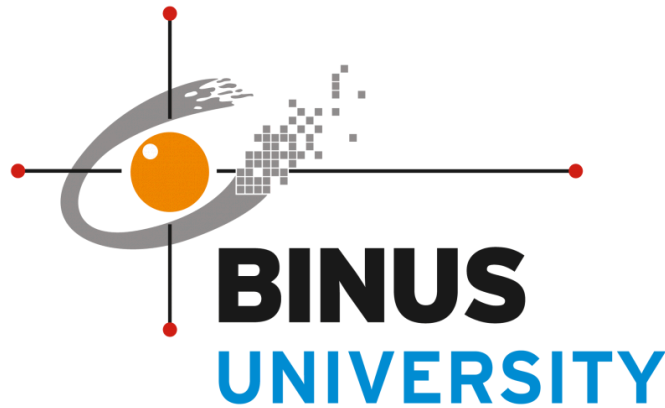


LAPORAN PROJECT TEXT MINING



Youtube Comment Sentiment Analysis and Summarization from 'Coursera Review' Videos

Felix Cristiano Bungaran - 2602170612

Kayla Masayuningtyas - 2602141871

Meisa Kamilia - 2602135446

BINUS UNIVERSITY

2024

DAFTAR ISI

BAB 1 PENDAHULUAN.....	3
1.1 Latar Belakang.....	3
1.2 Rumusan Masalah.....	5
1.3 Tujuan dan Manfaat.....	5
1.3.1 Tujuan.....	5
1.3.2 Manfaat.....	5
1.4 Metode.....	6
BAB 2 TINJAUAN PUSTAKA.....	7
2.1 Analisis sentimen.....	7
2.2 Massive Open Online Course.....	7
2.3 BERT.....	7
2.4 ROBERTA.....	8
BAB 3 METODE IMPLEMENTASI.....	9
3.1 Kerangka berpikir.....	9
3.2. Metode riset.....	10
3.3 Pengumpulan data.....	10
3.4 Pra Proses data.....	10
3.5 Modelling.....	10
3.6 Evaluasi model.....	11
BAB 4 HASIL DAN EVALUASI.....	12
4.1 Analisis Sentimen.....	12
4.2 Text Summarization.....	12
4.3 Evaluasi.....	13
BAB 5 KESIMPULAN.....	14
BAB 6 DAFTAR PUSTAKA.....	15

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Feedback pengguna yang berlebihan muncul akibat semakin populernya *Massive Open Online Courses* (MOOC), karena MOOC telah berkembang dari 300.000 pengguna pada tahun 2011 menjadi 220 juta pada tahun 2021 (McKinsey, 2022), sehingga menjadi tantangan bagi platform seperti Coursera untuk memindai dan menganalisis *review* secara manual secara efisien. Coursera adalah salah satu platform digital terkemuka yang menawarkan berbagai kursus dan sertifikasi secara global (Coursera, 2024). Seiring dengan semakin populernya platform ini, memahami pengalaman pengguna melalui ulasan menjadi penting untuk meningkatkan kualitas kursus dan kepuasan siswa. Analisis sentimen memainkan peran penting dalam proses ini dengan mengevaluasi dan mengklasifikasikan *Feedback* pengguna secara otomatis, sehingga memberikan wawasan yang berharga.

Penelitian terkini telah menekankan perlunya analisis sentimen dalam mengelola volume data yang sangat besar ini. Platform kini dapat secara otomatis mengkategorikan masukan pengguna dan memperoleh pemahaman yang lebih mendalam tentang pengalaman pengguna dengan bantuan teknologi ini. Misalnya, analisis sentimen terbukti meningkatkan pemahaman tentang pengalaman pengguna dengan memungkinkan platform menemukan dan memperbaiki masalah dalam skala besar (Tan et al. 2020). Data untuk analisis sentimen yang kami gunakan berisi informasi mendalam tentang kesenangan pengguna, kualitas kursus, dan efektivitas instruktur, tetapi sulit untuk menarik kesimpulan yang berguna dan pola yang signifikan dari teks tak terstruktur dalam jumlah besar tanpa menggunakan teknologi otomatis seperti analisis sentimen.

Dalam studi Lazrig dan Humpherys's (2022), algoritma *machine learning* seperti Naive Bayes dan Random Forest digunakan untuk melakukan analisis sentimen pada ulasan siswa. Studi tersebut menunjukkan bahwa algoritma ini dapat mengklasifikasikan sentimen secara akurat dengan tingkat keberhasilan yang mencengangkan, yakni 98% kasus. Namun, metodologi tersebut terutama berfokus pada pembedaan antara umpan balik positif dan negatif, dengan mengabaikan pemikiran netral. Dengan menggunakan metode ini, algoritma tersebut mampu mengklasifikasikan sentimen dengan tingkat presisi yang tinggi, yang menunjukkan potensi pembelajaran mesin untuk menilai komentar pengguna secara andal dalam konteks pendidikan.

Dalam sebuah studi tentang analisis sentimen untuk *Massive Open Online Courses* (MOOCs), Mrhar et al. (2021) menggabungkan penggunaan teknik Bayesian dengan *Convolutional Neural Networks* (CNN) dan jaringan *Long Short-Term Memory* (LSTM). Menurut penelitian mereka, akurasi model gabungan ini lebih tinggi daripada model CNN atau LSTM secara individual. Menurut Mrhar et al.

(2021), teknik hibrida menunjukkan kemanjuran yang luar biasa dalam mengurangi ketidakpastian prediksi, menghasilkan estimasi probabilitas yang lebih dapat diandalkan untuk kategorisasi sentimen, dan mengelola umpan balik pengguna yang rumit atau ambigu secara efektif.

Chen et al. (2022) mengintegrasikan pendekatan *deep learning* Bayesian dengan arsitektur *Convolutional Neural Networks* (CNN) dan *Long Short-Term Memory* (LSTM) untuk menyajikan solusi bagi analisis sentimen dalam MOOCs. Sasaran dari strategi ini adalah untuk mengatasi masalah prediksi ketidakpastian sentimen, yang sangat penting untuk mengidentifikasi siswa yang mungkin berisiko putus sekolah. Program ini dapat mengidentifikasi reaksi emosional siswa dalam posting forum dengan lebih tepat dengan mengkarakterisasi ketidakpastian ini secara tepat. Hal ini memungkinkan untuk menyediakan modifikasi dan intervensi kursus yang lebih individual untuk meningkatkan keterlibatan dan retensi siswa.

Dalam studi ini, kami melakukan analisis sentimen dan *text summarization* pada evaluasi ulasan Coursera menggunakan BERT (*Bidirectional Encoder Representations from Transformers*). BERT memiliki kapasitas canggih untuk mengumpulkan konteks dari kedua arah dalam sebuah kalimat, BERT digunakan dalam studi ini untuk analisis sentimen pada ulasan Coursera. BERT dapat memahami nuansa linguistik lebih menyeluruh daripada teknik yang lebih konvensional seperti LSTM, CNN atau *machine learning*, terutama dalam kalimat yang rumit atau ambigu. Dibandingkan dengan model sebelumnya, arsitektur pembelajaran mendalam yang telah dilatih sebelumnya menawarkan kinerja yang lebih baik dalam kategorisasi sentimen dan pemahaman kontekstual, yang secara signifikan meningkatkan akurasinya dalam memproses materi berskala besar atau tidak terstruktur.

Di era digital saat ini, volume data yang tersedia secara online terus meningkat, mulai dari artikel berita, laporan penelitian hingga konten media sosial, pengguna seringkali berhadapan dengan tantangan untuk menemukan pemahaman menyeluruh dari data yang tersedia (APJII, 2024). Dalam hal ini, *text summarization* menjadi metode yang berguna untuk membantu pengguna mengolah dan memahami informasi secara efisien. Metode ini tidak hanya menghemat waktu tetapi juga membantu pengambilan keputusan dengan menyajikan informasi kunci secara ringkas.

Dalam beberapa tahun terakhir, kemajuan bidang *Natural Language Processing* (NLP) berkembang signifikan terutama dalam teknik *text summarization*. Salah satu pencapaian terbesar adalah dengan hadirnya model berbasis *deep learning*, salah satunya model BERT (*Bidirectional Encoder Representations from Transformers*), yang pertama diperkenalkan oleh Google pada tahun 2018. BERT terbukti sangat efektif dalam berbagai tugas NLP seperti klasifikasi teks, analisis sentimen dan pemahaman konsep (Liu & Lapata, 2019).

Dalam konteks *text summarization*, kemampuan BERT memahami hubungan semantik antar kalimat dan memahami konteks dengan lebih luas dan dapat menghasilkan ringkasan yang lebih relevan dan berkualitas tinggi. Dengan memanfaatkan mekanisme pada model BERT, *text summarization* dapat menganalisis dan menyaring informasi penting serta memastikan bahwa ringkasan yang dihasilkan mencakup inti dari keseluruhan teks (Zhang et al., 2019).

1.2 Rumusan Masalah

Dari beberapa uraian yang telah dipaparkan pada latar belakang, penelitian ini bertujuan untuk:

- a) Seberapa baik akurasi metode RoBERTa dan Vader dalam menangani analisis sentimen?
- b) Bagaimana cara mengidentifikasi informasi inti dari sebuah video dengan cepat untuk membantu pengguna memahami konteks dan relevansi video secara keseluruhan?
- c) Apakah Coursera bisa menjadi rekomendasi *Massive Open Online Courses* bagi para pelajar?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

- a) Menilai keakuratan metode roBERTa dan VADER dalam menangani analisis sentimen.
- b) Metode *text summarization* membantu pengguna untuk memperoleh pemahaman yang menyeluruh mengenai video *review* tentang Coursera.
- c) Mengevaluasi kualitas Coursera berdasarkan analisis sentimen untuk rekomendasi.

1.3.2 Manfaat

Dari tujuan-tujuan tersebut, manfaat potensial dari pelaksanaan penelitian ini adalah sebagai berikut:

- a) Untuk Peneliti:
 - Mempelajari analisis sentimen dalam ulasan Coursera dengan menerapkan teknik roBERTa dan VADER untuk membuat pilihan yang lebih baik.
 - Mengidentifikasi model yang optimal untuk analisis sentimen.
- b) Untuk Coursera:
 - Mengidentifikasi faktor-faktor kepuasan pelajar yang penting yang mempengaruhi Coursera.
 - Dapat meningkatkan kualitas Coursera berdasarkan ulasan dan sentimen dari pelajar.

c) Untuk Pelajar:

- Membantu menemukan kursus daring terbuka massal berkualitas tinggi.
- Memberikan wawasan tentang kualitas dan efektivitas Coursera.

1.4 Metode

Dalam metode penelitian ini, kami mengumpulkan informasi atau komentar dari video Youtube yang membahas *review* coursera menggunakan teknik *scraping* untuk memberikan lebih banyak konteks. Metode yang digunakan dalam proses analisis sentimen adalah RoBERTa. Untuk mendeteksi sentimen dengan lebih baik, RoBERTa atau model berbasis *deep learning* menghasilkan embedding teks kemudian menerapkan lapisan klasifikasi untuk mengidentifikasi sentimen lebih akurat. Keakuratan metode ini dalam menilai sentimen komentar bertujuan untuk memastikan metode ini bisa efektif dalam mendeteksi opini pengguna mengenai Coursera.

Untuk *text summarization*, kami menggunakan metode BERT yang mampu menghasilkan ringkasan informasi video mengenai *review* coursera. Kami mengekstraksi fitur tekstual dan membuat ringkasan berdasarkan hubungan kalimat dan konteks dengan mengoptimalkan model BERT yang telah di *pre-trained*. Dengan menggunakan metode ini, kami dapat meringkas poin-poin utama dari video *review* tentang Coursera dan memperoleh pemahaman yang menyeluruh mengenai video-video tersebut.

BAB 2 TINJAUAN PUSTAKA

2.1 Analisis sentimen

Analisis sentimen adalah teknik pemrosesan bahasa alami yang digunakan untuk memutuskan sentimen dan mengurainya menjadi positif, negatif, atau netral (Aftab et al., 2023). Analisis sentimen dapat mengungkapkan perasaan dan pemikiran masyarakat umum terhadap suatu barang, layanan, atau produk. Analisis sentimen lebih dari sekadar penambangan sentimen, tetapi juga penambangan kontekstual teks, yang menemukan dan mengekstrak informasi subjektif dari materi sumber dan membantu perusahaan memahami sentimen sosial merek, produk, atau layanannya sambil mengawasi diskusi daring. Alat kategorisasi teks yang paling populer adalah analisis sentimen, yang menentukan apakah opini utama yang diungkapkan dalam pesan masuk bersifat positif, netral, atau negatif (Murthy et al., 2020.)

2.2 Massive Open Online Course

Massive Open Online Courses (MOOCs) adalah bentuk pendidikan baru yang muncul sebagai hasil dari pengembangan pembelajaran jarak jauh (Nurutdinova et al. 2023). Perkembangan terbaru dalam pembelajaran jarak jauh adalah diperkenalkannya *Massive Open Online Courses* pada tahun 2012. Melalui penerapan kemajuan ilmiah dan teknologi pada pendidikan daring, MOOCs meningkatkan lingkungan pembelajaran sehingga siswa dapat mengambil kursus pada berbagai disiplin ilmu dengan kecepatan mereka sendiri dengan kewajiban dan batasan biaya yang minimal (Ngo et al. 2024).

2.3 BERT

Model BERT merupakan model yang semua parameternya disesuaikan menggunakan data berlabel dari pekerjaan hilir setelah pertama kali diinisialisasi menggunakan parameter yang telah dilatih sebelumnya (Devlin et al. 2019). BERT merupakan struktur yang dirancang untuk menggunakan representasi teks dua arah (kiri-kanan dan kanan-kiri) secara bersamaan dan menggabungkannya dengan metode *Mask Language Model* (MLM) dengan metode *Next Sentence Prediction* (NSP) secara bersamaan. Oleh karena itu, BERT dapat disimpulkan sebagai metode terbaik untuk memahami teks

dengan konten atau konteks yang kompleks (Nabiilah et al. 2023). Khusus untuk suatu tugas, desain BERT menyediakan opsi untuk mengekspresikan sepasang frasa sebagai frasa tunggal atau sebagai urutan token yang berurutan. Representasi untuk token tertentu dibuat dengan menggabungkan token, segmen, dan embedding yang relevan. Lapisan encoder terakhir dihubungkan ke lapisan yang terhubung penuh, dan kata pertama dalam urutan tersebut diberi token yang berbeda untuk tugas klasifikasi. Dengan menggunakan lapisan *softmax*, kalimat atau pasangan kalimat dapat diklasifikasikan (Mandhasiya et al. 2023).

2.4 ROBERTA

Adaptasi paradigma BERT, RoBERTa (*Robustly Optimized BERT Approach*) bertujuan untuk meningkatkan kinerja dalam tugas-tugas yang melibatkan *Natural Language Processing* (NLP). Perbedaannya ditemukan dalam penggunaan kumpulan data yang lebih besar, ukuran *batch* yang lebih besar, dan pra-pelatihan yang lebih lama dan lebih ekstensif. Untuk meningkatkan efisiensi dan keluaran model, RoBERTa juga menyingkirkan pengaturan *Masked Language Model* (MLM) yang digunakan BERT dan berkonsentrasi pada pengoptimalan hiperparameter. Modifikasi ini telah memungkinkan RoBERTa untuk melampaui BERT dalam sejumlah tolak ukur NLP, termasuk pemahaman teks dan analisis sentimen. (Liu et al., 2019)

BAB 3 METODE IMPLEMENTASI

3.1 Kerangka berpikir

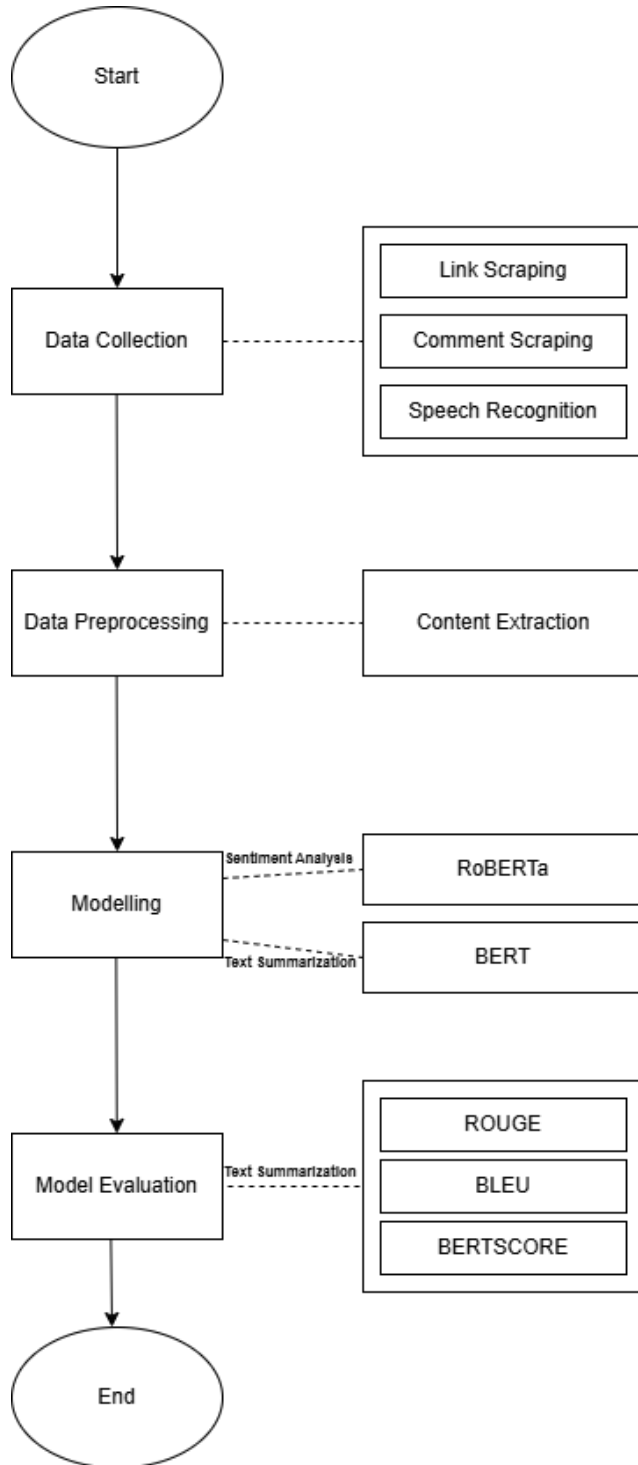


Figure 1. Flowchart

3.2. Metode riset

Pengumpulan data akan menjadi langkah pertama dari lima langkah dalam prosedur penelitian ini, yang akan dilanjutkan dengan praproses data. Persiapan data melibatkan pembersihan data dan mengubahnya menjadi token. Selanjutnya, untuk membandingkan kedua model, data akan dimodelkan menggunakan RoBERTa dan VADER selama tahap pemodelan data untuk *sentiment analysis* dan BERT untuk *text summarization*. Langkah terakhir adalah menilai model yang dibuat dengan membandingkan hasil dari datanya.

3.3 Pengumpulan data

Link scraping, *comment scraping*, dan *speech recognition* yang memanfaatkan Google Speech-to-Text API semuanya digunakan dalam proses pengumpulan data untuk studi ini. Ulasan URL dikumpulkan menggunakan *comment scraping* Selenium, dan komentar YouTube di Coursera diperoleh menggunakan *comment scraping*. Audio dari evaluasi video ditranskripsikan dengan cepat dan tepat ke dalam teks menggunakan pengenalan ucapan. Semua data ini disiapkan untuk ringkasan teks, yang secara efektif menghasilkan wawasan ulasan yang penting.

3.4 Pra Proses data

Lima kata kunci utama yang menjadi fokus *data preprocessing* adalah 'coursera', 'kursus daring', 'platform', 'kursus', dan 'pembelajaran'. Untuk memastikan analisis tetap relevan dengan evaluasi dan percakapan mengenai Coursera dan pembelajaran daring, prosedur ini melibatkan tokenisasi teks dan penyaringan data menggunakan kata kunci ini.

3.5 Modelling

Sentiment analysis dan *text summarization* merupakan dua proses utama dalam pemodelan. metode yang digunakan untuk analisis sentimen: RoBERTa, model berbasis Transformer yang telah diajarkan untuk memahami konteks ulasan dengan akurasi tinggi. BERT, yang dapat memberikan ringkasan kontekstual dari data teks dengan mempertimbangkan hubungan antara frasa, kemudian digunakan untuk melakukan ringkasan teks. Metodologi ini menjamin analisis dan Coursera *review summary* yang komprehensif dan relevan.

3.6 Evaluasi model

Mengevaluasi model hasil analisis sentimen menggunakan RoBERTa, kita bisa mengukur jumlah prediksi yang masuk ke dalam masing-masing kelas sentimen seperti 'Positive', 'Negative', dan 'Neutral'. Dengan menghitung distribusi hasil klasifikasi dari model, kita dapat mengetahui berapa banyak data yang diprediksi dengan masing-masing label ini, yang akan memberikan gambaran seberapa baik model dalam mengidentifikasi masing-masing kelas.

Mengevaluasi model hasil text summarization menggunakan BERT, kita dapat menggunakan beberapa *metrics* evaluasi seperti BLEU , BERTScore, dan ROUGE. Ketiga metrik ini memberikan pandangan yang komprehensif mengenai kualitas ringkasan yang dihasilkan oleh model BERT.

BAB 4 HASIL DAN EVALUASI

4.1 Analisis Sentimen

Pada penelitian ini, analisis sentimen menggunakan metode RoBERTa menghasilkan distribusi

Table 1. Sentiment Distribution by RoBERTa

	Labels	Counts
0	Positive	228
1	Negative	180
2	Neutral	179

Dari hasil distribusi di atas, metode RoBERTa menunjukkan distribusi sentimen yang relatif seimbang. Dengan hasil yang cenderung seimbang antara label sentimen positif, negatif dan netral dapat mengindikasikan bahwa metode RoBERTa mampu memahami nuansa sentimen yang lebih kompleks. Model ini tidak hanya sensitif terhadap teks positif, tapi juga mampu mengidentifikasi teks dengan sentimen negatif dan netral.

4.2 Text Summarization

1	title	summarized text
2	Coursera Review Our Experience and How it Works	Coursera is a super popular openonline education platform launched in 2012 and partnered with major universities and compan
3	Coursera Review My Thoughts After 5 Years and 40 Online Courses	Richard Wallace has been on Coursera for about the last five years and taken around 40 courses. In this video he'll be doing a co
4	Top 5 Online Learning Platforms 2024 Review of Coursera SkillShare Udemy EdX LinkedIn	if you want to study something new but don't have the time or money to enroll in a college degree program you should conside
5	Coursera Review The Best Elearning Site	Coursera is a learning platform it's an e-learning platform but it is very different from other e- learning platforms like LinkedIn le
6	Palo Alto Networks Cybersecurity Professional Certificate Review Coursera	Palo Alto networks cybersecurity professional certificate offered on Coursera. Designed to provide Learners with essential skills in
7	ONLY UNSPONSORED Review of the Google Cybersecurity Certificate From Coursera	The Google cybersecurity certificate is designed to prepare you for a career as a cyber security analyst with a professional certific
8	Coursera Review Why Is It the Best Online Learning Platform	my favorite online learning platforms so let's get started call Sarah. Call Sarah partners with over 2 7 Scompanies and universities
9	Coursera Plus 2023 Review 7000 Online Courses shortvideo	unlock unlimited learning with Coursera plus hey today we're exploring Corsair plus a fantastic resource for lifelong Learners like
10	Coursera Review Best Platform for Courses	In this video we're going to be taking a look at Coursera and seeing if it's truly good as people say it is. Once you have signed up

Text summarization menggunakan metode BERT (*Bidirectional Encoder Representations from Transformers*) telah dilakukan dalam penelitian ini untuk merangkum konten dari video-video di Youtube yang membahas mengenai *review* dari Coursera. Model BERT digunakan dalam proses ini untuk memahami konteks dari kalimat yang panjang, kemudian menghasilkan ringkasan yang singkat dengan tetap mempertahankan informasi penting dari kalimat tersebut.

Dalam penelitian ini, *text summarization* menghasilkan ringkasan sebanyak 19 video Youtube yang membahas mengenai *review* dari Coursera. Hasil diatas menunjukan bahwa BERT dapat memahami konteks kata dalam kalimat secara dua arah dan menghasilkan ringkasan yang bersifat informatif dan juga relevan meskipun kalimatnya panjang.

4.3 Evaluasi

Berdasarkan hasil analisis sentimen menggunakan metode RoBERTa, model ini menunjukkan performa yang seimbang dalam mengidentifikasi berbagai kategori sentimen, dengan distribusi sentimen positif, negatif, dan netral yang relatif merata. Hal ini menunjukkan kemampuan RoBERTa dalam memahami nuansa sentimen yang lebih kompleks, berkat pendekatan berbasis transformer yang digunakan. Keunggulan ini juga terlihat dari sensitivitasnya dalam mendeteksi emosi negatif dan netral, yang sering kali lebih sulit diidentifikasi oleh model analisis sentimen tradisional.

Metrics	Score
Rouge 1	0.22627951204776764
Rouge 2	0.1331196278333664
Rouge L	0.22578610479832696

- Rouge 1 memiliki skor 22,6%, menunjukan bahwa kata-kata penting dalam ringkasan yang dihasilkan tidak terlalu sesuai. Hal ini dapat terjadi jika ringkasan yang dihasilkan menggunakan kata-kata yang berbeda.
- Rouge 2 memiliki skor 13,3% yang menunjukan bahwa struktur kalimat pada ringkasan yang dihasilkan cukup berbeda dengan referensi.
- Rouge L memiliki skor 22,5% yang menunjukan bahwa urutan kata dalam ringkasan belum sepenuhnya mencerminkan struktur teks referensi.

Model tidak cukup menangkap struktur penting dari asal teks karena rendahnya skor ROUGE yang artinya memiliki sedikit kecocokan dalam coverage kata dan struktur dibandingkan dengan referensi.

Bleu: 0.015665937215089798

Bleu memiliki skor 1,56% menunjukan bahwa ada sedikit atau hampir tidak ada kesesuaian kata secara langsung antara ringkasan yang dihasilkan dengan referensi.

Bertscore : 0.8653887510299683

Bertscore memiliki skor 86,5%, lebih tinggi dibandingkan ROUGE dan BLEU. Ini menunjukkan bahwa meskipun ada perbedaan kata atau struktur, makna dan inti dari ringkasan yang dihasilkan sangat relevan dengan referensi.

BAB 5 KESIMPULAN

Dengan menghitung jumlah prediksi untuk kelas "Positif," "Negatif," dan "Netral", evaluasi model analisis sentimen RoBERTa memberikan wawasan tentang distribusi sentimen kumpulan data. Ini membantu pemahaman kita tentang akurasi model dalam klasifikasi sentimen. Namun, evaluasi BERT terhadap model ringkasan teks menggunakan metrik seperti ROUGE, BLEU, dan BERTScore memberikan gambaran menyeluruh tentang kualitas ringkasan yang dihasilkan. Bersama-sama, ketiga ukuran ini memberikan gambaran menyeluruh tentang seberapa efektif model tersebut menghasilkan ringkasan yang akurat dan relevan. BERTScore memberikan penilaian yang lebih menyeluruh berdasarkan kesamaan semantik, sementara metrik BLEU dan ROUGE berkonsentrasi pada kesamaan n-gram.

BAB 6 DAFTAR PUSTAKA

- Aftab, F., Bazai, S. U., Marjan, S., Baloch, L., Aslam, S., Amphawan, A., & Neo, T. K. (2023). A Comprehensive Survey on Sentiment Analysis Techniques. *International Journal of Technology*, 14(6), 1288. <https://doi.org/10.14716/ijtech.v14i6.6632>
- Asosiasi Penyelenggara Jasa Internet Indonesia. (n.d.). <https://apjii.or.id/berita/d/apjii-jumlah-pengguna-internet-indonesia-tembus-221-juta-orang#:~:text=APJII%20Jumlah%20Pengguna%20Internet%20Indonesia,jiwa%20penduduk%20Indonesia%20tahun%202023>
- Chen, X., Wang, F. L., Cheng, G., Chow, M., & Xie, H. (2022). Understanding Learners' Perception of MOOCs Based on Review Data Analysis Using Deep Learning and Sentiment Analysis. *Future Internet*, 14(8), 218. <https://doi.org/10.3390/fi14080218>
- Chiny, M., Chihab, M., Chihab, Y., Laboratory of Computer Sciences, Ibn Tofail University, Bencharef, O., Department of Computer Sciences, & Cadi Ayyad University. (2021). LSTM, VADER and TF-IDF based Hybrid Sentiment Analysis Model. *International Journal of Advanced Computer Science and Applications*, Vol. 12(No. 7), 265–266. <https://www.ijacsa.thesai.org>
- Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., & Google AI Language. (n.d.). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT 2019* (pp. 4171–4186) [Conference-proceeding]. Association for Computational Linguistics. <https://aclanthology.org/N19-1423.pdf> (Original work published 2019)
- Ibrahim Lazrig, & Sean L. Humpherys. (2022). Using Machine Learning Sentiment Analysis to Evaluate Learning Impact. *Information Systems Education Journal (ISEDJ)*, 1545-679X.
- Liu, Y., & Lapata, M. (2019). Text Summarization with Pretrained Encoders. *Cornell University*. <https://doi.org/10.18653/v1/d19-1387>
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019, July 26). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. arXiv.org.

<https://arxiv.org/abs/1907.11692v1>

- Mandhasiya, D. G., Murfi, H., & Bustamam, A. (2024, January 1). *The hybrid of BERT and deep learning models for Indonesian sentiment analysis*. Guna Mandhasiya | Indonesian Journal of Electrical Engineering and Computer Science. <https://ijeecs.iaescore.com/index.php/IJEECS/article/view/33790/17898>
- Mrhar, K., Benhiba, L., Bouekkache, S., & Abik, M. (2021). A Bayesian CNN-LSTM Model for Sentiment Analysis in Massive Open Online Courses MOOCs. *International Journal of Emerging Technologies in Learning (iJET)*, 16(23), 216–232. <https://doi.org/10.3991/ijet.v16i23.24457>
- Murthy, N. D. G. S. N., Allu, N. S. R., Andhavarapu, N. B., & Belusonti, N. M. B. M. (2020). Text based Sentiment Analysis using LSTM. *International Journal of Engineering Research And*, V9(05). <https://doi.org/10.17577/ijertv9is050290>
- N. D. Infante, M. Lazar, S. Ram and A. Ray, Demand for online education is growing. Are providers ready?, Jul. 2022, [online] Available: <https://www.mckinsey.com/industries/education/our-insights/demand-for-online-education-is-growing-are-providers-ready#/>.
- Nabiilah, G. Z., Prasetyo, S. Y., Izdihar, Z. N., & Girsang, A. S. (2023). BERT base model for toxic comment analysis on Indonesian social media. *Procedia Computer Science*, 216, 714–721. <https://doi.org/10.1016/j.procs.2022.12.188>
- Ngo, T. T. A., Tran, T. T., An, G. K., & Nguyen, P. T. (2023). Students' Perception Towards Learning Massive Open Online Courses on Coursera Platform: Benefits and Barriers. *International Journal of Emerging Technologies in Learning (iJET)*, 18(14), 4–23. <https://doi.org/10.3991/ijet.v18i14.39903>
- Nurutdinova, A. R., Shakirova, D. S., Sabaeva, E. K., & Samarkina, N. O. (2023). The content modification within the framework of the massive open online courses (case study: International and Russian practices). In *IntechOpen eBooks*. <https://doi.org/10.5772/intechopen.1001342>
- Saoualih, A., Safaa, L., Bouhatous, A., Bidan, M., Perkumienė, D., Aleinikovas, M., Šilinskas, B., & Perkumas, A. (2024). Exploring the Tourist Experience of the Majorelle Garden Using

VADER-Based Sentiment Analysis and the Latent Dirichlet Allocation Algorithm: The Case of TripAdvisor Reviews. *Sustainability*, 16(15), 6378. <https://doi.org/10.3390/su16156378>

Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. *Applied Sciences*, 13(7), 4550. <https://doi.org/10.3390/app13074550>

Zhang, H., Cai, J., Xu, J., & Wang, J. (2019). Pretraining-Based Natural Language Generation for Text Summarization. *Acl Anthology*. <https://doi.org/10.18653/v1/k19-1074>