yang cepat pula (Alfandi Safira & Hasan, 2023).

Di sisi lain analisis sentimen dapat diterapkan dalam berbagai bidang di luar bisnis dan teknologi. Dalam bidang ekonomi, misalnya, analisis sentimen dapat digunakan untuk memprediksi pergerakan pasar berdasarkan sentimen publik terhadap isu-isu tertentu. Dalam politik, teknik ini bisa digunakan untuk mengukur opini publik terhadap kandidat atau kebijakan tertentu. Media sosial, seperti Twitter, sering menjadi sumber data utama karena kemampuannya untuk mencerminkan opini publik secara luas dan *real-time*. Analisis sentimen terhadap data ini dapat memberikan pandangan mendalam tentang tren dan pola pikir masyarakat.

2.3 Bidirectional Encoder Representation from Transformers

Bidirectional encoder representations from transformer (BERT) adalah model pembelajaran mesin yang berbasis arsitektur transformer, yang dirancang untuk memahami konteks kalimat secara lebih mendalam (Situmorang & Purba, 2024). BERT dikembangkan oleh *google* dan menjadi teknologi baru dalam perosesan bahasa alami karena mampus memahami hubungan antar kata dalam dua arah (Asiva Noor Rachmayani, 2015).



Gambar 2.1 Arsitektur BERT. Sumber: (Putria & Nofab, 2025)

Berbeda dengan model NLP tradisional yang hanya memproses kata secara sekuensial dari kiri ke kanan atau sebaliknya, BERT memproses kata dalam konteks kalimat secara simultan dari kedua arah. Ini memungkinkan BERT untuk menangkap makna yang lebih akurat dari suatu kata berdasarkan konteks sekitarnya. Misalnya, dalam kalimat "bank di pinggir sungai" dan "bank tempat menyimpan uang", BERT mampu membedakan arti kata "bank" dengan memahami konteks seluruh kalimat.

Berdasarkan Gambar 2.1 arsitektur BERT terdiri *pre-Training* dan *fine-tuning*. *Pre-Training* dilakukan dengan melatih model pada kumpulan data besar

untuk mempelajari pola umum dalam bahasa, kemudian pada tahap *fine-tuning* (Elankath & Ramamirtham, 2023). BERT disesuaikan dengan tugas spesifik seperti klasifikasi sentimen, penjawaban pertanyaan, atau ekstraksi informasi (Devlin et al., 2019). Kelebihan utama BERT adalah kemampuannya untuk menangani tugas-tugas yang membutuhkan pemahaman kontekstual tingkat lanjut, menjadikannya sangat efektif dalam berbagai aplikasi NLP, termasuk analisis sentimen, dan penerjemahan bahasa (Jayadianti et al., 2022).

2.4 Ceramah

Ceramah adalah bentuk komunikasi lisan yang biasanya disampaikan oleh seseorang kepada sekelompok orang dengan tujuan memberikan nasihat, informasi, atau pemahaman tentang suatu topik tertentu, terutama yang berkaitan dengan moral, agama, atau pendidikan. Dalam konteks agama, ceramah sering kali disampaikan oleh pemuka agama atau ulama untuk membimbing pendengar dalam hal-hal yang berhubungan dengan keimanan, etika, dan kehidupan spiritual (Iriyanto et al., 2023).

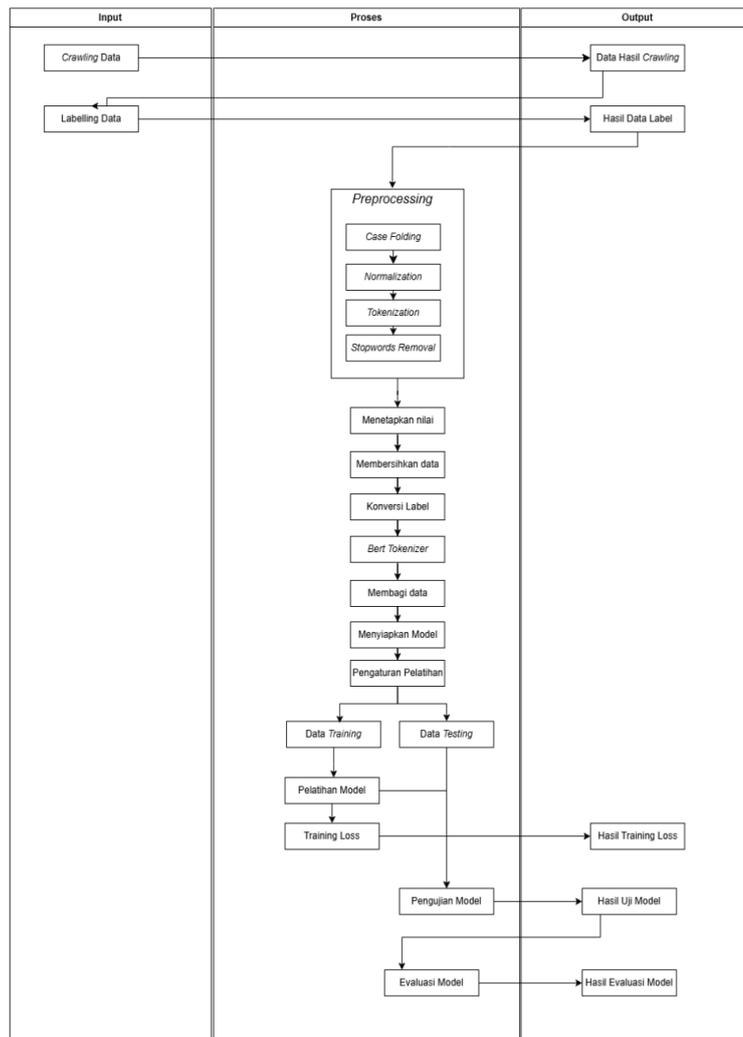
Ceramah umumnya bersifat satu arah, di mana pembicara memberikan pesan kepada audiens yang mendengarkan, meskipun terkadang ada sesi tanya jawab atau diskusi setelahnya. Topik ceramah dapat beragam, mulai dari ajaran agama, isu-isu sosial, nilai-nilai moral, hingga motivasi hidup (Lestari, 2024). Ceramah sering disampaikan dalam acara keagamaan, seperti pengajian, khutbah, atau seminar, dan menjadi media penting untuk menyebarkan pengetahuan dan inspirasi kepada masyarakat.

BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Desain sistem yang digunakan dalam penelitian ini bertujuan untuk menganalisis sentimen komentar Youtube menggunakan metode BERT. Tahapan dalam desain sistem dirancang agar alur data mulai dari pengambilan komentar hingga analisis sentimen dapat berjalan secara efisien.



Gambar 3.1 Desain Penelitian

Berdasarkan Gambar 3.1 alur sistem dimulai dengan pengumpulan data komentar dari video YouTube yang telah dipilih berdasarkan popularitas atau relevansinya terhadap penelitian. Komentar-komentar ini diambil melalui *API* YouTube yang memungkinkan sistem untuk secara otomatis mengunduh data komentar dalam jumlah besar, termasuk teks komentar, nama pengguna, waktu unggahan, dan jumlah suka. Setelah semua data terkumpul, langkah selanjutnya adalah labeling data, di mana setiap komentar diberikan label sentimen, seperti positif, negatif, guna mempersiapkannya untuk tahap analisis lebih lanjut.

Tahap berikutnya adalah *preprocessing* data, yang mencakup beberapa sub-proses yaitu *case folding*, yaitu mengonversi teks komentar menjadi huruf kecil. Hal ini dilakukan untuk memastikan konsistensi dalam analisis, guna menghindari perbedaan yang disebabkan oleh variasi penggunaan huruf besar dan kecil yang tidak relevan. Kemudian, dilakukan tahapan normalisasi, yaitu mengganti kata-kata singkatan atau tidak baku dengan bentuk baku, seperti mengganti kata "gk" menjadi "gak" atau "utk" menjadi "untuk". Setelah itu, dilakukan tahapan tokenization, di mana teks dipisahkan menjadi token-token atau kata-kata terpisah untuk mempermudah analisis lebih lanjut. Tahap terakhir adalah *stopwords removal*, yaitu menghapus kata-kata umum yang tidak memberikan kontribusi berarti terhadap analisis sentimen, seperti kata sambung atau kata yang terlalu sering muncul namun tidak signifikan dalam konteks analisis.

Setelah tahap *preprocessing*, data akan melewati beberapa langkah lagi sebelum digunakan untuk pelatihan model. Menetapkan nilai dilakukan untuk

menginisialisasi parameter yang diperlukan, diikuti dengan Membersihkan Data, yang memastikan data bebas dari nilai-nilai yang hilang atau tidak relevan.

Selanjutnya, konversi label digunakan untuk mengubah label kategori menjadi angka agar bisa diproses oleh model, sementara *BERT Tokenizer* mempersiapkan data teks untuk dimasukkan ke dalam model BERT. Setelah data siap, proses berikutnya adalah Membagi Data menjadi dua bagian utama data *Training* untuk pelatihan model dan data *testing* untuk evaluasi model. Data ini akan digunakan untuk Pelatihan Model, di mana model BERT dilatih menggunakan data pelatihan, dan selama pelatihan, *Training loss* dihitung untuk memantau seberapa baik model belajar dari data.

Setelah pelatihan selesai, dilakukan pengujian model pada data *testing* untuk mengevaluasi kinerja model yang sudah dilatih. Hasil dari pengujian ini mencakup hasil uji model yang menunjukkan seberapa akurat model dalam memprediksi data yang belum pernah dilihat sebelumnya. Proses terakhir adalah evaluasi model, yang menghasilkan hasil evaluasi model, seperti akurasi, presisi, dan metrik lainnya, yang digunakan untuk menilai keberhasilan model dalam melakukan tugas analisis sentimen atau tugas serupa.

3.2 Implementasi

Tahap implementasi dilakukan setelah seluruh proses perancangan sistem selesai disusun. Implementasi ini mencakup rangkaian tahapan teknis, mulai dari pengumpulan data komentar YouTube, *preprocessing* data, pemberian label sentimen, hingga pelatihan dan evaluasi model BERT.

3.2.1 Pengumpulan Data

Pada tahap ini, data dikumpulkan dari *platform* YouTube dengan memanfaatkan YouTube data *API*. Proses pengumpulan data dimulai dengan memilih video ceramah Islam berbahasa Indonesia yang relevan dengan penelitian. Komentar-komentar yang terdapat pada video tersebut diambil secara otomatis melalui *API*. Setiap komentar yang dikumpulkan dilengkapi dengan informasi tambahan seperti waktu posting, jumlah suka, dan nama pengguna. Dengan mengotomatisasi pengambilan data ini, peneliti dapat memperoleh dataset komentar yang luas dalam waktu yang relatif singkat, sehingga mempercepat proses analisis sentimen. Hasil dari pengumpulan data ini akan disimpan dalam format terstruktur seperti *xlsx* untuk memudahkan proses selanjutnya.

3.2.2 Labeling Data

Tahap pertama dalam proses ini adalah memberikan label pada data komentar yang telah dikumpulkan dari YouTube. Komentar-komentar ini diberi label untuk menentukan kategori sentimen, yaitu positif dan negatif. Proses pelabelan ini dapat dilakukan secara manual oleh tenaga ahli yang memahami konteks. Tujuan dari labelling adalah memastikan bahwa setiap data memiliki target *output* yang jelas, sehingga model dapat belajar dan mengenali pola sentimen dengan lebih baik selama proses pelatihan.

3.2.3 Preprocessing Data

Setelah data dilabeli, langkah selanjutnya adalah *Preprocessing* data untuk membersihkan dan menyiapkannya agar siap digunakan dalam model pembelajaran. Contoh *Preprocessing* dapat dilihat pada Tabel 3.1 dibawah ini.

Tabel 3.1 *Preprocessing* Data

No	Data sebelum
1	Ceramah ini sangat menginspirasi, sy belajar banyak!
2	Ceramah apa gibah sih
3	Wadaw

a. Case folding

Preprocessing melibatkan beberapa tahap, antara lain Mengonversi semua teks menjadi huruf kecil agar konsistensi data terjaga selama analisis dan menghapus karakter selain huruf dan spasi untuk menjaga hanya teks yang relevan. Contoh *case folding* dapat dilihat pada Tabel 3.2 dibawah ini.

Tabel 3.2 *Case Folding*

Tahapan	Data sebelum	Data sesudah
<i>Case folding</i>	"Ceramah ini sangat menginspirasi, sy belajar banyak!"	"ceramah ini sangat menginspirasi, sy belajar banyak!"
	"Ceramah apa gibah sih"	"ceramah apa gibah sih"
	"Wadaw"	"wadaw"

b. Normalisasi

Mengganti kata-kata singkatan dengan bentuk standarnya. Misalnya, mengganti "utk" menjadi "untuk", "gk" menjadi "gak", dan sebagainya. Contoh normalisasi dapat dilihat pada Tabel 3.3 dibawah ini.

Tabel 3.3 Normalisasi

Tahapan	Data sebelum	Data sesudah
Normalisasi	"ceramah ini sangat menginspirasi, sy belajar banyak!"	"ceramah ini sangat menginspirasi, saya belajar banyak!"
	"Ceramah apa gibah sih"	"ceramah apa gibah sih"
	"Wadaw"	"wadaw"

c. Tokenisasi

Tokenisasi adalah proses memecah teks menjadi unit-unit kecil yang disebut token (bisa berupa kata, tanda baca, atau simbol lainnya). Agar kata tersebut lebih mudah dipahami. Contoh tokenisasi dapat dilihat pada Tabel 3.4 dibawah ini.

Tabel 3.4 Tokenisasi

Tahapan	Data sebelum	Data sesudah
Tokenisasi	"ceramah ini sangat menginspirasi, saya belajar banyak!"	"[ceramah] [ini] [sangat] [menginspirasi] [,] [saya] [belajar] [banyak] [!]"
	"Ceramah apa gibah sih"	"[ceramah] [apa] [gibah] [sih]"
	"Wadaw"	"[wadaw]"

d. *Stopword removal*

Stopwords adalah kata-kata yang sering muncul dalam teks namun tidak membawa informasi penting untuk analisis (misalnya kata "dan", "yang", "di", dll). Contoh *stopword removal* dapat dilihat pada Tabel 3.5 dibawah ini.

Tabel 3.5 *Stopword removal*

Tahapan	Data sebelum	Data sesudah
<i>Stopword removal</i>	"[ceramah] [ini] [sangat] [menginspirasi] [,] [saya] [belajar] [banyak] [!]"	"[ceramah] [ini] [sangat] [menginspirasi] [saya] [belajar] [banyak]"
	"Ceramah apa gibah sih"	"[ceramah] [gibah]"
	"Wadaw"	"[wadaw]"

3.2.4 Menetapkan Nilai

Fungsi `set_seed()` digunakan untuk menetapkan angka acak yang konsisten pada berbagai pustaka yang digunakan dalam eksperimen, seperti *PyTorch* dan *NumPy*, dengan tujuan memastikan hasil eksperimen dapat direproduksi. Dalam eksperimen *machine learning*, penting untuk menetapkan *seed* agar hasil eksperimen tidak bervariasi setiap kali dijalankan. Fungsi ini mengatur *seed* pada tiga bagian, `torch.manual_seed(seed_value)` digunakan untuk menetapkan nilai

pada operasi acak yang dilakukan oleh *PyTorch* pada *CPU*. `np.random.seed(seed_value)` mengatur nilai untuk operasi acak pada pustaka *NumPy* yang sering digunakan dalam pengolahan data dan perhitungan. `torch.cuda.manual_seed_all(seed_value)` mengatur *seed* pada *GPU* jika model dilatih menggunakan perangkat keras akselerasi, seperti *GPU*, yang memungkinkan eksperimen tetap konsisten meskipun menggunakan perangkat keras yang berbeda. Dengan menetapkan *seed* yang sama di setiap tahap, eksperimen akan menghasilkan hasil yang serupa setiap kali dijalankan, sehingga memungkinkan eksperimen yang dapat diulang dan hasil yang konsisten. Baris kode yang menangani penetapan nilai/*seed* dinyatakan dalam kode pada *pseudocode* dibawah ini.

```
def set_seed(seed_value=42):  
    torch.manual_seed(seed_value)  
    np.random.seed(seed_value)  
    torch.cuda.manual_seed_all(seed_value)  
    set_seed(42)
```

3.2.5 Membersihkan Data

Selanjutnya, dilakukan beberapa langkah untuk membersihkan dan memformat data yang akan digunakan dalam penelitian. `df = df.dropna(subset=[text_col, label_col])` berfungsi untuk menghapus baris-baris yang memiliki nilai kosong pada kolom yang berisi teks dan label. Hal ini penting untuk memastikan bahwa hanya data yang lengkap yang digunakan dalam pelatihan dan pengujian model, menghindari masalah yang mungkin timbul jika model

diberikan data yang tidak lengkap. Selanjutnya, `df[text_col] = df[text_col].astype(str)` dan `df[label_col] = df[label_col].astype(str)` mengubah tipe data pada kolom teks dan label menjadi tipe data string. Baris kode yang menangani pembersihan data dinyatakan dalam kode pada *pseudocode* dibawah ini.

```
df = df.dropna(subset=[text_col, label_col])
df[text_col] = df[text_col].astype(str)
df[label_col] = df[label_col].astype(str)
```

3.2.6 Konversi Label

Setelah membersihkan data dilakukan proses pengubahan label teks menjadi angka agar dapat diproses oleh model. `label_map = {label: idx for idx, label in enumerate(df[label_col].unique())}` berfungsi untuk membuat sebuah pemetaan antara setiap label unik yang terdapat pada kolom label dengan sebuah angka indeks. Fungsi `enumerate()` digunakan untuk memberikan indeks numerik pada setiap label yang unik, yang kemudian disimpan dalam dictionary `label_map`. Misalnya, jika label dalam data adalah "positif" dan "negatif", maka masing-masing label akan dipetakan menjadi angka seperti 0 untuk "positif" dan 1 untuk "negatif". Setelah pembuatan pemetaan, `df['encoded_label'] = df[label_col].map(label_map)` digunakan untuk mengganti setiap label pada kolom `label_col` dengan nilai numerik yang sesuai berdasarkan `label_map`. Proses ini menghasilkan kolom baru dalam dataset yang disebut `encoded_label`, yang berisi label yang telah dikonversi menjadi angka. Baris kode yang menangani konversi label dinyatakan dalam kode pada *pseudocode* dibawah ini.

```
label_map = {label: idx for idx, label in
enumerate(df[label_col].unique())}
df['encoded_label'] = df[label_col].map(label_map)
```

3.2.7 BERT Tokenizer

Sebelum diproses dengan model BERT, Selanjutnya, *tokenizer* = *BertTokenizer.from_pretrained('indobenchmark/indobert-base-p1')*, digunakan untuk memuat *BertTokenizer* yang telah dilatih sebelumnya dari model BERT *indobenchmark/indobert-base-p1*. *BertTokenizer* adalah komponen dari pustaka *transformers* yang berfungsi untuk mengonversi teks mentah menjadi token, yaitu unit terkecil yang dapat diproses oleh model BERT. Model *indobenchmark/indobert-base-p1* adalah model BERT yang telah dilatih khusus untuk bahasa Indonesia menggunakan berbagai sumber data, termasuk artikel berita, media sosial, dan lain-lain. *BertTokenizer* didokumentasikan pada *pseudocode* dibawah ini:

```
tokenizer BertTokenizer.from_pretrained('indobenchmark/indobert-
base-p1')
```

Proses tokenisasi itu sendiri melibatkan beberapa tahapan (Israel, 2024) dilakukan sebagai berikut:

a. Tokenisasi

Proses tokenisasi dimulai dengan membagi setiap kalimat menggunakan *WordPiece* menjadi kata-kata atau sub-kata. Setiap kata yang ditokenisasi akan diperiksa apakah sudah ada dalam *vocabulary*. Jika kata tersebut tidak ditemukan

dalam *vocabulary*, *tokenizer* akan membagi kata tersebut menjadi sub-kata yang memiliki kemungkinan muncul paling besar dalam *vocabulary*. Kata yang tidak ditemukan dalam *vocabulary* akan digantikan dengan token [UNK]. Meskipun demikian, penggantian kata-kata yang tidak ada dalam *vocabulary* dengan token [UNK] dapat mengurangi informasi yang hilang dari teks tersebut. Untuk mengatasi hal ini, *tokenizer* akan memecah kata yang tidak ditemukan menjadi dua sub-kata. Sub-kata pertama akan berupa token yang ada dalam *vocabulary*, sementara sub-kata kedua akan diawali dengan simbol ##.

Contoh kalimat : " Ceramah ini sangat menginspirasi, saya belajar banyak!"

Hasil tokennya : ['Ceramah', 'ini', 'sangat', 'menginspirasi', 'saya', 'belajar', 'banyak']



Gambar 3.2 Tokenisasi

Gambar 3.2 menggambarkan proses tokenisasi yang dilakukan menggunakan BERT pada kalimat "Ceramah ini sangat menginspirasi, saya belajar banyak." Dalam gambar, kalimat tersebut dibagi menjadi token-token yang lebih kecil, dengan setiap token diberi highlight warna biru. Tokenisasi ini memisahkan setiap kata atau sub-kata yang ada.

b. Penambahan Token

Setiap kalimat diberi token khusus di bagian awal yang berupa [CLS] dan di bagian akhir berupa [SEP]. Token [CLS] berfungsi sebagai penanda awal kalimat dan digunakan dalam analisis saat melakukan klasifikasi sentimen, karena token ini

berfungsi untuk mengumpulkan informasi rata-rata dari setiap kata dalam kalimat untuk menghasilkan vektor. Sementara itu, token [SEP] berfungsi untuk memisahkan kalimat satu dengan yang lainnya. Setelah diberikan token [CLS] dan [SEP], kalimat tersebut disebut sebagai token *embeddings*. Penambahan token dapat dilihat pada Gambar 3.3 di bawah ini.



Gambar 3.3 Penambahan Token [CLS] dan [SEP]

c. Mengubah Token

Setiap kata atau sub-kata diberikan nomor unik atau ID berdasarkan indeksnya dalam *vocabulary*, yang diurutkan berdasarkan frekuensi kemunculannya. Setiap kata memiliki ID yang berbeda. Proses ini dilakukan karena selama *pre-Training*, model hanya dapat memproses ID dari token. Sebagai contoh, dalam *vocabulary* model ‘indobenchmark/indobert-base-p1’, token [UNK] memiliki ID 1, token [CLS] memiliki ID 2, dan token [SEP] memiliki ID 3, dan [MASK] memiliki ID 4. Penomoran ID ke token sudah menjadi ketetapan. Contoh penomoran token pada kalimat dapat dilihat pada Gambar 3.4 berikut.



Gambar 3.4 Mengubah Token ID

d. Penambahan Token *Padding*

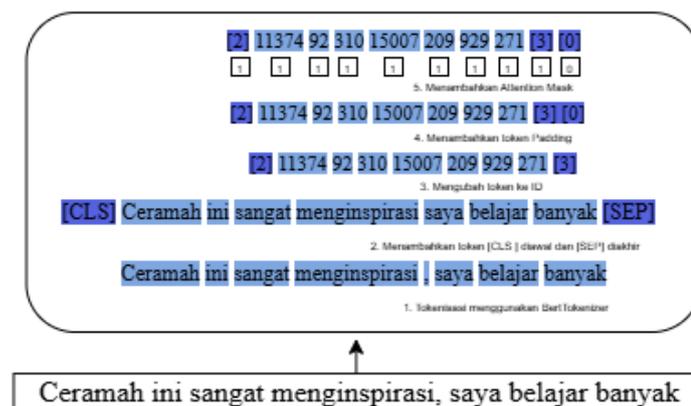
Selanjutnya, kalimat tersebut disesuaikan dengan panjang urutan maksimal yang telah ditentukan, dengan cara mengurangi jumlah token atau menambahkan *Padding* menggunakan token khusus, yaitu [PAD]. Pada Gambar 3.5 menunjukkan tahapan pemberian token *padding*, proses penambahan *padding* hingga mencapai panjang maksimal yang ditetapkan, yaitu 15. Karena panjang kalimat input lebih pendek dari panjang urutan maksimum, token [PAD] ditambahkan di akhir kalimat setelah token [SEP].



Gambar 3.5 Penambahan Token Padding

e. *Sentence Embedding*

Selanjutnya untuk membedakan antara kalimat pertama, kalimat berikutnya, atau *padding*, digunakan *sentence embedding*. Pada contoh ini, angka 1 diberikan untuk kalimat pertama, sementara angka 0 diberikan untuk token [PAD]. Tokenizer akan membedakan kalimat pertama, kalimat selanjutnya, atau *padding* dengan mengacu pada token [SEP], yang berfungsi sebagai pemisah antar kalimat. Pada Gambar 3.6 menunjukkan di mana angka 1 diberikan pada token kalimat pertama, dan hanya token [PAD] yang diberi angka 0 dalam tahap *sentence embedding*.

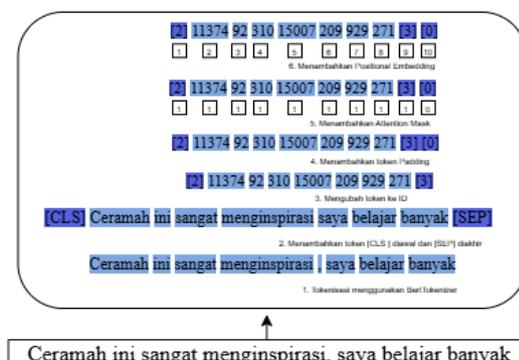


Gambar 3.6 *Sentence Embedding*

f. *Positional Embedding*

Pada tahap selanjutnya, yaitu *positional embedding*, setiap token akan diberikan angka yang menggambarkan posisinya dalam kalimat. Tahap ini penting untuk mengetahui urutan setiap kata dalam kalimat. Dengan demikian, apabila ada kata yang sama tetapi memiliki makna yang berbeda, BERT dapat mengidentifikasi makna yang berbeda berdasarkan posisi kata tersebut dalam kalimat. Contoh

penambahan *positional embedding* pada kalimat dapat dilihat pada gambar 3.7 dibawah ini.



Gambar 3.7 *Positional Embedding*

3.2.8 Pembagian Data

Selanjutnya, $split_ratios = [0.1, 0.3, 0.5]$ mendefinisikan beberapa rasio pembagian data untuk eksperimen. Variabel $split_ratios$ berisi tiga nilai yang masing-masing menunjukkan proporsi data yang akan digunakan untuk pengujian, yaitu 10%, 30%, dan 50% dari total data. Rasio-rasio ini digunakan untuk menguji seberapa baik model dapat melakukan generalisasi pada data pengujian yang berbeda-beda. Kemudian, dalam perulangan *for split in split_ratios:*, dataset dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian. $train_test_split(df, test_size=split, random_state=42)$ adalah fungsi dari pustaka *sklearn* yang membagi dataset df menjadi dua satu untuk pelatihan dan satu untuk pengujian. Parameter $test_size=split$ menentukan berapa persen dari data yang akan digunakan sebagai data pengujian, berdasarkan nilai yang ada pada $split_ratios$. Misalnya, jika $split = 0.1$, maka 10% data akan digunakan untuk pengujian, sedangkan sisanya (90%) digunakan untuk pelatihan. $random_state=42$ memastikan bahwa

pembagian data dilakukan secara acak namun tetap konsisten pada setiap eksperimen. Baris kode yang menangani pembagian data dinyatakan dalam kode pada *pseudocode* dibawah ini.

```
split_ratios = [0.1, 0.3, 0.5]
train_df, test_df = train_test_split(df, test_size=split,
random_state=42)
```

3.2.9 Menyiapkan Data dan Model

Selanjutnya membersihkan teks dan menyiapkan dataset untuk digunakan dalam pelatihan dan pengujian model. Pertama, teks pada dataset pelatihan dan pengujian dibersihkan dengan menggunakan $train_texts = [str(t).strip() \text{ for } t \text{ in } train_df[text_col].tolist() \text{ if } isinstance(t, str) \text{ and } t.strip() \neq ""]$ dan $test_texts = [str(t).strip() \text{ for } t \text{ in } test_df[text_col].tolist() \text{ if } isinstance(t, str) \text{ and } t.strip() \neq ""]$. Pada langkah ini, setiap teks dalam kolom yang ditentukan (misalnya, komentar atau ulasan) dikonversi menjadi string dan kemudian dipangkas dari spasi ekstra di awal dan akhir (menggunakan *metode strip()*). Selain itu, kode ini memastikan bahwa hanya teks yang valid dan tidak kosong yang dipertahankan untuk pelatihan dan pengujian model.

Setelah teks dibersihkan, langkah berikutnya adalah menyiapkan label yang sesuai untuk dataset pelatihan dan pengujian. $train_labels = train_df['encoded_label'].tolist()[:len(train_texts)]$ dan $test_labels = test_df['encoded_label'].tolist()[:len(test_texts)]$ digunakan untuk memastikan bahwa label yang diambil memiliki jumlah yang sama dengan jumlah teks yang

telah dibersihkan. Setiap label yang sebelumnya telah dikodekan dalam kolom *encoded_label* akan dipetakan sesuai dengan jumlah teks setelah pembersihan.

train_dataset dan *test_dataset* dibuat dengan menggunakan kelas *CommentDataset*, yang mengonversi teks dan label menjadi format yang sesuai dengan model BERT. Kelas ini menggunakan *tokenizer* untuk mempersiapkan teks agar bisa diproses oleh model BERT.

Dalam penelitian ini, menggunakan model BERT untuk melakukan tugas klasifikasi teks dengan memanfaatkan kelas *BertForSequenceClassification* dari *Hugging Face Transformers*. Model ini telah dilatih sebelumnya dan dioptimalkan untuk memprediksi label berdasarkan *input* teks. Fungsi *from_pretrained('indobenchmark/indobert-base-p1')* digunakan untuk memuat model BERT yang sudah dilatih sebelumnya. Model yang digunakan adalah varian IndoBERT yang dilatih khusus untuk bahasa Indonesia, dengan nama model *indobenchmark/indobert-base-p1*, yang dapat diakses melalui *Hugging Face Model Hub*. Dengan menggunakan metode *from_pretrained*, penulis memuat bobot dan konfigurasi model BERT yang telah disesuaikan untuk tugas klasifikasi teks dalam bahasa Indonesia.

Argumen *num_labels* pada model berfungsi untuk menentukan jumlah kelas yang akan diprediksi oleh model. Nilai *num_labels* diambil dari hasil fungsi *len()* yang diterapkan pada *label_map*, digunakan memetakan setiap label teks ke indeks numerik. Fungsi *len()* sendiri merupakan fungsi bawaan *Python* yang digunakan untuk menghitung jumlah elemen dalam suatu objek, seperti list. Dalam kasus ini, *len(label_map)* mengembalikan jumlah nilai di dalam *label_map*, yang

merepresentasikan jumlah total kategori atau kelas dalam dataset. Sebagai contoh, jika dataset terdiri dari tiga kategori seperti "positif", "negatif", dan "netral", maka `num_labels` akan bernilai 3, menyesuaikan dengan jumlah kelas tersebut. Baris kode yang menyiapkan data dan model dinyatakan dalam kode pada *pseudocode* dibawah ini.

```
train_texts = [str(t).strip() for t in
train_df[text_col].tolist() if isinstance(t, str) and t.strip()
!= '']

    test_texts = [str(t).strip() for t in
test_df[text_col].tolist() if isinstance(t, str) and t.strip()
!= '']

    train_labels =
train_df['encoded_label'].tolist()[0:len(train_texts)]

    test_labels =
test_df['encoded_label'].tolist()[0:len(test_texts)]

    train_dataset = CommentDataset(train_texts,
train_labels, tokenizer)

    test_dataset = CommentDataset(test_texts, test_labels,
tokenizer)

    for epoch in epochs_list:
        output_dir = f"./bert_split{int((1-
split)*100)}_epoch{epoch}"

        model =
BertForSequenceClassification.from_pretrained(
```

```

        'indobenchmark/indobert-base-pl',
        num_labels=len(label_map)
    )

```

3.2.10 Pengaturan Pelatihan

Selanjutnya, pada pengaturan pelatihan, *TrainingArguments* digunakan untuk mengonfigurasi berbagai parameter penting yang dibutuhkan dalam proses pelatihan model *BERT*. Beberapa aspek utama yang diatur meliputi jumlah *epoch*, ukuran *batch*, serta strategi pencatatan (*logging*). Parameter *output_dir* menentukan lokasi penyimpanan model yang telah dilatih, sedangkan *num_train_epochs* mengatur berapa kali model akan dilatih menggunakan seluruh *dataset*. Ukuran *batch* untuk proses pelatihan dan evaluasi ditentukan oleh *per_device_train_batch_size* dan *per_device_eval_batch_size*, yang dalam penelitian ini diset sebesar 8 sampel per *batch*.

Selain itu, strategi pencatatan *log* diatur menggunakan *logging_strategy="epoch"*, yang berarti pencatatan dilakukan di akhir setiap *epoch*. Model hanya disimpan sekali di akhir pelatihan karena *save_strategy* diatur ke *"no"*. Frekuensi pencatatan *log* selama pelatihan diatur melalui *logging_steps=10*. Sementara itu, *load_best_model_at_end=False* menunjukkan bahwa model terbaik tidak akan dimuat secara otomatis setelah pelatihan selesai. Adapun *report_to=None* menandakan bahwa pelaporan tidak diarahkan ke sistem eksternal seperti *Weights & Biases (W&B)* atau *TensorBoard*. Pengaturan ini seluruhnya didokumentasikan pada *pseudocode* sebagai berikut.

```

Training_args = TrainingArguments(

```

```
        output_dir=output_dir,  
        num_train_epochs=epoch,  
        per_device_train_batch_size=8,  
        per_device_eval_batch_size=8,  
        logging_strategy="epoch",  
        save_strategy="no",  
        logging_dir=f"{output_dir}/logs",  
        logging_steps=10,  
        load_best_model_at_end=False,  
        report_to=None,  
    )
```

3.2.11 Pelatihan Model

Selanjutnya Pelatihan model, objek *trainer* dibuat dengan menggunakan kelas *CustomTrainer*, yang merupakan turunan dari kelas *Trainer* dari pustaka *transformers*. Kelas *CustomTrainer* ini digunakan untuk melatih model BERT dengan tambahan fitur pencatatan *loss* per langkah (step), yang telah dimodifikasi sebelumnya. Objek *trainer* diinisialisasi dengan parameter-parameter berikut: *model=model* yang menunjukkan model BERT yang telah didefinisikan sebelumnya, *args=Training_args* yang berisi pengaturan pelatihan seperti jumlah *epoch*, ukuran *batch*, dan strategi *logging*, serta *train_dataset=train_dataset* dan *eval_dataset=test_dataset* yang masing-masing berisi dataset pelatihan dan pengujian yang telah disiapkan. Setelah objek *trainer* dibuat, proses pelatihan dimulai dengan pemanggilan *trainer.train()*, yang akan melatih model

menggunakan data pelatihan dan mencatat *loss* selama setiap langkah pelatihan.

Pelatihan model didokumentasikan pada *pseudocode* di bawah ini.

```
trainer = CustomTrainer(  
    model=model,  
    args=Training_args,  
    train_dataset=train_dataset,  
    eval_dataset=test_dataset,
```

3.2.12 Pengujian Model

Selanjutnya dilakukan proses prediksi dilakukan setelah model dilatih menggunakan dataset pengujian. Fungsi *trainer.predict(test_dataset)* digunakan untuk menghasilkan prediksi dari model pada dataset pengujian *test_dataset* yang telah dipersiapkan sebelumnya. Fungsi ini menghasilkan output yang berisi nilai *logit* (*output* mentah) untuk setiap kelas, yang menunjukkan seberapa kuat keyakinan model terhadap setiap kelas untuk setiap *input* data.

Setelah itu, hasil prediksi yang berupa *logit* diubah menjadi kelas yang diprediksi *torch.argmax(torch.tensor(preds_output.predictions), axis=1)*. Fungsi *torch.argmax()* digunakan untuk memilih kelas dengan nilai *logit* tertinggi untuk setiap *input* data. *axis=1* menunjukkan bahwa kita ingin memilih nilai tertinggi pada setiap baris setiap prediksi dari hasil *logit* yang dihasilkan untuk setiap kelas. Dengan demikian, setiap *input* dalam dataset pengujian akan dipetakan ke kelas dengan probabilitas tertinggi, yang menjadi hasil prediksi model.

Hasil akhir dari proses ini adalah *preds*, yang berisi indeks kelas yang dipilih sebagai prediksi untuk setiap sampel dalam dataset pengujian. Prediksi ini

kemudian dapat digunakan untuk menghitung matriks evaluasi model, seperti akurasi, dan untuk melakukan analisis lebih lanjut mengenai performa model pada data yang tidak terlihat selama pelatihan.

Sebagai contoh, dalam kasus prediksi sentimen dengan tiga kelas (positif, negatif, netral), jika model menghasilkan *logit* berikut untuk masing-masing kelas *Logit* yang diberikan adalah:

1. Positif: 0.1591
2. Negatif: -0.0185
3. Netral: -0.0998

Proses *softmax* dimulai dengan menghitung eksponensial dari setiap *logit*, e adalah bilangan Euler, yang merupakan bilangan konstanta matematika e adalah sekitar 2.71828. maka hasilnya sebagai berikut:

$$e^{0.1591} = 1.172$$

$$e^{-0.0185} = 0.981$$

$$e^{-0.0998} = 0.905$$

Selanjutnya, jumlahkan eksponensial dari semua kelas:

$$sum = 1.172 + 0.981 + 0.905 = 3.058$$

Setelah itu, hitung probabilitas untuk setiap kelas dengan membagi eksponensial dari *logit* masing-masing kelas dengan jumlah total eksponensial:

$$\text{Negatif : } \sigma(0) = \frac{0.981}{3.058} = 0.321$$

$$\text{Netral : } \sigma(1) = \frac{0.905}{3.058} = 0.296$$

$$\text{Positif: } \sigma(0) = \frac{1.172}{3.058} = 0.383$$

Hasilnya menunjukkan probabilitas untuk setiap kelas, yang selanjutnya dapat digunakan untuk menentukan prediksi model. Berdasarkan probabilitas yang diperoleh, teks ini akan diprediksi sebagai positif, karena kelas positif memiliki probabilitas tertinggi. Pengujian model didokumentasikan pada *pseudocode* berikut.

```

preds_output = trainer.predict(test_dataset)

        preds =

torch.argmax(torch.tensor(preds_output.predictions), axis=1)

```

3.2.13 Evaluasi Model

Evaluasi model BERT dilakukan untuk mengevaluasi kemampuan model dalam mengklasifikasikan sentimen komentar YouTube. Evaluasi ini penting untuk memastikan bahwa model dapat memprediksi sentimen dengan akurat (positif, negatif, atau netral). Beberapa metrik yang digunakan dalam pengujian ini, seperti akurasi, presisi, *recall*, dan *F1-Score*, merupakan hasil perhitungan yang didasarkan pada *confusion matrix*. *Confusion matrix* sendiri berfungsi untuk memberikan gambaran rinci mengenai jumlah prediksi yang benar (*true positive* dan *true negative*) serta prediksi yang salah (*false positive* dan *false negative*), sehingga membantu dalam mengevaluasi kinerja model secara komprehensif (Arifqi et al., 2024).

1. Akurasi yaitu mengukur seberapa sering model memberikan prediksi yang benar dibandingkan dengan total jumlah data. Rumusnya adalah:

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{total}} \quad (3.2)$$

2. *Presisi* yaitu mengukur seberapa banyak prediksi positif yang benar-benar positif. Rumusnya adalah:

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

3. *Recall* yaitu mengukur seberapa banyak prediksi positif yang benar-benar ditemukan oleh model dari seluruh data positif. Rumusnya adalah

$$Recall = \frac{TP}{TP+FN} \quad (3.4)$$

4. *F1-Score* adalah ukuran gabungan dari presisi dan *recall* yang memberikan gambaran tentang keseimbangan antara keduanya. Rumusnya adalah:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3.5)$$

Keterangan :

- a) *TP* = *true positive*
- b) *TN* = *true negative*
- c) *FP* = *false positive*
- d) *FN* = *false negative*

True Positive merupakan jumlah data yang benar-benar termasuk dalam kelas tertentu (misalnya, negatif, netral, atau positif) dan diprediksi oleh model sebagai kelas yang sama. *True Negative* yaitu jumlah data yang benar-benar tidak termasuk dalam kelas tertentu, dan diprediksi dengan benar sebagai kelas lain. *False Positive* merupakan jumlah data yang sebenarnya tidak termasuk dalam kelas tertentu, tetapi diprediksi oleh model sebagai kelas tersebut. *False Negative* adalah jumlah data yang sebenarnya termasuk dalam kelas tertentu, tetapi diprediksi oleh model sebagai kelas lain.

Sebagai contoh, data sampel berikut digunakan untuk menguji performa model BERT dalam mengklasifikasikan sentimen komentar. Data ini mencakup

komentar, label sentimen manual, dan hasil prediksi model, yang kemudian dijadikan dasar dalam perhitungan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Rincian data tersebut dapat dilihat pada Tabel 3.6.

Tabel 3.6 Evaluasi Model

No	Komentar	Label Sentimen	Prediksi
1	Ceramah ini sangat menginspirasi, saya belajar banyak!	Positif	Positif
2	Saya tidak setuju dengan pendapat yang disampaikan.	Negatif	Positif
3	Topik ceramah ini menarik, tapi saya perlu penelitian lebih lanjut.	Netral	Netral

Setelah melakukan prediksi menggunakan model BERT, selanjutnya menghitung metrik pengujian berdasarkan hasil prediksi menggunakan rumus sebagaimana berikut :

1. Jumlah prediksi yang benar = 2 (Komentar 1 dan komentar 3)

Jumlah total data = 3 Komentar

$$\text{Akurasi} = \frac{2}{3} = 0.66$$

2. Positif benar = 1 (komentar 1)

Positif salah = 1 (komentar 2, yang harusnya negatif)

$$\text{Precision} = \frac{1}{1+1} = 0.5$$

3. Positif benar = 1 (Komentar 1)

Negatif salah = 0 (Tidak ada komentar positif yang terlewatkan)

$$\text{Recall} = \frac{1}{1+0} = 1.0$$

4. *F1-Score*

Menggunakan precision dan *recall* yang sudah dihitung :

$$\text{F1 - Score} = 2 \times \frac{0.5 \times 1.0}{0.5 + 1.0} = 0.67$$

BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini, akan dipaparkan hasil implementasi model *Bidirectional Encoder Representations from Transformers* (BERT) dalam analisis sentimen komentar YouTube pada video ceramah Islam. Penjelasan mencakup tahapan pengolahan data, hasil analisis sentimen, evaluasi kinerja model, serta pembahasan mengenai temuan yang diperoleh dengan mengintegrasikan perspektif Islam dalam penelitian ini.

4.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan komentar dari video ceramah Islam di *platform* YouTube. Komentar tersebut diambil dari video ceramah yang disampaikan oleh Habib Bahar bin Smith, dan diperoleh melalui proses *web crawling* menggunakan YouTube Data *API v3* yang sebelumnya harus mendaftar di *Google Cloud* untuk mendapatkan *API Keys*. Jumlah data *crawling* sebanyak 1117 komentar selanjutnya data tersebut di *labeling* secara manual oleh ahli bahasa Indonesia sarjana Pendidikan Bahasa Indonesia Universitas Brawijaya yang sekarang aktif dalam menulis sebuah karya. *Labeling* dilakukan dengan 3 kriteria yang ditentukan yaitu Negatif, Netral, dan Positif. Setelah dilakukan labelling menghasilkan 672 komentar negatif, 274 komentar netral, dan 171 komentar positif.

Tabel 4.1 menyajikan 5 komentar teratas dan 5 komentar yang terbagi menjadi tiga kategori sentimen utama positif, negatif, dan netral. Komentar positif mencerminkan perasaan mendukung, mendoakan, atau mengungkapkan rasa

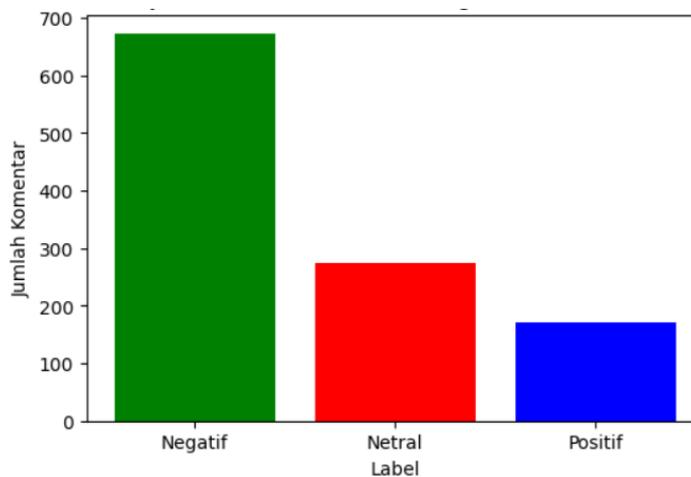
syukur, seperti pada komentar “I love u habib♥” dan “Amiin🙏🙏”, yang menunjukkan ekspresi cinta dan doa. Sebaliknya, komentar negatif berisi keluhan atau ketidakpuasan, seperti pada komentar “Bicaranya sampe ngeden2 gitu apa dak capek tu”, yang menunjukkan kritik terhadap situasi atau perilaku tertentu. Terakhir, komentar netral berisi pernyataan yang tidak mengandung emosi kuat, seperti “*Wallohu alam bisowab*”, yang lebih bersifat pernyataan atau doa tanpa penilaian emosional yang jelas. Pembagian komentar ini penting untuk analisis sentimen, memberikan wawasan tentang bagaimana masyarakat merespons berbagai topik atau peristiwa yang terjadi.

Tabel 4.1 Data Berlabel

No	Komentar	Label_Manual
1	Bicaranya sampe ngeden2 gitu apa dak capek tu	Negatif
2	Nanya sendiri dijawab sendiri	Negatif
3	Amit qmit	Negatif
4	<i>I love u habib♥</i>	Positif
5	<i>Wallohu alam bisowab</i>	Netral
6	Semoga acaranya lancar dan sukses	Positif
7	Amiin🙏🙏	Positif
8	Nderek tabarukan majlis maulid-ilmu¶ kyai-masaikh-habaib	Positif
9	Alhamdulillah - magelang nyimak bib..	Positif
10	Alhamdulillah 🙏🙏	Positif

Dataset tersebut dapat diilustrasikan melalui grafik yang terlihat pada gambar 4.1 dibawah. Gambar ini menunjukkan grafik batang yang menggambarkan jumlah komentar berdasarkan kategori sentimen, yaitu positif, negatif, dan netral. Dari grafik, terlihat bahwa komentar dengan label negatif memiliki jumlah yang paling tinggi, mencapai lebih dari 600 komentar, yang ditandai dengan warna hijau. Di sisi lain, komentar netral berjumlah sekitar 200, ditandai dengan warna merah,

sementara komentar positif memiliki jumlah paling sedikit, dengan sekitar 100 komentar, yang ditandai oleh warna biru.



Gambar 4.1 Grafik Komentar

4.2 *Prepossessing Data*

Setelah data dikumpulkan dan dilakukan proses *labelling*, langkah berikutnya adalah *Preprocessing* data untuk memastikan bahwa teks komentar siap digunakan dalam analisis sentimen. Proses ini bertujuan untuk membersihkan data dari karakter yang tidak relevan, menyamakan format teks, serta menghapus kata-kata yang tidak memiliki makna signifikan dalam analisis sentimen. Tahapan *Preprocessing* yang dilakukan meliputi *Case folding*, normalisasi, tokenisasi, dan *Stopword removal*.

4.2.1 *Case folding*

Case folding adalah tahap awal dalam *preprocessing* data yang bertujuan untuk menyamakan format teks dengan mengubah semua huruf menjadi huruf kecil serta menghapus karakter yang tidak diperlukan. Tujuan dari proses ini adalah

untuk menghindari perbedaan akibat penggunaan huruf besar dan kecil yang dapat memengaruhi hasil analisis sentimen. Proses ini dilakukan dengan menerapkan sebuah fungsi yang bekerja dalam tiga langkah utama. Pertama, semua huruf dalam teks dikonversi menjadi huruf kecil menggunakan `text.lower()`, sehingga tidak ada perbedaan antara huruf besar dan kecil dalam analisis. Kedua, karakter selain huruf seperti angka, tanda baca, dan simbol dihapus menggunakan `re.sub(r'^a-z\s]', '', text)`, agar teks hanya berisi kata-kata yang bermakna. Terakhir, spasi ganda yang mungkin terdapat dalam teks dihapus menggunakan `re.sub(r'\s+', ' ', text).strip()`, sehingga teks lebih rapi dan konsisten.

Tabel 4.2 menunjukkan hasil dari tahap *Case folding* dalam proses *Preprocessing* data komentar. Pada tahap ini, semua teks diubah menjadi huruf kecil untuk menghindari perbedaan antara kata yang memiliki kapitalisasi berbeda, seperti yang terlihat pada komentar "I love u habib♥" yang diubah menjadi "i love u habib". Selain itu, karakter selain huruf, seperti angka, tanda baca, dan simbol, juga dihapus untuk memastikan hanya kata-kata yang relevan yang diproses lebih lanjut. Sebagai contoh, komentar "Amiin🙏🙏" diubah menjadi "amiin", dan "Alhamdulillah👏👏" menjadi "alhamdulillah".

Tabel 4.2 menunjukkan hasil dari tahap *Case folding* dalam proses *Preprocessing* data komentar. Pada tahap ini, semua teks diubah menjadi huruf kecil untuk menghindari perbedaan antara kata yang memiliki kapitalisasi berbeda, seperti yang terlihat pada komentar "I love u habib♥" yang diubah menjadi "i love u habib". Selain itu, karakter selain huruf, seperti angka, tanda baca, dan simbol, juga dihapus untuk memastikan hanya kata-kata yang relevan

yang diproses lebih lanjut. Sebagai contoh, komentar "Amiin 🙏🙏" diubah menjadi "amiin", dan "Alhamdulillah 🙏🙏" menjadi "alhamdulillah".

Tabel 4.2 Hasil *Case Folding*

No	Komentar	Case folding
1	Bicaranya sampe ngeden2 gitu apa dak capek tu	bicaranya sampe ngeden gitu apa dak capek tu
2	Nanya sendiri dijawab sendiri	nanya sendiri dijawab sendiri
3	Amit qmit	amit qmit
4	I love u habib ❤️	❤️
5	Wallohu alam bisowab	wallohu alam bisowab
6	Semoga acaranya lancar dan sukses	semoga acaranya lancar dan sukses
7	Amiin 🙏🙏	amiin
8	Nderek tabarukan majlis maulid-ilmu¶ kyai-masaikh-habaib	nderek tabarukan majlis maulid-ilmuamp para kyai masaikh habaib
9	Alhamdulillah - magelang nyimak bib..	alhamdulillah magelang nyimak bib
10	Alhamdulillah 🙏🙏	alhamdulillah

4.2.2 Normalisasi

Normalisasi adalah proses dalam *Preprocessing* data yang bertujuan untuk mengubah kata-kata tidak baku atau singkatan menjadi bentuk yang lebih standar dalam bahasa Indonesia. Proses ini penting untuk memastikan bahwa model dapat memahami teks dengan lebih baik tanpa dipengaruhi oleh variasi penulisan yang tidak seragam. Dalam penelitian ini, normalisasi dilakukan dengan menggunakan kamus normalisasi yang berisi daftar kata tidak baku dan padanannya dalam bentuk standar. Proses ini bekerja dengan mencocokkan setiap kata dalam teks dengan daftar kata dalam kamus normalisasi, jika kata tersebut ditemukan dalam kamus, maka akan digantikan dengan bentuk bakunya.

Sebagai contoh, kata-kata seperti "utk" akan diubah menjadi "untuk", "gk" menjadi "gak", "mntp" menjadi "mantap", "dg" menjadi "dengan", dan "tdk" menjadi "tidak". Kata-kata lain seperti "cramh" yang merupakan singkatan dari

"ceramah" serta "blm" yang berarti "belum" juga akan dinormalisasi agar teks lebih mudah dipahami oleh model.

Tabel 4.3 menunjukkan hasil dari tahap Normalisasi dalam *Preprocessing* data komentar. Pada tahap ini, kata-kata tidak baku atau singkatan diubah menjadi bentuk yang lebih standar dalam bahasa Indonesia.

Tabel 4.3 Hasil Normalisasi

No	Komentar	Normalisasi
1	Bicaranya sampe ngeden2 gitu apa dak capek tu	bicaranya sampe ngeden gitu apa dak capek tu
2	Nanya sendiri dijawab sendiri	nanya sendiri dijawab sendiri
3	Amit qmit	amit amit
4	I love u habib♥	i love u habib
5	Wallohu alam bisowab	wallohu alam bisowab
6	Semoga acaranya lancar dan sukses	semoga acaranya lancar dan sukses
7	Amiin 🙏🙏	amiin
8	Nderek tabarukan majlis maulid-ilmu¶ kyai-masaikh-habaib	nderek tabarukan majlis maulid-ilmuamp para kyai masaikh habaib
9	Alhamdulillah - magelang nyimak bib..	alhamdulillah magelang nyimak bib
10	Alhamdulillah 🙏🙏	alhamdulillah

4.2.3 Tokenisasi

Tokenisasi adalah proses memecah teks menjadi unit-unit kecil yang disebut token, biasanya berupa kata atau frasa pendek. Proses ini penting untuk membantu model analisis sentimen memahami teks lebih efektif. Dalam penelitian ini, tokenisasi dilakukan menggunakan pustaka *spaCy*, yang secara otomatis memisahkan setiap kata dalam teks yang telah dinormalisasi.

Proses tokenisasi dilakukan dengan menggunakan model bahasa Indonesia dari *spaCy*. Fungsi yang digunakan membaca teks yang telah dinormalisasi dan mengonversinya menjadi daftar token. Setiap kata dalam teks akan dipisahkan sebagai token individu agar dapat dianalisis lebih mudah oleh model. Proses ini diterapkan pada setiap teks dalam dataset dengan menggunakan metode *apply()*.

Tabel 4.4 menunjukkan hasil dari tahap Tokenisasi dalam *Preprocessing* data komentar. Pada tahap ini, setiap komentar dipecah menjadi token atau kata-kata terpisah. Misalnya, komentar "Bicaranya sampe ngeden2 gitu apa dak capek tu" dipecah menjadi ['bicaranya', 'sampe', 'ngeden', 'gitu', 'apa', 'dak', 'capek', 'tu'], dan komentar "I love u habib♥" menjadi ['i', 'love', 'u', 'habib♥'].

Tabel 4.4 Hasil Tokenisasi

No	Komentar	Tokenisasi
1	Bicaranya sampe ngeden2 gitu apa dak capek tu	['bicaranya', 'sampe', 'ngeden', 'gitu', 'apa', 'dak', 'capek', 'tu']
2	Nanya sendiri dijawab sendiri	['nanya', 'sendiri', 'dijawab', 'sendiri']
3	Amit qmit	['amit', 'amit']
4	I love u habib♥	['i', 'love', 'u', 'habib♥']
5	Wallohu alam bisowab	['wallohu', 'alam', 'bisowab']
6	Semoga acaranya lancar dan sukses	['semoga', 'acaranya', 'lancar', 'dan', 'sukses']
7	Amin 🙏🙏	['amiin']
8	Nderek tabarukan majlis maulid-ilmu¶ kyai-masaikh-habaib	['nderek', 'tabarukan', 'majlis', 'maulid', 'ilmuamp', 'para', 'kyai', 'masaikh', 'habaib']
9	Alhamdulillah - magelang nyimak bib..	['alhamdulillah', 'magelang', 'nyimak', 'bib']
10	Alhamdulillah 🙏🙏	['alhamdulillah']

4.2.4 *Stopword removal*

Stopword removal adalah proses menghapus kata-kata umum yang tidak memiliki makna signifikan dalam analisis sentimen, seperti "dan", "yang", atau "di". Tujuan tahap ini adalah untuk meningkatkan akurasi model dengan menghilangkan kata-kata yang tidak berkontribusi pada pemahaman konteks utama teks.

Dalam penelitian ini, proses ini dilakukan menggunakan pustaka Sastrawi, yang menyediakan daftar kata umum dalam bahasa Indonesia untuk dihapus dari

teks. Dengan menerapkan *Stopword removal*, teks menjadi lebih bersih dan lebih fokus pada kata-kata yang memiliki makna dalam analisis sentimen.

Tabel 4.5 menunjukkan hasil dari tahap *Stopword removal* dalam *Preprocessing* data komentar. Pada tahap ini, kata-kata yang dianggap tidak memiliki makna penting, seperti kata sambung atau kata umum, dihapus dari setiap komentar. Misalnya, pada komentar "Bicaranya sampe ngeden2 gitu apa dak capek tu", kata "apa" dan "dak" dihapus, sehingga hasilnya menjadi ['bicaranya', 'sampe', 'ngeden2', 'gitu', 'capek']. Begitu juga pada komentar "Semoga acaranya lancar dan sukses", kata "dan" dihapus, menghasilkan ['semoga', 'acaranya', 'lancar', 'sukses'].

Tabel 4.5 Hasil *Stopword Removal*

No	Komentar	<i>Stopword removal</i>
1	Bicaranya sampe ngeden2 gitu apa dak capek tu	['bicaranya', 'sampe', 'ngeden2', 'gitu', 'capek']
2	Nanya sendiri dijawab sendiri	['nanya', 'sendiri', 'dijawab']
3	Amit qmit	['amit']
4	I love u habib ❤️	['i', 'love', 'u', 'habib']
5	Wallohu alam bisowab	['wallohu', 'alam', 'bisowab']
6	Semoga acaranya lancar dan sukses	['semoga', 'acaranya', 'lancar', 'sukses']
7	Amiin 🙏🙏	['amiin']
8	Nderek tabarukan majlis maulid-ilmu¶ kyai-masaikh-habaib	['nderek', 'tabarukan', 'majlis', 'maulid', 'ilmuamp', 'para', 'kyai', 'masaikh', 'habaib']
9	Alhamdulillah - magelang nyimak bib..	['alhamdulillah', 'magelang', 'nyimak', 'bib']
10	Alhamdulillah 🙏🙏	['alhamdulillah']

4.3 Pembagian Data

Dalam penelitian ini, data komentar yang telah melalui proses *Preprocessing* berjumlah 1.117 komentar, yang diperoleh dari video ceramah Habib Bahar bin Smith di *platform* YouTube. Setelah tahap pembersihan dan pengolahan awal selesai dilakukan, seluruh data tersebut kemudian dibagi menjadi dua bagian, yaitu data latih (*Training set*) dan data uji (*testing set*).

Tujuan dari pembagian ini adalah untuk mengevaluasi kemampuan model BERT dalam menggeneralisasikan hasil pelatihan terhadap data baru yang belum pernah dilihat sebelumnya. Untuk memastikan pengujian dilakukan secara menyeluruh dan adil, dilakukan eksperimen dengan beberapa skenario pembagian data yang berbeda menggunakan pendekatan proporsional tertentu.

Proses pembagian dilakukan secara acak dengan memanfaatkan fungsi *train_test_split()* dari *library Scikit-learn*, disertai parameter *random_state = 42* agar hasil pembagian bersifat tetap saat dijalankan ulang. Strategi ini digunakan untuk menghindari bias dalam evaluasi dan memastikan keakuratan hasil eksperimen.

Tabel 4.6 menunjukkan pembagian data antara data latih dan data uji dengan berbagai perbandingan proporsi. Pembagian ini bertujuan untuk melatih model menggunakan data latih dan menguji kinerjanya menggunakan data uji yang terpisah. Pada perbandingan 90:10, 90% dari total 1117 data digunakan untuk data latih, yaitu sebanyak 1005 data, sementara 10% sisanya (112 data) digunakan untuk data uji. Pada perbandingan 70:30, data latih berjumlah 782 dan data uji sebanyak 335, sementara pada perbandingan 50:50, data latih berjumlah 558 dan data uji 559. Pembagian ini mempengaruhi proses pelatihan dan evaluasi model, dengan proporsi data latih yang lebih besar memberikan kesempatan bagi model untuk belajar lebih banyak, sementara data uji yang cukup tetap diperlukan untuk mengukur kinerja model dengan akurat.

Tabel 4.6 Pembagian Data

No	Train-Test <i>Split</i>	Jumlah data latih	Jumlah data uji	Total
1	90:10	1005	112	1117
2	70:30	782	335	1117
3	50:50	558	559	1117

Dengan melakukan variasi pembagian seperti ini, diharapkan dapat diketahui sejauh mana jumlah data latih memengaruhi performa model, serta menguji stabilitas prediksi model pada proporsi data uji yang berbeda. Hasil dari eksperimen ini menjadi dasar dalam menentukan kombinasi parameter terbaik untuk klasifikasi komentar menggunakan model BERT.

4.4 Pelatihan model BERT

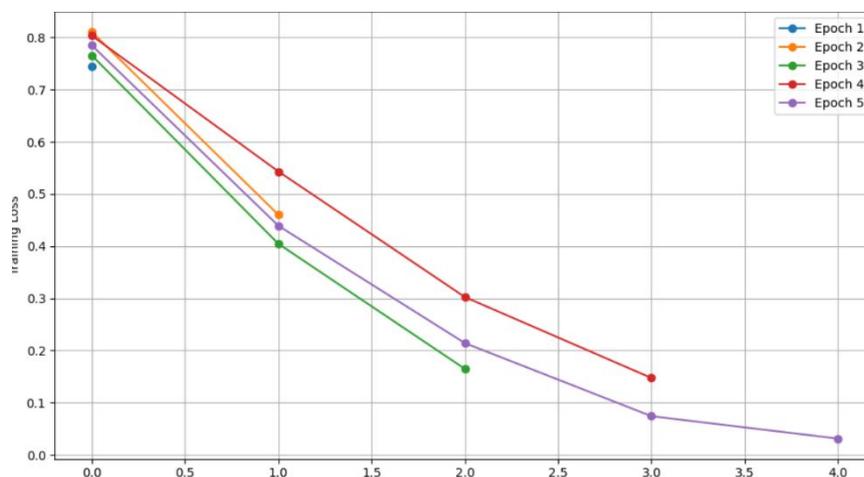
Pelatihan model dilakukan untuk mengoptimalkan kinerja BERT dalam melakukan klasifikasi sentimen terhadap komentar YouTube. Dalam penelitian ini, model yang digunakan *IndoBERT-base-pl* yang dilatih ulang menggunakan dataset hasil *Preprocessing*. Proses pelatihan ini bertujuan untuk menyesuaikan model dengan karakteristik data yang digunakan, sehingga model dapat mengenali pola-pola bahasa yang berkaitan dengan kategori sentimen secara lebih spesifik.

Proses pelatihan dilakukan menggunakan pustaka *Transformers* dari *HuggingFace* dan *API Trainer*, yang dirancang untuk mempermudah pelatihan model berbasis transformer. Dataset dibagi ke dalam dua bagian, yaitu data latih dan data uji, dengan tiga skenario pembagian: 90:10, 70:30, dan 50:50. Selain itu, dilakukan variasi jumlah *epoch* dari 1 hingga 5 pada setiap skenario untuk mengevaluasi pengaruhnya terhadap akurasi model.

4.4.1 Skenario 1

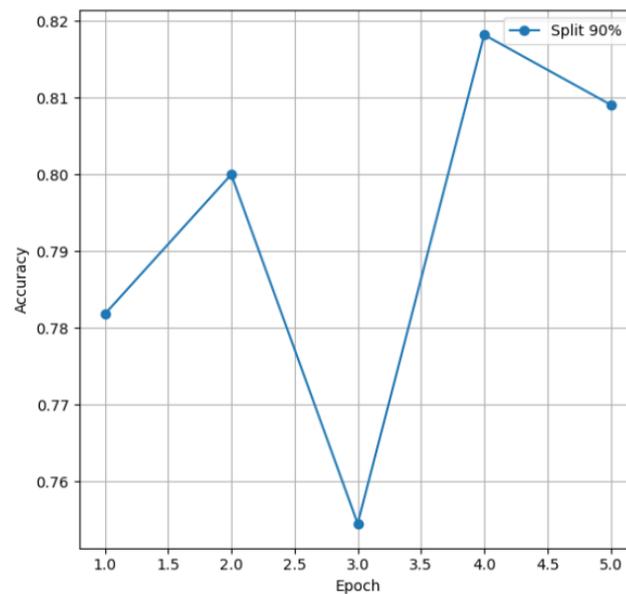
Pada Skenario 1 dengan pembagian data 90:10, pembelajaran model selama 5 *epoch* menunjukkan tren yang signifikan dalam hal *Training Loss* dan *Accuracy*. Pembagian data ini memberikan 90% data untuk pelatihan (data latih) dan 10%

untuk pengujian (data uji). Berikut adalah analisis lebih rinci terkait perkembangan model pada setiap *epoch*.



Gambar 4.2 *Training Loss Split 90*

Pada Gambar 4.2 terdapat grafik *Training Loss* Pada *epoch* pertama, *Training Loss* dimulai dengan nilai 0.74. Pada *epoch* kedua, nilai *Training Loss* menurun menjadi 0.46, yang menunjukkan bahwa model mulai belajar dengan lebih baik dari data pelatihan. Pada *epoch* ketiga, *Training Loss* turun lagi menjadi 0.16, menunjukkan bahwa model semakin mengurangi kesalahan dalam prediksi. Nilai *Training Loss* tetap stabil di 0.14 pada *epoch* keempat, namun pada *epoch* kelima, *Training Loss* mencapai nilai terendah yaitu 0.03, yang menunjukkan peningkatan efisiensi model dalam mengurangi kesalahan.

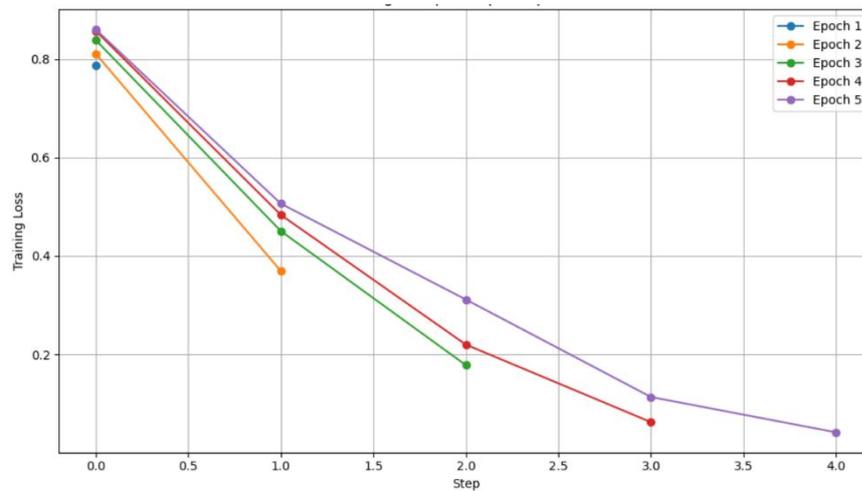


Gambar 4.3 Grafik Akurasi *Split* 90

Pada Gambar 4.3 terdapat grafik akurasi, terlihat *Accuracy* mengalami peningkatan yang konsisten seiring berjalannya waktu. Pada *epoch* pertama, *Accuracy* tercatat 0.78, yang menunjukkan performa awal model. Pada *epoch* kedua, *Training Loss* menurun, *Accuracy* tetap naik pada angka 0.80, menunjukkan bahwa model mulai mengurangi kesalahan, performa prediksi lebih signifikan. Pada *epoch* ketiga, *Accuracy* menurun menjadi 0.75. Pada *epoch* keempat, *Accuracy* meningkat lebih lanjut menjadi 0.81, dan pada *epoch* kelima, *Accuracy* menurun di angka 0.80.

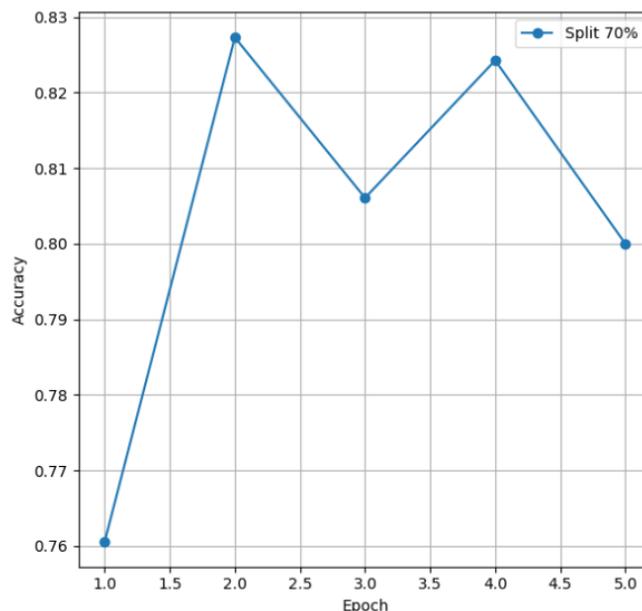
4.4.2 Skenario 2

Pada Skenario 2 dengan pembagian data 70:30, yang berarti 70% data digunakan untuk pelatihan (data latih) dan 30% digunakan untuk pengujian (data uji), model dilatih selama 5 *epoch*. Proses ini menunjukkan bagaimana *Training Loss* dan *Accuracy* berkembang dari *epoch* pertama hingga kelima.



Gambar 4.4 Grafik *Training Loss Split 70*

Pada Gambar 4.4 terdapat grafik *Training Loss* menunjukkan penurunan yang cukup signifikan selama lima *epoch*. Pada *epoch* pertama, *Training Loss* dimulai dengan nilai 0.78, yang menunjukkan bahwa model masih cukup jauh dari prediksi yang akurat. Namun, pada *epoch* kedua, *Training Loss* menurun menjadi 0.36, yang mengindikasikan adanya perbaikan dalam performa model. Pada *epoch* ketiga, *Training Loss* kembali turun menjadi 0.17, menunjukkan bahwa model semakin efisien dalam mengurangi kesalahan. Pada *epoch* keempat, *Training Loss* sedikit menurun menjadi 0.06, dan akhirnya mencapai nilai terendah 0.04 pada *epoch* kelima, yang menunjukkan peningkatan yang signifikan dalam efisiensi model.



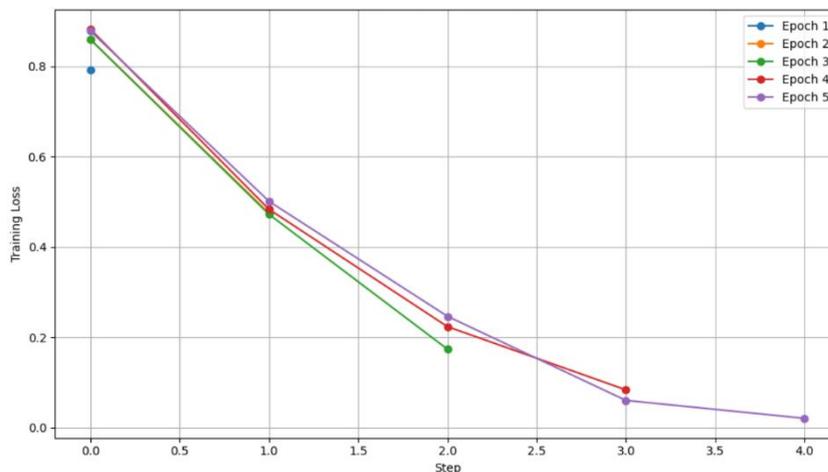
Gambar 4.5 Grafik Akurasi *Split 70*

Pada Gambar 4.5 terdapat grafik *Accuracy* untuk *Split 70%*, pergerakan menunjukkan fluktuasi yang menarik. Pada *epoch* pertama, *Accuracy* dimulai dengan 0.76, yang cukup rendah jika dibandingkan dengan beberapa pembagian data lainnya. Namun, pada *epoch* kedua, meskipun *Training Loss* menurun, *Accuracy* meningkat menjadi 0.82, menunjukkan bahwa model lebih akurat dalam memprediksi data pelatihan. Pada *epoch* ketiga, *Accuracy* sedikit menurun menjadi 0.80. Pada *epoch* keempat, *Accuracy* mengalami kenaikan sedikit menjadi 0.82. Pada *epoch* kelima, *Accuracy* kembali menurun menjadi 0.80, menunjukkan bahwa model berhasil mencapai performa yang cukup stabil dan akurat meskipun terjadi fluktuasi pada beberapa *epoch* sebelumnya.

4.4.3 Skenario 3

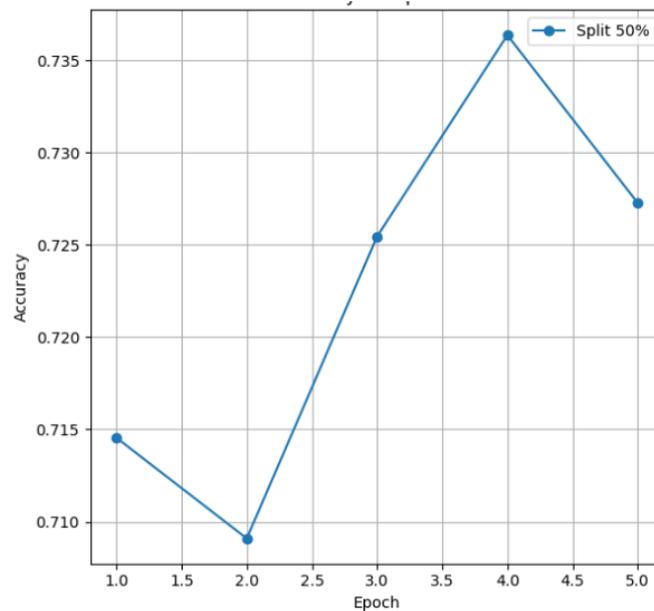
Pada Skenario 3 dengan pembagian data 50:50, di mana 50% data digunakan untuk pelatihan dan 50% untuk pengujian, grafik *Training Loss* dan

Accuracy menunjukkan perkembangan yang berbeda dibandingkan dengan pembagian data sebelumnya (90:10 dan 70:30).



Gambar 4.6 Grafik *Training Loss Split 50*

Pada Gambar 4.6 terdapat grafik *Training Loss* untuk *Split 50%*, terlihat bahwa setiap *epoch* menunjukkan *Training Loss* menunjukkan penurunan yang signifikan dari *epoch* pertama hingga *epoch* kelima. Pada *epoch* pertama, *Training Loss* dimulai dengan nilai 0.79, yang menunjukkan bahwa model belum cukup efisien dalam memprediksi data. Pada *epoch* kedua, *Training Loss* menurun menjadi 0.48, menunjukkan adanya perbaikan dalam kemampuan model untuk mengurangi kesalahan. Pada *epoch* ketiga, *Training Loss* kembali menurun menjadi 0.17, yang menunjukkan bahwa model semakin baik dalam mempelajari data pelatihan. Pada *epoch* keempat, *Training Loss* turun lebih jauh menjadi 0.08, dan akhirnya mencapai nilai terendah 0.02 pada *epoch* kelima, menunjukkan bahwa model semakin efisien dalam meminimalkan kesalahan.



Gambar 4.7 Grafik Akurasi *Split 50*

Pada Gambar 4.7 terdapat grafik *Accuracy* untuk *Split 50%*, dapat dilihat bahwa *Accuracy* menunjukkan fluktuasi yang tidak stabil. Pada *epoch* pertama, *Accuracy* tercatat 0.71, yang menunjukkan performa awal yang cukup rendah. Pada *epoch* kedua, *Training Loss* menurun, *Accuracy* juga menurun menjadi 0.70. Pada *epoch* ketiga, *Accuracy* naik sedikit di angka 0.72, meskipun *Training Loss* terus menurun. Pada *epoch* keempat, *Accuracy* sedikit meningkat menjadi 0.73, tetapi pada *epoch* kelima, *Accuracy* kembali menurun menjadi 0.72. Penurunan ini mungkin menunjukkan bahwa meskipun *Training Loss* menurun drastis, model tidak dapat memanfaatkan data uji dengan baik, yang bisa disebabkan oleh overfitting atau ketidakseimbangan dalam data pelatihan dan pengujian.

4.5 Evaluasi Model

Evaluasi dilakukan untuk menilai performa model BERT dalam mengklasifikasikan komentar ke dalam tiga kategori sentimen: negatif, netral, dan

positif. Evaluasi dilakukan menggunakan *confusion matrix* yang disajikan dalam Tabel 4.7.

Tabel 4.7 Hasil Evaluasi

Skenario	Epoch	Akurasi	Negatif			Positif			Netral		
			P	R	F1	P	R	F1	P	R	F1
90:10	1	0.78	0.76	0.90	0.83	0.94	0.89	0.92	0.70	0.47	0.56
90:10	2	0.80	0.80	0.90	0.85	0.89	0.89	0.89	0.73	0.53	0.62
90:10	3	0.75	0.78	0.84	0.81	0.88	0.79	0.83	0.61	0.57	0.59
90:10	4	0.82	0.83	0.93	0.88	0.93	0.74	0.82	0.73	0.63	0.68
90:10	5	0.81	0.81	0.92	0.86	0.94	0.84	0.89	0.71	0.57	0.63
70:30	1	0.76	0.77	0.93	0.84	0.71	0.82	0.76	0.81	0.28	0.42
70:30	2	0.83	0.87	0.90	0.88	0.79	0.75	0.77	0.74	0.68	0.71
70:30	3	0.81	0.87	0.87	0.87	0.81	0.68	0.74	0.65	0.71	0.68
70:30	4	0.82	0.86	0.89	0.88	0.85	0.75	0.80	0.70	0.68	0.69
70:30	5	0.80	0.82	0.91	0.86	0.84	0.70	0.77	0.70	0.57	0.63
50:50	1	0.71	0.73	0.94	0.82	0.75	0.65	0.69	0.53	0.19	0.28
50:50	2	0.68	0.77	0.83	0.80	0.77	0.62	0.69	0.49	0.46	0.47
50:50	3	0.73	0.76	0.88	0.82	0.76	0.65	0.70	0.54	0.37	0.44
50:50	4	0.74	0.78	0.86	0.82	0.75	0.73	0.74	0.56	0.43	0.49
50:50	5	0.73	0.77	0.88	0.82	0.72	0.68	0.70	0.57	0.38	0.46

Pada Tabel 4.7 akurasi tertinggi pada skenario 90:10 tercatat 0.82 pada *epoch* ke-4, sementara akurasi terendah adalah 0.75 pada *epoch* ke-3. Pada skenario 70:30, akurasi tertinggi adalah 0.83 pada *epoch* ke-2, dan akurasi terendah adalah 0.76 pada *epoch* ke-1. Di skenario 50:50, akurasi tertinggi mencapai 0.74 pada *epoch* ke-4, dengan akurasi terendah 0.68 pada *epoch* ke-2. Secara keseluruhan, skenario 70:30 memberikan akurasi terbaik, diikuti oleh 90:10, sementara 50:50 menunjukkan hasil terendah.

4.6 Pembahasan

Hasil evaluasi menunjukkan bahwa model BERT berhasil melakukan klasifikasi sentimen pada dataset dengan performa yang baik pada pembagian data

70:30 dan epoch ke-2. Akurasi model mencapai 83%, yang menunjukkan kinerja model dalam mengklasifikasikan data dengan baik.

Pada kelas 0 (Negatif), model menunjukkan performa yang baik. Dari 207 komentar yang sebenarnya negatif, model berhasil mengklasifikasikan 186 komentar dengan benar, memberikan presisi sebesar 87% dan recall sebesar 90%. Presisi sebesar 87% berarti bahwa 87% dari komentar yang diklasifikasikan sebagai negatif oleh model benar-benar merupakan komentar negatif. Namun, terdapat 22 komentar yang salah diklasifikasikan sebagai netral dan 6 komentar yang salah diklasifikasikan sebagai positif.

Pada kelas 1 (Positif), model berhasil mengklasifikasikan 33 komentar dengan benar dari 44 komentar yang sebenarnya positif, memberikan recall sebesar 75%. Meskipun presisi untuk kelas ini 79%, recall yang lebih rendah dibandingkan dengan kelas negatif menunjukkan bahwa model perlu ditingkatkan dalam mendeteksi komentar positif dengan lebih akurat.

Pada kelas 2 (Netral), model mampu mengklasifikasikan dengan benar 54 komentar dari 79 komentar yang sebenarnya netral, memberikan recall sebesar 68%. Meskipun presisi untuk kelas ini cukup tinggi 74%, recall yang lebih rendah menunjukkan bahwa model kesulitan dalam mendeteksi komentar netral dengan akurat.

Perbandingan hasil dari validator menunjukkan bahwa dari 330 data yang diuji, sebanyak 273 data memiliki hasil yang sama antara validator dan sistem klasifikasi, menghasilkan persentase sebesar 83%. Sementara itu, terdapat 57 data

yang hasilnya berbeda antara validator dan sistem klasifikasi, yang menunjukkan persentase 17%.

Secara keseluruhan, hasil yang dicapai menunjukkan bahwa model BERT yang digunakan dalam penelitian ini telah berhasil memahami dan mengklasifikasikan komentar dengan konteks sentimen yang berbeda secara efektif, serta dapat digunakan sebagai pendekatan yang baik untuk analisis sentimen dalam Bahasa Indonesia, khususnya pada topik ceramah Islam.

4.7 Integrasi Sains dan Islam

Integrasi antara sains dan Islam dalam penelitian ini tercermin dari pemanfaatan teknologi kecerdasan buatan, seperti model BERT, untuk memahami opini masyarakat terhadap ceramah Islam di YouTube. Hal ini tidak hanya mendukung perkembangan ilmu pengetahuan, tetapi juga memperkuat nilai dakwah Islam di era digital.

Dalam Tafsir Ibnu Katsir Edisi 1 hal 67, Allah Swt. berfirman dalam Al-Qur'an Surah *Al-Baqarah* ayat 14-15:

وَإِذَا لَقُوا الَّذِينَ آمَنُوا قَالُوا آمَنَّا وَإِذَا خَلَوْا إِلَىٰ شَيَاطِينِهِمْ قَالُوا إِنَّا مَعَكُمْ إِنَّمَا نَحْنُ مُسْتَهْزِئُونَ ۗ اللَّهُ يَسْتَهْزِئُ بِهِمْ وَيَمُدُّهُمْ فِي طُغْيَانِهِمْ يَعْمَهُونَ ﴿١٥﴾

“Apabila mereka berjumpa dengan orang yang beriman, mereka berkata, “Kami telah beriman.” Akan tetapi apabila mereka menyendiri dengan setan-setan (para pemimpin) mereka, mereka berkata, Sesungguhnya kami bersama kamu, kami hanya pengolok-olok. Allah Swt. akan membalas dan membiarkan mereka terombang-ambing (bingung) dalam kesesatan. (QS. Al-Baqarah: 14-15).

Ayat ini menggambarkan sikap orang-orang yang dengan sengaja memperolok-olok pihak lain, dengan tujuan untuk merendahkan mereka. Ketika

seseorang berpura-pura mendukung atau mengakui kebenaran di hadapan orang lain, tetapi secara diam-diam merendahkan atau menghina pihak tersebut di belakangnya, maka hal ini termasuk dalam bentuk kemunafikan yang dikecam dalam ayat ini.

Allah Swt. memperingatkan bahwa orang-orang yang memperolok-olok orang lain dengan cara seperti ini, baik itu dalam bentuk hinaan atau sarkasme, akan menerima balasan dari-Nya. Allah Swt. berfirman bahwa Dia akan membalas olok-olokan tersebut dan membiarkan orang tersebut terombang-ambing dalam kesesatan mereka. Artinya, orang yang merendahkan dan mengolok-olok tanpa dasar yang jelas atau dengan niat untuk merusak kehormatan pihak lain, pada akhirnya akan mengalami kebingungannya sendiri dan terhalang dari petunjuk-Nya. (Al-Sheikh, n.d.)

Ayat lain dijelaskan dalam Tafsir Ayat-Ayat Pendidikan hal 102, Allah Swt. berfirman dalam Surah *Al-Hujurat* (49) ayat 11:

يَا أَيُّهَا الَّذِينَ آمَنُوا لَا يَسْخَرْ قَوْمٌ مِّنْ قَوْمٍ عَسَىٰ أَن يَكُونُوا خَيْرًا مِّنْهُمْ وَلَا نِسَاءٌ مِّنْ نِّسَاءٍ عَسَىٰ
 أَن يَكُنَّ خَيْرًا مِّنْهُنَّ وَلَا تَلْمِزُوا أَنفُسَكُمْ وَلَا تَنَابَزُوا بِالْأَلْقَابِ بِئْسَ الْإِسْمُ الْفُسُوقُ بَعْدَ الْإِيمَانِ
 وَمَنْ لَّمْ يَتُبْ فَأُولَٰئِكَ هُمُ الظَّالِمُونَ

“Wahai orang-orang yang beriman! Janganlah suatu kaum mengolok-olok kaum yang lain, (karena) boleh jadi mereka (yang diolok-olok) lebih baik dari mereka; dan jangan pula perempuan-perempuan (mengolok-olok) perempuan lain, (karena) boleh jadi perempuan (yang diperolokkan) lebih baik. Janganlah kamu saling mencela satu sama lain dan janganlah saling memanggil dengan gelar-gelar buruk. Seburuk-buruk panggilan adalah (panggilan) yang buruk sesudah iman, dan barang siapa tidak bertobat, maka mereka itulah orang-orang yang zalim”
 (QS. *Al-Hujurat*: 11).

Ayat ini menegaskan pentingnya menjaga etika dalam komunikasi, termasuk dalam ranah digital. Kata yashkar/memperolok-olokkan dalam ayat ini, yaitu menyebut kekurangan pihak lain dengan tujuan menertawakan yang bersangkutan, baik dengan ucapan, perbuatan, atau tingkah laku. Komentar-komentar negatif atau yang merendahkan di media sosial bisa menciptakan suasana yang tidak kondusif bagi penyebaran dakwah. Teknologi seperti analisis sentimen dapat membantu menyaring konten yang bertentangan dengan nilai-nilai Islam (Dr. Listiawati, 2017).

Surah Al-Hujurat ayat 11 diturunkan terkait dengan beberapa peristiwa penghinaan terhadap sahabat Nabi yang miskin, seperti Ammar, Khabbab, dan Bilal, oleh delegasi Bani Tamim. Mujahid menyebutkan bahwa penghinaan ini terjadi antara kaum kaya dan miskin. Ibnu Zaid menambahkan bahwa orang yang dosanya ditutupi oleh Allah Swt. tidak berhak merendahkan orang lain yang dosanya terbuka, karena terbukanya dosa di dunia lebih baik daripada di akhirat.

Riwayat lainnya menyebutkan bahwa ayat ini diturunkan untuk mengingatkan agar tidak memanggil seseorang dengan gelar yang menyakitkan, seperti yang terjadi pada seorang laki-laki dari Bani Salamah yang tidak suka dipanggil dengan salah satu nama panggilan. Intinya, ayat ini mengajarkan umat Islam untuk menghindari penghinaan dan merendahkan sesama (Ayat, 2024).

Dari sisi hadis, Rasulullah ﷺ bersabda:

عَنْ عَبْدِ اللَّهِ بْنِ مَسْعُودٍ قَالَ: قَالَ رَسُولُ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ: سَبَابُ الْمُسْلِمِ فُسُوقٌ، وَقِتَالُهُ كُفْرٌ

Dari Abdullah bin Mas'ud RA berkata, "Rasulullah bersabda, 'Mencela seorang muslim merupakan kefasikan dan memeranginya merupakan kekufuran.'"

Dalam hadis tersebut, Rasulullah ﷺ menggunakan kata “*Sibabun*” yang berarti mencela, dan para ulama membedakan antara “*Sibabun*” dan “*Sabbun*” yang berarti mencela seseorang dengan aib yang memang ada pada dirinya, sementara berarti mencela seseorang tanpa mempedulikan apakah aib tersebut ada pada dirinya atau tidak. Inilah yang disebut sebagai kefasikan, yang lebih parah daripada sekadar kemaksiatan. Kefasikan sendiri dalam bahasa Arab berarti keluar dari kebenaran, dan dalam istilah syariat, itu lebih berat dibandingkan dengan perbuatan maksiat biasa.

Hadis ini menegaskan bahwa mencela atau mengolok-olok sesama muslim, baik dengan menyebutkan aib yang ada atau tidak ada, adalah perbuatan tercela dan kefasikan. Islam mengajarkan kita untuk menutupi aib saudara kita, bukan mengumbar atau menuduhnya dengan hal yang tidak ada. Rasulullah ﷺ juga mengingatkan pentingnya menjaga lisan, karena lisan yang tidak terjaga dapat menimbulkan perbuatan buruk, seperti mencela atau memerangi sesama muslim. Selain itu, Dalam buku Khotbah Jum’at Allah Swt. berfirman dalam Surah Al-Isra ayat 36:

وَلَا تَقْفُ مَا لَيْسَ لَكَ بِهِ عِلْمٌ إِنَّ السَّمْعَ وَالْبَصَرَ وَالْفُؤَادَ كُلُّ أُولَٰئِكَ كَانَ عَنْهُ مَسْئُولًا

“Dan janganlah kamu mengikuti sesuatu yang tidak kamu ketahui. Karena pendengaran, penglihatan dan hati nurani, semua itu akan diminta pertanggungjawabannya.” (QS. Al-Isra: 36).

Ayat ini menjelaskan mengenai tanggung jawab dalam menerima dan menyebarkan informasi (Alallah, 2024). Dalam konteks analisis sentimen komentar YouTube ceramah Islam, ayat ini mengingatkan kita bahwa setiap informasi yang

kita dengar atau baca, termasuk komentar-komentar yang beredar di platform media sosial, haruslah disaring dengan hati-hati. Sebagai pengguna media sosial dan masyarakat digital, kita harus mampu menyaring informasi yang valid dan menghindari mengikuti atau menyebarkan informasi yang tidak jelas kebenarannya.

Komentar ceramah Islam di YouTube sering memuat opini beragam yang dapat memengaruhi pemahaman dakwah. Karena itu, penting menyaring informasi yang benar dan bermanfaat, terutama dalam isu keagamaan. Penelitian ini menerapkan analisis sentimen berbasis teknologi untuk mengklasifikasikan komentar, agar pengguna lebih bijak dalam memilah informasi dan menghindari penyebaran kebencian.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini berhasil mengimplementasikan model BERT untuk menganalisis sentimen komentar pada video ceramah Islam di *platform* YouTube. Model BERT terbukti efektif dalam mengklasifikasikan sentimen komentar ke dalam kategori positif, negatif, dan netral, dengan hasil yang menunjukkan kinerja yang baik (Geografi et al., 2022). Evaluasi model mengonfirmasi bahwa BERT mampu memberikan hasil yang baik dalam mengidentifikasi sentimen komentar, meskipun terdapat beberapa kesalahan klasifikasi, terutama pada kategori sentimen netral.

Model BERT yang digunakan dalam penelitian ini melalui beberapa tahapan, mulai dari pengumpulan data komentar, *Preprocessing*, pelatihan model, hingga evaluasi menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-Score*. Berdasarkan hasil evaluasi, model mencapai akurasi tertinggi sebesar 83%, pada skenario 70:30 *epoch* 2 yang menunjukkan bahwa BERT dapat mengenali pola sentimen dalam komentar-komentar dengan baik. Meskipun demikian, ada potensi untuk perbaikan, khususnya dalam meningkatkan presisi pada kategori sentimen positif dan memperbaiki akurasi pada sentimen netral. Selain itu, perbandingan hasil dari validator juga menunjukkan bahwa 273 data dari 330 yang diuji memiliki hasil yang sama antara validator dan sistem klasifikasi, menghasilkan persentase sebesar 83%. Sementara itu, 57 data lainnya menunjukkan perbedaan antara

validator dan sistem klasifikasi, yang menunjukkan persentase 17%. Hasil ini menegaskan bahwa model mampu dengan baik melakukan klasifikasi sentimen, dengan sebagian besar data berhasil diprediksi dengan benar, meskipun ada sedikit data yang terklasifikasi berbeda.

Walaupun model BERT telah menunjukkan kemampuan yang baik dalam memahami sentimen komentar berbahasa Indonesia, tantangan dalam menganalisis komentar yang menggunakan bahasa informal, slang, atau singkatan tetap ada. Oleh karena itu, model ini perlu terus dikembangkan untuk lebih memahami konteks bahasa sehari-hari yang lebih variatif, terutama pada komentar-komentar yang berisi ekspresi tidak baku atau bahasa daerah.

Selain itu, integrasi antara teknologi kecerdasan buatan seperti analisis sentimen dengan nilai-nilai Islam memberikan kontribusi penting dalam pengembangan dakwah Islam di era digital. Dengan menggunakan teknologi ini, kita dapat memastikan bahwa pesan-pesan dakwah yang disampaikan melalui media sosial seperti YouTube diterima secara positif, sekaligus menghindari penyebaran komentar negatif yang tidak sesuai dengan etika komunikasi Islam.

5.2 Saran

Berdasarkan hasil penelitian ini, ada beberapa hal yang perlu dipertimbangkan untuk pengembangan lebih lanjut, baik dari segi model maupun aplikasi praktisnya. Berikut adalah beberapa saran yang dapat dijadikan pertimbangan:

- a. Pengembangan Model dengan Dataset Lebih Besar: Meskipun penelitian ini menggunakan dataset yang cukup besar, pengembangan lebih lanjut dapat

dilakukan dengan mengumpulkan dataset yang lebih beragam. Komentar-komentar dengan variasi bahasa yang lebih kompleks, slang, atau bahkan dialek daerah yang digunakan oleh masyarakat Indonesia dapat meningkatkan kemampuan model untuk memahami berbagai macam sentimen dengan lebih baik.

- b. Penerapan di Berbagai *platform* Sosial: Penelitian ini berfokus pada analisis sentimen di YouTube, namun model yang telah dibangun bisa diterapkan di *platform* media sosial lainnya, seperti Twitter, Facebook, atau Instagram. Dengan demikian, pemahaman mengenai opini publik terhadap isu-isu tertentu, baik yang bersifat sosial, keagamaan, maupun politik, dapat lebih luas dan mendalam. Selain itu, mengingat berkembangnya media sosial dalam berbagai bentuk dan bahasa, model ini dapat diperluas cakupannya ke tingkat global.

DAFTAR PUSTAKA

- Al-Sheikh, D. A. B. M. bin A. B. I. (n.d.). *Tafsir Ibnu Katsir QS-002 Al-Baqarah*.
- Alallah, M. (2024). *Khotbah Jumat*.
https://www.google.co.id/books/edition/Khotbah_Jumat/-ChAEQAAQBAJ?hl=id&gbpv=0
- Alfandi Safira, & Hasan, F. N. (2023). Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier. *ZONasi: Jurnal Sistem Informasi*, 5(1), 59–70. <https://doi.org/10.31849/zn.v5i1.12856>
- Andriani Marshanda Putria, Widya Khafa Nofab, D. A. P. H. (2025). *Penerapan metode bert untuk analisis sentimen ulasan pengguna aplikasi segari di google play store*. 4(1), 89–104.
- Areshey, A., & Mathkour, H. (2023). Transfer Learning for Sentiment Classification Using Bidirectional Encoder Representations from Transformers (BERT) Model. *Sensors*, 23(11). <https://doi.org/10.3390/s23115232>
- Arham, A., Swedia, E. R., Cahyanti, M., & Septian, M. R. D. (2022). Implementasi sentiment analysis pada opini masyarakat indonesia di twitter terhadap virus covid-19 varian omicron dengan algoritma naïve bayes, decision tree, dan support vector machine. *Sebatik*, 26(2), 565–572. <https://doi.org/10.46984/sebatik.v26i2.1961>
- Arifqi, T., Suarna, N., & Prihartono, W. (2024). Penggunaan Naive Bayes Dalam Menganalisis Sentimen Ulasan Aplikasi Mcdonald'S Di Indonesia. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1949–1956. <https://doi.org/10.36040/jati.v8i2.8740>
- Asiva Noor Rachmayani. (2015). *Konsep BERT pada NLP*. 6.
- Ayat, A. Q. S. A. (2024). *Rasisme di Media Sosial: Studi Tafsir Al-Munir Karangan Wahbah*. 9(2), 141–153.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-Training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1(Mlm), 4171–4186.
- Dr. Listiawati, M. H. I. (2017). *Tafsir Ayat-Ayat Pendidikan: Edisi 1*. Kencana. <https://books.google.co.id/books?id=H-VNDwAAQBAJ>
- Elankath, S. M., & Ramamirtham, S. (2023). Sentiment analysis of Malayalam tweets using bidirectional encoder representations from transformers: a study. *Indonesian Journal of Electrical Engineering and Computer Science*, 29(3), 1817. <https://doi.org/10.11591/ijeecs.v29.i3.pp1817-1826>

- Geografi, J., Ilmu, F., Unp, S., Lima, K., & Kota, P. (2022). *ANALISIS KERAWANAN LONGSOR MENGGUNAKAN BAHASA PEMROGRAMAN PYTHON DI KECAMATAN SULIKI KABUPATEN LIMA PULUH KOTA*. 4.
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/jti.v14i2.679>
- Harpizon, H. A. R., Kurniawan, R., Iwan Iskandar, Salambue, R., Budianita, E., & Syafria, F. (2022). Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes. *JNKTI (Jurnal Nasional Komputasi Dan Teknologi Informasi)*, 5(1), 131–140. <http://repository.uin-suska.ac.id/59746/>
- Iriyanto, D., Suarna, N., & Raditya Danar Dana, R. D. D. (2023). Peningkatan Mutu Pembelajaran Gerak Dasar Tari Kelas Vii Melalui Video Animasi 2D Menggunakan Metode Addie. *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, 11(1), 16–23. <https://doi.org/10.21063/jtif.2023.v11.1.16-23>
- Israel, B. P. (2024). *IMPLEMENTASI MODEL BERT PADA ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP AKSI*. 12(3), 2335–2342.
- Jayadianti, H., Kaswidjanti, W., Utomo, A. T., Saifullah, S., Dwiyanto, F. A., & Drezewski, R. (2022). Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN. *ILKOM Jurnal Ilmiah*, 14(3), 348–354. <https://doi.org/10.33096/ilkom.v14i3.1505.348-354>
- Kumari, A., & Arockiaraj, M. C. (2023). *YouTube Comment Scraping and Sentiment Analysis Using Python and Machine Learning*. 8(2), 13–17. <https://doi.org/10.46610/jowdwd.2023.v08i02.003>
- Lestari, N. (2024). Optimalisasi Dakwah Di Era Modern Melalui Media Sosial. *Qawwam: The Leader's Writing*, 5(1), 35–42.
- Mualfah, D., Ramadhoni, Gunawan, R., & Mulyadipa Suratno, D. (2023). Analisis Sentimen Komentar YouTube TvOne Tentang Ustadz Abdul Somad Dideportasi Dari Singapura Menggunakan Algoritma SVM. *Jurnal Fasilkom*, 13(01), 72–80. <https://doi.org/10.37859/jf.v13i01.4920>
- Putra, S. A., & Gata, W. (2020). *Analisis Sentimen Komentar Youtube Terhadap Ceramah Ning Umi Laila Sindir Rhoma Irama Menggunakan Algoritma LSTM*. 0877.
- Rani, S. (2023). Transformasi Komunikasi Dakwah dalam Era Digital: Peluang dan Tantangan dalam Pendidikan Islam Kontemporer. *AL-MIKRAJ Jurnal Studi Islam Dan Humaniora (E-ISSN 2745-4584)*, 4(1), 207–216. <https://doi.org/10.37680/almikraj.v4i1.3513>
- Singgalen, Y. A. (2022). Analisis Performa Algoritma NBC, DT, SVM dalam

Klasifikasi Data Ulasan Pengunjung Candi Borobudur Berbasis CRISP-DM. *Building of Informatics, Technology and Science (BITS)*, 4(3). <https://doi.org/10.47065/bits.v4i3.2766>

Situmorang, G. F., & Purba, R. (2024). *Deteksi Potensi Depresi dari Unggahan Media Sosial X Menggunakan Teknik NLP dan Model IndoBERT*. 6(2), 649–661. <https://doi.org/10.47065/bits.v6i2.5496>

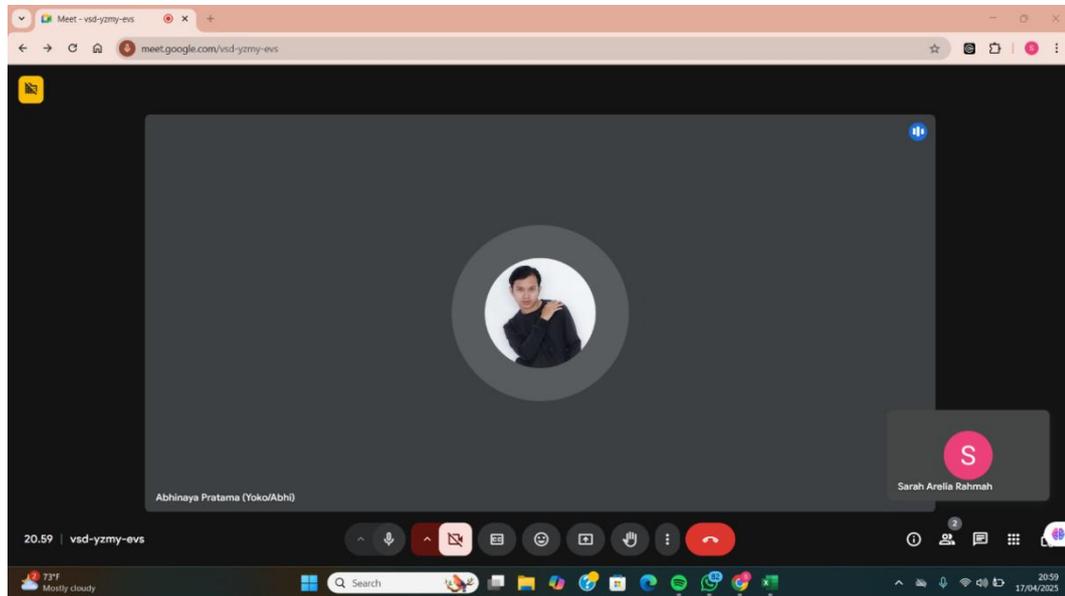
Suharsono, J. P., & Nurahman, D. (2024). Pemanfaatan Youtube Sebagai Media Peningkatan Pelayanan Dan Informasi. *Ganaya*, 7(1), 298–304. <https://doi.org/10.37329/ganaya.v7i1.3157>

Vidya Chandradev, I Made Agus Dwi Suarjaya, & I Putu Agung Bayupati. (2023). Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT. *Jurnal Buana Informatika*, 14(02), 107–116. <https://doi.org/10.24002/jbi.v14i02.7244>

LAMPIRAN

LAMPIRAN 1

Meet bersama Validator



LAMPIRAN 2

Data hasil labeling Validator dan Prediksi terbaik dari Model

Komentar	Label_Manual	Predicted_Label
jangan bahas lain dulu mending tes dna dulu lah	2	2
lanjutkan bib jga kesehatan mu jga istikomahan mu semoga perjuangan mu berhasil habibana	1	1
satu tambah satu berapa bang	2	2
lebih parah lu atas dasar nasab palsu ngemis rakyat	0	0
orang ceramah gak sih bisanya gibah melulu	0	0
tanah ane minjem ane tawar jt baru nawar aja uang nya kalo disini ada org per org rb aja besok ane bayar haji taupik mau ngasi berapa jt takbir	2	2
maju terus habib bahar suaramu menyambung suara rakyat lah isi hati rakyatsemoga panjang umur habib bahar lindungi allah lancar dakwah nyasalam bangsa aceh darusallam	1	1
panas kipas	2	0
ceramah apa gibah sih	0	0
min minta video buat reupload ada buat beli roko jarum coklat	2	2
mondok pesantren biar lebih paham agama	0	0

Komentar	Label_Manual	Predicted_Label
bahar bin smith mahluk bisnis agama sungguh sangat miris orang respect sama si bahar	0	0
beuh indonesia sangat berisik krn bahar jeng riziq hmm tes dna dulu deh jangan ceramah lain dulu non sen ada keberkahan	0	0
semoga lekas sembuh kasian rakyat	2	1
selau bahasanya ada urusan	0	0
subhannal masya allah tabarakallah semoga ustadz habib bahar binsmit dipanjangkan umurnya allah swt diberi kesehatan slalu dilapangkan rezekinya amin	1	1
sekalian basmi org belokan sejarah	0	0
lebih rusak kamu hartipu mukibin pakai nama nabi	0	0
mohon maaf dari pihak prsj pengurus rumah sakit jiwa mengakatan anda rawat mohon sembuhkan dulu prsj anda komen di sini	0	2
kebenaran gimana nasabnya gak jelas	0	2
isi ceramah nya judul nya menggibah	0	0
rasainilh buah arab sebenarnya terkutuk	0	0
habib sarinah ayam siksa bpk gw sakit sangat perih paru hukum mrk lap pritik spritus ku isik lap mesi sabu sabun me tegeteh tikus ku bengutk bapak tikus toloktaruh makanan tikus ampok ne langsing bapak coba teriak dijawa bib tegaltak lawan sampean bib	0	0
manusia jadi jadian akhlak nya bobrok	0	0
kun fayakun silahkan pesan kafani bang silahkan check out	2	2
tumben pro presiden sekarang	0	2
loe paham gak arti nahi munkar kaya gimana	2	0
aku suka ceramah habib bahar bin smith hrs tegas membela rakyat kecil pda ceramah lemah lembut bikin ngantuk lemah lembut tentu baik hatinya orang omongannya keras hatinya lembut prinsipku ambil baiknya buang buruknya	1	1
bejad mata duitan wkwkwkw	0	0
alhamdulillah	1	1
allah hu akbar smoga sehat selalu alhabib	1	1
lu cuap doang tong bukti tunjukin bukti	0	0
ceramahnya bahar ky kornet bis	0	0
bahar bin dedemit bin memedi	0	0
wkwkw bahar	2	2

Komentar	Label_Manual	Predicted_Label
bahar smith keturunan komunis pki yaman	0	0
bahar bangsat tenan ikijancuk bahart	0	0
mantap	1	1
beda sama elo trik pinter kalau kamu bodoh terik bilang lo gila	2	0
mediaserbagunamukibul takdir	2	2
haduh tes dna hela baru tablig keilmuan lain	2	0
terimah kasih buat si demit berkat mu ulama berkedok habib putus kali engkau para cecunguk pda kelabakan tantrum	0	0
emang kamu berani sama pak prabowo ngaca bang	0	2
godbles you wal habib bahar bin smith	1	1
manusia kyk gni ko di yang mau mendengarkan hadeuhh manfaat nya apa isi ceramahnya	0	0
kelamaan	2	2
soak mantan cucu nabi	0	0
rusak nya rakyat karna lu bahar samit	0	0
hee bahar iraha sadar na maneh aing ngadoakeun mun sia juriah rosul sing teereh sadar mun sia teu nyambung sbagai juriah rosul sing hina keun we ka hirupan sia ku jamaah sia cam kan	0	0
enapa klu persiden habib rijik wakil bahar bin shemitcobah jadi pemerita bln aja gemana kaya rskyat satuan janda pakir	2	0
nasab palsu byk bacot	0	0
cinta monyet	2	0
si bakar muja prabowo takut iya	2	2
orang gila	0	0
mau muntah aku mendengar ocehan mu jahanam	0	0
memet aku hamil	2	2
ceramah isinya apa	0	0
wong iki maneh muak	0	0
wadaw	2	2
dulu gua mukhibin garis kerastapi semenjak lihat si bahar teriakquot jual airgua mulai sadar lebih cinta ulama kyai nusantara	0	0
semua di omongin kaumnya	2	0

Komentar	Label_Manual	Predicted_Label
hidup habib bahar	1	1
emangnya kamu manusia apa	2	2
sangat bersyukur ada habib bahar	1	1
muak orang	0	0
emalmahameruini tanah ane minjem ane tawar jt baru nawar aja uang nya kalo disini ada org per org rb aja besok ane bayar haji taupik mau ngasi berapa jt takbir	2	2
masyaallah sehat selalu bib	1	1
kok ceramah ini yang dengerin udah stres kali jemaah panitianya orang tukang caci maki kok undang	0	0
masa ceramah jadi politisi	0	0
ambil bang makasi bang bisyaroh nya berkah	1	1
maksa aja panitianya mudahan jamaah nya dengerinnya masuk kuping kanan keluarin jangan dikeluarin kuping kiri	2	0
dasar bujur banar bib ai	2	2
habib bahar semua teguran biar pemerintah jadi baik insya allah menjaga nkrikalau banyak baik mungkin santay aja makasih habib	1	1
mending sumbangin buat kamu ajatrus tarok kamar nya	0	0
mudah mudahan aja si bahar kena azabngeritik pemerintah segala emang si bahar udah berbuwat apa negara	0	0
izin reupload min buat ngopi terima kasih	2	2
ada harus bingungin	2	2
betul kata habib memang yang alami	1	1
sekalian sama kamu jg tangkap jgpemecah belah bangsa	0	0
ceramah hebat kh zainuddin mz nyinggung pemerintah tidak brutal	2	0
lawak lu ah	2	0
lalu sambo bagmna	2	2
nkri kalo ada habib makmurkrna ada pengemis paksa	0	0
hadeh gajadi ah mau ngundang	0	0
bola panas keluar mulut bahar menghantam ke dirinya bahar lihat segera kun fayakun	0	0
gak jelas isi ceramah nya apa	0	2
ceramahnya memprovokasi terus	0	0
ceramah habib bahar bin smith isinya daging semua	1	0

Komentar	Label_Manual	Predicted_Label
mkhoerudilahn lah kalo mulut nya kotor buat apa mulut lidah lebih rusah para koruptor	0	0
ulama kan beda adabyf keras yang lembut	2	1
manusia perusakk streesz smitty	0	0
maskapan ngaji nyakitab baca bukan nya cuma koar koar doankkebanyakan ceramah ngaji nolpalsu palsu	0	0
es campur nya bang	2	2
hu antum udah beli pengaman gak nanti malemm	0	2
modal teriak laku ya	0	0
ceramah ko ngomel marah	0	0
semoga habib tmbah sehat penuh keberkahan aamiin	1	1
ceramah nya selalu marah aneh dech	0	0
iya tunggu tanggal mainnya lu bahar tangkap	0	0
takbir	1	1
semoga panjang umur kyai imad	1	1
orang manggil si bahar mah sama stresnya	0	0
ceramah provokasi	0	0
jelas loe bib	0	0
heran cuman modal teriak teriak ko laku jadi penceramah	0	0
sebenarnya bahar ikut andil menyengsarakan muhibbin sumbangan donasinyadan bahar menyamar jadi habib jadi jadian tujuanya untuk perut dunia buktinya bahar sering pamer makan minum layani jongosdan bahar sering flexing kemewahan rumahnya koleksi mobil mobilnyajadi ceramah bahar justru menyindir diri nya sendiridasar koplak	0	0
esihkurniasih betul sekali bu pilihlah ulama dibenci rezim zolim yang bisa dibungkam uang	1	2
habib habib gila isinya adu domba	0	0
teriak keruan bermutu bahar	0	0
kata sapa wkwkwk	2	2
kemarin bahar bilang prabowo penghianat	2	2
sombong kamu bahar	0	0
kamu rusak ngaku anak nabi	0	0
pada pemakan kotak amal pemakan uang sumbangan	2	0
ayoo teriak prabowo penghianattakut	2	2

Komentar	Label_Manual	Predicted_Label
tumben loh bahar binsemit bicaranya benar	0	0
saya bodohitu bukan ceramahitu orasi demo	0	0
kayanya kejadian nasab baalawigak bakalan berkoar koar	0	0
lebih rusak lo ngaku cucu nabib	0	0
dakwah apa ngurusin pemerintah kalau mau dakwah bicara lah agama pemerintah ada ngurus	0	0
bahrulkamal namanya lihat negara baikquot siapa tidak marahmafia judi dilindungiyg ditangkap malah krocoquot tidak pernah tuntas kasus apapunkorupsi terangquotanpadahal dampak kedua kejahatan sistematisdgn judul l bisa malingutang pinjolkeluarga berantakanbahkan polisi bakar suamibegitu korupsi terus dgn kayak gitu gak marahsitu waras kan anda bilang ceramah kok triakquot tidak lembutyg jelas esensi ceramahnya habib bahar bin smith amar maruf nahi munkar itu ajaran islam	1	0
ulama muda berani tegas cerdas sulit mencari ulama beliau	1	1
orang segera disingkirkan	0	0
bukan maling anda merusak agama cuma ngomong tanpa bukti	0	0
manusia apa binatangkog persis binatang yanajis lihatnya	0	0
ceramah tidak perlu Teriak Teriak ceramah santai	2	0
ceramahnya bermutu	0	1
heran ko ada yang ngasih panggung	0	0
ngotot jadi cucu nabi quot bantah data sejarah tes dna manuskrip filologi pustaka naqobah ansab quot	2	2
namanya pengadu domba ulama anggap dekat pemerintahan dari jaman di setiap pemerintahan pun kerajaan ada ulama kiyai penasehat	0	0
hidup kan petrus bungkus tikus	0	0
aja yang mau mendengarkan ocehan gelandangan penjual ayat tuhan	0	0
sir wala taqif	1	2
org gila ky gini di dengerin	0	0
ngapain si wooy ngapain Teriak triaklu pikir orang budeg apa	0	0
ceramah politik mulu pidato apa pengajian	0	0
bukan ceramahtp orasi politik	0	0
orang berceramah punya ciri khas masing cara penyampainya itu dijalur kebenaran bangsa inisemoga	1	1

Komentar	Label_Manual	Predicted_Label
habib diberikn kesehatan panjang umur berkah doanya dunia akhirat		
bahar ada sopan santun wagil pak prabowo prabowo ae betar tangkap habib palsu yang di kandangin	0	0
basmi baalwipenipu nasab pemalsu kuburantukang syubhattahayul	0	0
sang pembela kebenaran	1	2
kamu omongan sia bukan ceramah agama tp politik	0	0
ceramahnya kalah jauh sama uah	0	0
ujisimatupanglogin sini	2	2
ceramah isinya sampah	0	0
ulama lurus tegas apa adanya	1	1
habib cucu nabi hahahaha mimpi ato menghayal gak cocok jadi kenyataan	0	0
kok capek	2	0
ceramah kok teriak materi disampaikan gak jelas	0	0
customerru bahar gak benar kib skrg udah sakit kejiwaannya punya gen israel	0	0
katanya bahar udah muak indonesia ko ceramah indonesia bikin rusak gerombolan bahar demit ente kkg usah ikut campur urusan negara kpk lebih tau mana koruptor	0	0
ambil hikmah nya aja bagus dengerin jlek buang aja simpel	2	0
jangan tangkap buang aja negara timur tengah	0	0
kaya paling suci si bahar	0	0
bahar smith benar benar dahini orang emang di ruqyahnggak cuma pake ayat suci al qur antambah kembang rupa plus air sumur biar liar omonganya	0	0
namanya zuriyah palsu kyak gini	0	0
kibin	2	2
bahar kalah main kelereng makanya ngamuk	0	0
di kasih panggung aja orang bgtuu	0	0
bahar bin bakwali bin yaman bin palsu bin wadul kaleng rombeng	0	0
nih gurunya	2	2
ujung ujung nya duit minta sumbangan	0	0
sungguh mengherankan makhluk jenis bebas berkeliaran mana marwah pemimpin hapus spesies pirang	0	0
wkwkwk babu nya bahar	0	0

Komentar	Label_Manual	Predicted_Label
parah bahar pirang cucung antek belandapak presiden prabowo takangkap penipu besrorban bahar pirang	0	0
panitia ngundang gelayg undang lebih geloornng sperti injgnkasih panggung	0	0
gak takut pegat apah yahh uratnya ngomongnya kenceng kenceng gitu	0	0
nah kelebihanannya yik bahar biar jin ada badannya lari	2	0
bahar ceramah eggah majunya	0	0
sadar wahai panitia penyelemgar jangan lah mengundang penceramah sll memecah belah umat	0	1
mengundang yang undang samastresnya	0	0
i love u habib	1	1
usah perintah bahar pal prabowo berantas korupsi bahar fitnah pak jokowi pencemaran nama baik pak jokowie wahai projo wahai pendukungnya pak jokowi segera laporkan bahar b smith atas propokssi fitnah nya pak jokowi	0	0
iya si eta mah sok gak puguh puguh ceuceuleuweungan olangan	0	2
lu gak punya iman	0	0
bukan gak jelas ceramah beliau otak lo aja gak jelas menangkap ilmu beliau sampaikan beliau menyampaikan kebenaran m negakan keadilan memberantas koruptor	1	1
ente becanda model gini diblg ulama tai	0	0
bahar nin tompel	0	0
ngomong opo to koe ki bib ra jeas arah omongan mu ki susah pahami orang uang goblok kaya	0	0
yaa allah senang dakwah begini orang aja indonesia begini masallah	1	0
munafikun jangan bawa mukibin indonesia karewna saudara hayuu lah bakaalawi murni jangan bawa mukibin lawan senggol hayang perang tah salam banten	0	0
ceramah apa ngajak berantem senang ustad pribumi bicaranya halus mudah mengerti nakutin gitu	0	0
gak bahaya ta	2	2
gua bingung orag ada ngikutin	0	0
kalo msh ceramah misuhmenghina hinaudhlah tamat	0	0
habib pedut bermanfaat	0	1
efek minum pil koplo butir ceramah	0	0
rakyat susah pemasukkan kurang	2	0
takut bib	0	2

Komentar	Label_Manual	Predicted_Label
dia si bahar demit dakwahnya selalu provokator rakyat pemerintah	0	0
mulut mu yang rusak	0	0
tahu bulat	2	2
kallau mau berjuang berjuanglah palestinedisana jelas butuh peerjuangan	2	2
orang sakit jiwa gini kok ada ndengerin	0	0
kalian stress	0	0
habib palsu	0	0
sang komentar gila hadir nih	2	2
minta video min syiar	2	2
bahar tae kering	0	0
siap terima kasih	1	1
mungkin suka judi online	0	2
hey hey hey ini gak urusan	2	2
pelawak dunia	0	2
muhammad sifat heii orang sllu menebarkan kasih sayang muhammad bukan menebar kebencian	1	0
koyok panitia sorga ae mitdemit	0	0
pengikut elopun har bahar banyak jg main slot urus dl pengikut lo br urusin lain kalo emng benar apa lo tuduhkan mereka kau sebut korupsi apapun laporkan jangan bacot kasi jg buktinya kalo cmn teriak bayi thnpun lawak lawak	0	0
tau hadis di kitab durotunnasohin	2	2
ngaji ngapa iki telake ora metu	0	0
orang apa nya berpangku tangan dibelakang	2	0
rusaknya ulama bukan pemimpin pemerintahan tetapi oknum baklawi sendiri bertujuan membelokan sejarah bangsa	0	0
usir si bahar nkri	0	0
fitnah paling bisa maafkan menuduh pak jokowi korupsi pak jokowi paling hati hatu apbn minta maaf lah kaalian wahai rakyatku pak jokowi kalian terus memfitnah pak jokowi niscaya kalian di bersihkan bumi nusantara	0	0
setan bin demit	0	0
gak pegel apa ngomong growokan mulu ceramah tuh yng kalem apa har har	0	0

Komentar	Label_Manual	Predicted_Label
kamu komen nya jangan gitu coba sekali kali komen tahu bulat goreng mobil lima ratusan enak	2	2
kau percaya yakin dalam tubuhnya habib bahar bin smith mengalir darah baginda rasulullah saw kau ikut cinta sama habib bahar bin smith biar masuk surga biar syafaat baginda rasulullah saw jangan kau menghina kau kewalat masuk neraka	1	0
hahahasi jago provokator kluar urat matanya kaya dah mau ksurupan	0	0
katanya habib kalo mau tes dna yang paling depan tp nyatanya mana omonquot aja demit	0	2
kamu siapa mengaturquot org indonesia hay har bahar kamu emg pintar ngomong atas nama nabi muhammad tp kau sndiri bejat	0	0
terus kamu nya siapa dong kalo pemberontak	2	2
goblok	0	0
amit amit	0	1
kalo cucu nabi bacot nya jangan kaya gituh dawir	0	0
ceramah isinya ngurusin pemritah ilmunya di gaji	0	0
aparat daerah kepolian langsung mengamankan provokasi kayak gini wong bahar mbok pateni wae lo aku ra piye rakyat malah buahagia	0	0
sia bahar hayang jadi presiden sarua we sia ge no hiji mah perut bohong sia rizek mawa muhamad	0	0
mantaptegas lurus takbir	1	1
boroknya ketahuan busuk dulu mah dielu elu disanjungquotcie	0	0
sahadatnya benar	0	1
preman pasar cari muka	0	0
lanjutkan perjuangan guruku	1	1
bisanya koar koar tok mana perjuangan bahar cocote tok	0	0
kamu aja sono cramah sambih nyanyi gendong mana	0	0
ceramah ape mau berantem bahar smith	0	0
lanjut bib	1	1
betul bibgk ulama berani mengkritik pemerintah dzolimselain habib bahar meluruskan umat memang tugas nya ulama	0	1
kata siapa nabi muhammad diam ke zoliman kemungkaran kalo tau sejarah islam diam oke hbib bahar jalan benar sehat selalu bib terus bib berjuang di jalan allah	1	1
si bahar nya hasih air kencing noh biar suaranya pelan dikit	0	0

Komentar	Label_Manual	Predicted_Label
bahar jangan memecah belah rakyat	0	0
ceramah yh bukan ngelurin kata kat pantas mlh ngomongin polotik menjelek jelekin orang parah pemerintah jangan diam orang satu	0	0
dulu saya sampaikan hatihati ceramah seperti hatiquote tuturkatanya tidak lembut saya mohon jangan terdoktrin wahay sodarasodarku habib negeri ini kacau mantap bib	0	0
	1	1
serak gerong sia bahar munafik	0	0
mantap	1	1
si bahar kalo ceramah begini mulu	0	0
amiinallohuma solli alla muhammad	1	2
kalau jaman pa hartu udah psti ilang tuh ngebacot se enak nya dewek msa keturunan nabi ceramah nya	0	0
ceramah apaujuk rasaatau jual obatparah bnr orang	0	0
bahar semit memang stres	0	0
betul yang rumasa jadi suka habib bahar	0	0
lo kamu muhibinnya si bahar bin demit	0	0
ada tema selain membicarakan pemerintah	0	2
ceramahnya anak kecil	0	0
nonton bingung apa dbahas kok malah membuat orang bingung alasan bginda nabi isinya amburadul	0	0
bahar tukang obor mungkin hancur diri sendiri penuh cacian memalukan	0	0
cucu ubed lg bersiul	2	0
saya heran kok orang kaya gini pengikutnya banyak	0	0
allah jagalah selalu habib bahar bin smith	1	1
knalpot resing	2	0
tanah ane minjem ane tawar jt baru nawar aja uang nya kalo disini ada org per org rb aja besok ane bayar haji taupik mau ngasi berapa jt takbir	2	2
ceramah orasiini bukan keturunan rasulullah karna keturunan rasulullah pernah mengakui mengetahui dirinya keturunan nabi quothadist shahihquote	0	0
inlah orang paling rusak merusak umat islamsemoga dijauhkan mahluk ini	0	0
rusaknya ahlak warga indonesia rusak si bahar	0	0
sama stres sama ngundang	0	0
setuju buang aja nusa kambangan	0	0

Komentar	Label_Manual	Predicted_Label
ampun mberek bae ceramahe kt mah watir rongga tenggorokane pecah bae cung kacang	0	0
bukan begini cara orang ceramah bib	0	0
ythimigran yaman yahudi jago baca selawat baalawisuruh baca kitab imigran yaman yahudi pade nangis	0	0
tangerang kota hadir bib	2	2
betul betul	2	1
hp ku ipone hp mu apa	2	2
masing masing karakter nya ceramah	2	2
habib gila	0	0
uhuyy si paling benar takut ambeien kambuh ngeden terus	2	2
masih orang di indonesia guys	0	2
politik	2	2
woy ceramah ceramah bukan memperopokasi rakuat memecah belah doangkalo jangan undang tuh si bahar menulis ketemu nya imam ghozali dgn nabi musa siapa	2	2
semoga allahh melindungi ulama ulama lurus prabowo bravo jangan mau kendalikan remot remot nya amanah amanah lillah krn allahh swt aamiin krn umur pemimpin kita tau pakk	1	1
mie instan harga nya berapa bang	2	2
woii orang bejad apanya habib bahar bin smith bejad tuh orang pribumi pengurus ponpes melecehkan santrinya menghamili santriwatinya yang namanya manusia bejad jahannan si penipu donatur kaya pikir dong sedikit salam waras	0	0
betul yang rumasa jadi suka habib bahar	0	0
ada bagian nya hukum ulama ngajarin baikan biar damai negeri	1	0
habib apa preman lu	0	0
mukibul takdir	2	2
yang begini msh yang manggil ceramah	0	0
minum arak dulu trs ceramahitu lah bahar semit	0	0
polri zaman sekarang suap diam koruptor tahanan malah kasi enak	0	0
bilang aja kamu takut ama prabowo	2	2
sore cod	2	2
kok ada ngikutin orang ini	0	0

Komentar	Label_Manual	Predicted_Label
jamaah nya bahar smit kok mau dengerin bingung	0	0
orang gila ko suruh ceramah	0	0
jadi presiden ajah lu coba bahas bis dedemit	0	0
samoempud percaya tidak bukan urusan loh bro	2	2
habib palsu koar	0	0
tanah ane minjem ane tawar jt baru nawar aja uang nya kalo disini ada org per org rb aja besok ane bayar haji taupik mau ngasi berapa jt takbir	2	2
santai aja haranda bukan habib di hormati	0	0
pujiayulestarixzmakanya lu ngaji monyet biar jelass	0	0
luar biasa ilmu hb ceramah penuh kepedulian kpd rakyat	1	1
ceramah taek	0	0
sama presiden ada hormat nya manggil asal nama aja	0	2
bahar sikat juabahar ne ap x garang ceramah ky tahi palat	0	0
karepmu har	2	0
kenape lu sewot kepanasan lu	0	0
benar sih di tangkap koruptor ecek ecek kls teri raja korupsi jokowi keluarga nya coba berani menangkap jokowi prabowo mah udah kebaca lah	2	2
lanjutkan bib berjuangjadilah habib lurus menjadi ulama istiqomah	1	1
ceramah nya cuma ngehujat pemerintah aja tidak ngerti keturunan nabi ko gitu	0	0
pengajian dikaji apa ilmu agamanya mana	0	0
tauasiah apamau sidang koruptor	0	2
kamu gak takut tangan nya patah yaa kok ngetik nya sampek gitu	2	2
hey jowo tadi mau ikut mandi kok ajak sih wis selingkuh sama si tiya tuh jowo gak mau sih wis putus	2	0
kaya benar aja idup pencangkok nasab penipu makam	0	0
tes sound chek satu slamet audio rimex jeng jeng piw piw jeng dar	2	2
golput bertanggung jawab atas negara	2	2
kecap hitam lada micin kilo nasi kilok cabai garam secukupnya masukin semuanya	2	2
dibahas cuma pemerintahan doank giliran suruh baca kitab kuning bebabeba doank gak materi bahasannyayakin kalo orang yang boncengin politik ente ngikutin kalo ente mati emang masuk surga karna	0	0

Komentar	Label_Manual	Predicted_Label
ngikutin belum sadar woi nabi gak pernah ngajarin ngomong kurang ajar inimah bawaan yaman		
habis minum anggur kyk nya ngomong nya sampek ngeden mau eek bib	0	0
nderek tabarukan majlis maulidilmuampara kyaimasaikhhabaib	1	0
we alhamdulillah anda komen astagfirullah saya sukanya omaygaat maydarling supeng hiyo iyee hak ah iyeh	2	2
bahar stres punya iman jarang sholat so jago so suci aslinya bejat pengikutnya tau semua kelakuan bahar bahlul	0	0
satu nya berapa bang tambah tahu bulat nya empat sendok campur otak nya mangkok sama apa bang	2	2
yang disampingnya bising apa	0	0
halah apa mantaf cuman modal teriak teriak	0	0
waria	2	0
ceramahbohong mah	0	0
penjarakan bahar telah memprovokasi seluruh rakyat indonesia har bahar tolol	0	0
fikrimuhammad susah emang kalo udah kedoktrin mah percuma sekolah tinggi cuman jadi budak yaman	0	0

Keterangan :

Negatif : 0 Positif : 1 Netral : 2