

Prediksi Sentimen Pada Teks Media Sosial *Corporate University* Menggunakan RoBERTa

Yogie Oktavianus Sihombing, ^{a,1,*}, Nita Vibriyanti Situmorang ^{b,2}

^a Badan Kepegawaian Negara, Jl. Mayjen Sutoyo No. 12, Jakarta Timur, 13640

^b Badan Kepegawaian Negara, Jl. Mayjen Sutoyo No. 12, Jakarta Timur, 13640

¹ yogie.sihombing@bkn.go.id; ² nita.situmorang@bkn.go.id

* corresponding author

ARTICLE INFO

ABSTRACT / ABSTRAK

Article history

Received:

Revised:

Accepted:

Sentimen publik yang tercermin dalam media sosial memainkan peran sentral dalam membentuk citra sebuah *corporate university*. Penelitian ini bertujuan untuk melakukan prediksi sentimen pada data teks media sosial *corporate university* menggunakan salah satu model kecerdasan buatan untuk teknik *Natural Language Processing* (NLP) yaitu RoBERTa (*a Robustly Optimized BERT Pretraining Approach*). RoBERTa merupakan salah satu model NLP yang dilatih dalam korpus bahasa tertentu, kompleks dan mampu menghasilkan representasi teks dengan baik, yang berpotensi meningkatkan kinerja dalam memahami konteks dan niat di balik teks media sosial. Dalam kasus ini, RoBERTa dilatih dalam korpus Bahasa Indonesia yang disebut IndoRoBERTa. Pada penelitian ini, dikumpulkan dataset berbasis teks untuk dilakukan pra-pemrosesan data. Setelah itu, dilakukan konfigurasi hiperparameter yang dilanjutkan dengan *fine-tuning* dari model untuk dilatih dalam tugas analisis sentimen. Berdasarkan hasil *fine-tuning*, secara komparatif, IndoRoBERTa mengungguli varian IndoBERT lainnya dalam tugas analisis sentimen. Selanjutnya dilakukan evaluasi model IndoRoBERTa berdasarkan laporan klasifikasi, dan didapatkan bahwa model yang dilatih menghasilkan nilai rata-rata untuk akurasi tes 96.2% dan *F1-Score* 95.2% dari tiga kelas sentimen: negatif, netral, positif. Selain itu, untuk keseluruhan evaluasi model berdasarkan *confusion matrix*, sensitivitas, spesifisitas, MCC, indeks Kappa Cohen, dan kurva ROC, diperoleh hasil kinerja yang baik. Terakhir, model yang telah dievaluasi, dilakukan uji prediksi sentimen dari sejumlah sampel postingan teks media sosial yang berkaitan dengan *corporate university*, dan didapatkan nilai prediksi rata-rata dari sampel lebih dari 97%. Secara keseluruhan, hasil prediksi ini membantu bagaimana *corporate university* merespon publik terhadap layanan dan produk, khususnya layanan pendidikan dan pelatihan SDM.

Public sentiment reflected in social media plays a central role in shaping the image of a corporate university. This research aims to predict sentiment on corporate university social media text data using one of the artificial intelligence models for Natural Language Processing (NLP) techniques, namely RoBERTa (a Robustly Optimized BERT Pretraining Approach). RoBERTa is one of the NLP models trained in a specific, complex language corpus and is capable of producing good text representations, which has the potential to improve performance in understanding the context and intention behind social media texts. In this case, RoBERTa is trained in an Indonesian corpus called IndoRoBERTa. In this study, a text-based dataset was collected for data pre-processing. After that, hyperparameter configuration is performed followed by fine-tuning of the model to be trained in sentiment analysis tasks. Based on the fine-tuning results, comparatively, IndoRoBERTa outperforms other IndoBERT variants in sentiment analysis task. Next, an evaluation of the IndoRoBERTa model based on the classification report was conducted, and it was found that the trained model produced average values for test accuracy of 96.2% and F1-Score of 95.2% of the three sentiment classes: negative, neutral, positive. In addition, for the overall model evaluation based on confusion matrix, sensitivity, specificity, MCC, Cohen's Kappa index, and ROC curve, good performance results were obtained. Finally, the evaluated model was tested for sentiment prediction from a number of samples of social media text posts related to the corporate university, and the average prediction value of the samples was more than 97%. Overall, these prediction results help how corporate universities respond to the public on services and products, especially HR education and training services.

This is an open access article under the [CC-BY-SA](#) license.



Kata Kunci: sentimen, prediksi, RoBERTa, media sosial, *corporate university*

Keywords: sentiment, prediction, RoBERTa, social media, *corporate university*

1. Pendahuluan

Dalam kehidupan sosial, opini merupakan titik sentral dalam seluruh aktivitas manusia. Sebagian besar hasil dari opini dapat memberikan pengaruh dalam suatu pengambilan keputusan, kebijakan, dan tingkat reputasi suatu organisasi. Konsep dasar opini salah satunya adalah klasifikasi sentimen yang merupakan bagian dari studi analisis sentimen. Analisis sentimen merupakan disiplin ilmu komputasi yang bertujuan untuk mengidentifikasi secara otomatis keadaan pribadi dan pikiran manusia, seperti sentimen, emosi, perilaku, dan keyakinan, dari tekstual, audio, video, atau data lain atau kombinasinya (Poria et al., 2018). Lebih lanjut, Das (2023) menjelaskan, sentimen menunjukkan perasaan seseorang terhadap sesuatu dan secara luas diekspresikan, dirasakan, dikenali, diproses, serta ditafsirkan dalam bahasa alami. Klasifikasi sentimen biasanya terdiri dari tiga kelas sentimen: negatif, positif, dan netral. Sentimen positif dapat diinterpretasikan dengan berbagai emosi seperti sukacita, syukur, dan bangga. Demikian juga sentimen negatif, yang dikaitkan dengan beberapa emosi seperti kemarahan, ketakutan, dan kesedihan. Sedangkan sentimen netral menunjukkan sesuatu yang tidak positif maupun negatif. Dilihat dari sudut pandang Gasper et al (2021), menyatakan bahwa pengaruh netral bersifat independen dan dapat memiliki kecenderungan ke seluruh dimensi perasaan. Perasaan netral didefinisikan sebagai pengalaman yang dirasakan dimana seseorang merasa acuh tak acuh, tidak spesifik, dan tidak memiliki preferensi yang kuat. Analisis sentimen juga dikaitkan dengan pendeteksian sentimen dan emosi dalam ulasan dan komentar dari publik di media sosial yang mengekspresikan pendapat tentang produk atau peristiwa tertentu (Muthasima et al., 2019).

Dalam beberapa tahun terakhir, para ilmuwan telah menerapkan teknologi untuk mengklasifikasikan sentimen menggunakan kecerdasan buatan. Model dilatih dengan teknik *Natural Language Processing* (NLP) untuk mengenali aspek sentimen berbasis teks. NLP menurut Olaoye & Potter (2024) adalah cabang dari kecerdasan buatan yang berkaitan dengan kemampuan komputer untuk memahami dan memproses bahasa alami manusia. NLP menggabungkan ilmu bahasa dan komputasi dengan metode statistik serta pembelajaran mesin (*machine learning/deep learning*) untuk membantu sistem menganalisis teks dan audio manusia, memahami makna, serta menangkap maksud dan sentimen didalamnya.

Perkembangan studi analisis sentimen dan NLP bertepatan dengan pertumbuhan media sosial di Indonesia. Pada tahun 2023, secara global, pasar NLP tumbuh hingga 30%, hingga tahun 2030 diproyeksikan akan stabil diangka 10-15% (Statista, 2024). Berdasarkan data dari Kemp (2024), awal tahun 2024 tingkat penetrasi internet di Indonesia mencapai 66.5% dari total populasi pada awal tahun 2024 dari jumlah populasi, yaitu 276.4 juta penduduk Indonesia. Ditambah lagi, jumlah persentase pengguna media sosial di Indonesia mewakili 49,9% dari total populasi pada awal tahun 2024. Hal ini mengindikasikan bahwa analisis sentimen dengan pemanfaatan sistem kecerdasan buatan melalui media sosial dapat menjadi alat monitoring dan evaluasi yang cukup potensial bagi organisasi pemerintahan, tidak terkecuali sistem *corporate university* yang melekat pada organisasi pemerintahan.

Saat ini, regulasi *corporate university* di Indonesia telah diatur dalam Peraturan LAN RI Nomor 6 Tahun 2023 tentang Sistem Pembelajaran Pengembangan Kompetensi Secara Terintegrasi (*Corporate University*), yang dipahami sebagai pola sistem pengajaran terpadu untuk mengembangkan keterampilan Aparatur Sipil Negara (ASN) sesuai dengan peraturan perundang-undangan di bidang manajemen ASN (Setiawati et al., 2023). Tentunya pengaruh sentimen publik di media sosial memberikan dampak cukup signifikan terhadap citra organisasi pemerintahan yang menerapkan sistem *corporate university*. Sentimen publik di media sosial dapat mempengaruhi reputasi organisasi, sentimen negatif cenderung merusak, sedangkan sentimen positif cenderung memperbaiki dan meningkatkan reputasi (Etter et al., 2018; Anderson et al., 2021). Artinya, sentimen positif dapat meningkatkan reputasi dari *corporate university*. Jika *corporate university* memberikan kualitas layanan yang baik kepada peserta pelatihan, maka peserta pelatihan dapat memberikan apresiasi secara lisan maupun tertulis melalui komentar positif di media sosial, secara tidak langsung komentar tersebut membangun kepercayaan publik, sehingga reputasi dari *corporate university* pun juga meningkat. Sebaliknya, sentimen negatif seperti kritik atau ulasan buruk di media sosial, dapat mengurangi reputasi *corporate university*. Dampak negatifnya, dapat mengurangi minat calon peserta pelatihan untuk berpartisipasi dalam program pelatihan selanjutnya dari *corporate university*, bahkan dapat mempengaruhi pandangan publik terhadap citra organisasi secara keseluruhan. Organisasi pemerintahan yang menerapkan sistem *corporate university* sudah semestinya tanggap terhadap sentimen publik, sehingga dapat menyesuaikan strategi untuk mengatasi kritik

dan memperbaiki program pelatihannya. Respon yang cepat dan tepat terhadap sentimen negatif dapat membantu meminimalkan dampak negatifnya (Kapeš et al., 2022).

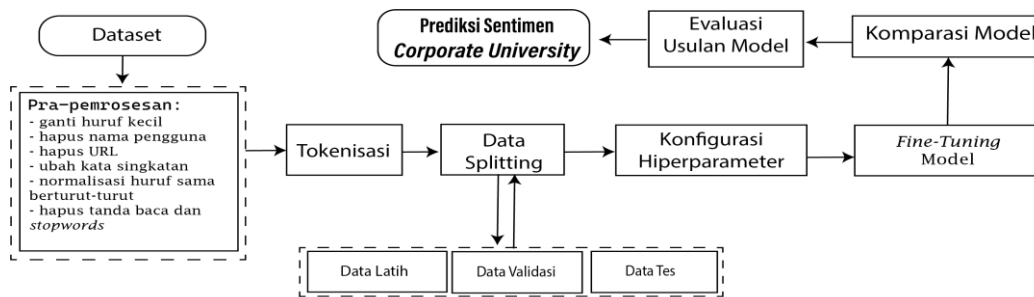
Oleh sebab itu, beberapa penelitian terbaru di bidang NLP Indonesia telah dilakukan untuk memprediksi sentimen berbasis teks dengan lebih efektif dan efisien, yaitu BERT (*Bidirectional Encoder Representations from Transformers*). Beberapa metode NLP yang dikembangkan dari model BERT secara khusus dilatih dalam korpus bahasa Indonesia, yang cukup dikenal adalah IndoBERT dari IndoNLU (Wilie et al., 2019) dan IndoLEM (Koto et al., 2020). Bahkan ada sebuah model yang dikembangkan dari BERT, yaitu RoBERTa (*Robustly Optimized BERT Pretraining Approach*). RoBERTa bekerja dengan melakukan modifikasi sederhana, seperti melatih model lama dengan jumlah data yang lebih banyak, menghapus fitur prediksi kalimat berikutnya (*next sentence prediction*) dan melatih kumpulan kalimat yang lebih panjang, dengan tujuan hasil yang lebih optimal dibandingkan BERT (Liu et al., 2019). Sebagai contoh, dalam penelitian sebelumnya yang dilakukan secara komparatif, mengenai klasifikasi teks *clickbait* oleh Rajapaksha et al (2021), RoBERTa mengungguli model lain, seperti BERT dan XLNet. Hal ini menunjukkan RoBERTa telah dilatih sebelumnya pada korpus data yang besar dibandingkan model lainnya termasuk data media berita untuk data pelatihannya. Selanjutnya, penelitian oleh Naseer et al (2021) yang membandingkan secara empiris 3 algoritma dalam hal verifikasi fakta: BERT, RoBERTa dan Electra, dengan hasil eksperimen menunjukkan bahwa RoBERTa mencapai akurasi dan *F1-Score* terbaik.

Pada penelitian sebelumnya terkait analisis sentimen menggunakan model BERT dan RoBERTa berbahasa Indonesia, yang mana telah dilakukan dalam penelitian dari Fransiscus & Girsang (2022), menggunakan IndoBERT untuk analisis sentimen dengan *F1-Score* 84% menggunakan dua kelas sentimen: positif dan negatif. Pada penelitian Sandra et al (2022) menggunakan model IndoRoBERTa untuk dua kelas sentimen: negatif dan positif, didapatkan hasil akurasi 91.73% dan *F1-Score* 86.42%. Secara garis besar, penelitian tersebut menggunakan sentimen kelas biner: negatif dan positif.

Pada penelitian kali ini dilakukan pendekatan klasifikasi sentimen multikelas untuk usulan model analisis sentimen, yaitu tiga kelas sentimen: negatif, netral dan positif. Tujuan penelitian adalah menganalisis hasil *fine-tuning* secara komparatif dari usulan model IndoRoBERTa dan varian IndoBERT dalam melakukan tugas analisis sentimen sehingga diketahui apakah model IndoRoBERTa lebih unggul dibandingkan varian IndoBERT lainnya atau sebaliknya. Setelah didapatkan hasil pelatihan model yang paling unggul, dilakukan evaluasi kinerja usulan model sentimen multikelas yang telah dilatih berdasarkan evaluasi laporan klasifikasi, *confusion matrix*, sensitivitas, spesifisitas, MCC (*Matthews Correlation Coefficient*), indeks Kappa Cohen dan kurva ROC (*Receiver Operating Characteristic*). Setelah model dievaluasi, dilanjutkan dengan uji prediksi pada sejumlah sampel teks berupa komentar dari konten media sosial *corporate university* berdasarkan model yang dibuat. Uji prediksi ini untuk mengetahui seberapa baik penerapan model dalam melakukan prediksi dari data sampel teks mentah media sosial dan mengetahui apakah model dapat menginterpretasikan makna dari sampel yang dilakukan uji prediksi berdasarkan kelas sentimen. Diharapkan dari penelitian ini dapat memberikan manfaat dalam bidang pengelolaan teknologi informasi dan komunikasi (TIK), sehingga model dapat diterapkan dan dikembangkan pada organisasi pemerintahan untuk meningkatkan kualitas pelatihan dan pengembangan SDM, khususnya *corporate university*.

2. Metodologi

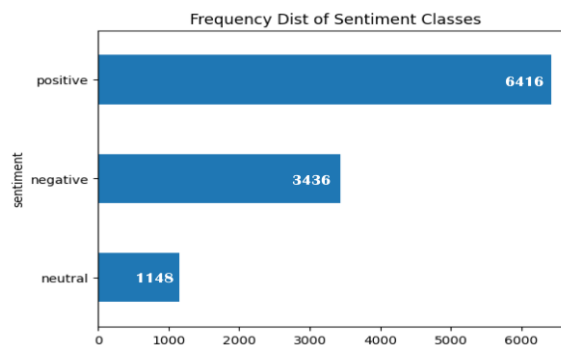
Desain penelitian menggunakan metode eksperimen komputasi perangkat lunak dengan algoritma untuk memproses, menghasilkan, dan menganalisis data secara efisien dan konsisten (Recker, 2021). Dalam penelitian ini memanfaatkan pemrograman *Python* dengan pustaka *Transformers* yang digunakan pada usulan model. Perangkat komputasi yang digunakan adalah Intel(R) Core(TM) i7-9750H CPU, NVIDIA GeForce RTX 2060 GPU dan 16 GB RAM. Untuk tahapan alur penelitian digambarkan pada Gambar 1.



Gambar 1. Alur Kerja Usulan Model Untuk Prediksi Sentimen *Corporate University*

2.1. Dataset

Tahap awal, dilakukan pengambilan dataset analisis sentimen, yang bersumber dari berbagai korpus berbahasa Indonesia dari penelitian Koto et al (2020, 2021), meliputi portal berita *online*, media sosial Twitter, artikel Wikipedia, dan berbagai sumber korpus bahasa Indonesia lainnya. Total dataset adalah 11000 baris data yang terbagi atas kelas negatif berjumlah 3436 data, netral 1148 data, dan positif 6416 data yang dapat dilihat pada Gambar 2.



Gambar 2. Distribusi Data Kelas Sentimen

2.2. Pra-Pemrosesan

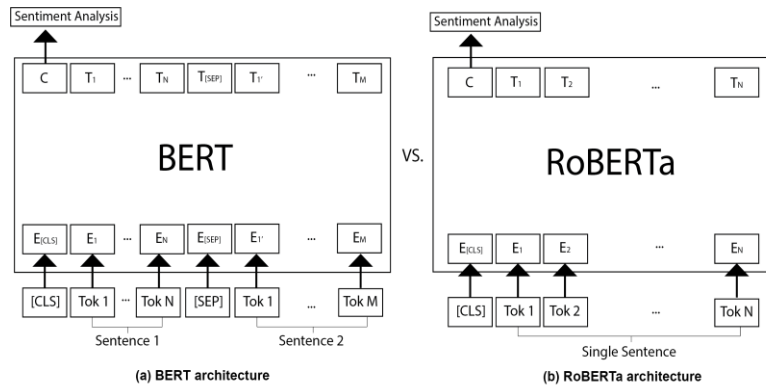
Tahap kedua, dilakukan pra-pemrosesan dari dataset. Pra-pemrosesan data tetap menjadi langkah penting dalam studi pembelajaran mesin. Perlu beberapa proses yang disesuaikan terhadap data yang tidak seimbang, sekurang-kurangnya agar data menjadi lebih bersih dan berkualitas selama proses pembelajaran (Felix & Lee, 2019). Ada pun langkah-langkah pra-pemrosesan yang dilakukan adalah:

- Langkah 1: Ubah setiap kata menjadi huruf kecil
- Langkah 2: Hilangkan nama pengguna
- Langkah 3: Hilangkan URL
- Langkah 4: Ubah kata singkatan menjadi kata baku
- Langkah 5: Normalisasi satu huruf yang berlebih berturut-turut
- Langkah 6: Menghapus semua tanda baca dan *stopwords*

Langkah pertama digunakan untuk mengubah kata yang mengandung huruf kapital menjadi huruf kecil contohnya 'Aku Benci Dia' menjadi 'aku benci dia'. Langkah kedua adalah menghilangkan kalimat yang mengandung unsur nama pengguna, seperti '@nama_pengguna'. Langkah ketiga digunakan untuk menghilangkan kalimat yang mengandung URL atau *link* website seperti "http://". Langkah keempat adalah mengubah kata yang dianggap sebagai kata singkatan menjadi kata baku, contohnya 'krn' menjadi 'karena'. Langkah kelima yaitu melakukan normalisasi kata yang mengandung lebih dari satu huruf yang sama, contohnya 'diaaaaa' dan 'minummm' menjadi 'dia' dan 'minum'. Langkah keenam digunakan untuk menghapus tanda baca yang tidak diperlukan, contohnya yaitu "!\"#\$%&()*+-./:;<=>?@[\\]^_`{|}~". Kemudian penghapusan *stopwords* atau kata yang tidak cukup bermakna dan kurang informatif dalam kalimat seperti 'dan', 'tapi', 'atau', dan sebagainya (Sarica & Luo, 2021).

2.3. Tokenisasi

Tahap ketiga adalah proses tokenisasi, yaitu proses memecah teks mentah menjadi unit-unit kecil yang disebut token. Adapun perbedaan tokenisasi pada arsitektur RoBERTa dan BERT untuk tugas analisis sentimen, dideskripsikan melalui Gambar 3.



Gambar 3. Arsitektur BERT dan RoBERTa

Berdasarkan Gambar 3 (a), proses tokenisasi pada BERT dimulai dengan memecah teks masukan menjadi token menggunakan *WordPiece* tokenizer. Token khusus [CLS] ditambahkan di awal teks, dan token [SEP] ditambahkan di akhir atau di antara dua kalimat. Setiap token kemudian diubah menjadi embedding melalui penjumlahan embedding token, embedding segmen, dan embedding posisi. Embedding ini diproses oleh beberapa lapisan transformer untuk menghasilkan representasi kontekstual dari setiap token (Devlin et al., 2018).

Sementara itu pada Gambar 3 (b), proses tokenisasi pada RoBERTa menggunakan *Byte-Pair Encoding* (BPE) tokenizer untuk memecah teks menjadi token. Token [CLS] ditambahkan di awal teks, tetapi RoBERTa tidak menggunakan token [SEP] untuk memisahkan kalimat. Setiap token juga diubah menjadi embedding melalui penjumlahan embedding token, embedding segmen, dan embedding posisi, kemudian diproses oleh beberapa lapisan transformer untuk menghasilkan representasi kontekstual. RoBERTa lebih fleksibel dalam menangani teks panjang karena tidak menggunakan [SEP] untuk memisahkan kalimat, membuatnya lebih efisien dalam memproses urutan teks yang panjang. Selain itu, RoBERTa sering dilatih dengan lebih banyak data dan urutan teks yang lebih panjang, yang membuatnya lebih efektif dalam beberapa tugas pemrosesan bahasa alami dibandingkan BERT (Liu et al., 2019).

2.4. Data Splitting, Konfigurasi Hiperparameter, Fine-Tuning dan Komparasi Model

Tahap selanjutnya adalah pemisahan data atau data *splitting* menjadi data latih, data validasi, dan data tes, masing-masing dengan rasio 60:20:20 dari total dataset. Dilanjutkan dengan konfigurasi hiperparameter yang telah ditentukan nilainya, terdiri dari $Epoch = 10$, $Learning Rate = 1e-05$, $Batch Size = 32$, $Optimizer = AdamW$ dan $Weight Decay = 0.01$.

Setelah dilakukan konfigurasi, dilanjutkan dengan proses *fine-tuning* model IndoRoBERTa dan beberapa varian IndoBERT. Hal ini dilakukan untuk menilai secara komparatif antara hasil *fine-tuning* model IndoRoBERTa dan beberapa varian IndoBERT berdasarkan *F1-Score* dan nilai akurasi tes untuk pengklasifikasian sentimen.

Untuk usulan model IndoRoBERTa berasal dari *Flax Community* yang dibuat oleh Wongso (2023) untuk tugas klasifikasi sentimen berdasarkan model dari Liu et al. (2019), yaitu *indonesian-roberta-base-sentiment-classifier*. Sedangkan, untuk varian IndoBERT yang dipakai adalah dari IndoNLU dan IndoLEM. IndoNLU dan IndoLEM merupakan pengembangan model BERT khusus bahasa Indonesia berdasarkan model dari penelitian Wilie et al (2019) dan Koto et al (2020, 2021) yang saat ini menjadi tolak ukur dalam melakukan evaluasi *fine-tuning* untuk tugas dalam bahasa Indonesia (Sebastian et al., 2022). Pada IndoNLU, varian IndoBERT yang digunakan adalah *indobert-base-p1* dan *indobert-base-p2*, sedangkan pada IndoLEM, digunakan varian *indobert-base-uncased* dan *indobertweet-base-uncased*.

2.5. Evaluasi Usulan Model & Uji Prediksi Sentimen Universitas Korporat.

Pada tahap akhir, dilakukan evaluasi dari usulan model yang paling optimal dari hasil *fine-tuning* berdasarkan laporan klasifikasi, *confusion matrix*, sensitivitas, spesifisitas, MCC, indeks Kappa Cohen dan kurva ROC. Setelah itu, model yang telah dilatih dilakukan pengujian prediksi sentimen pada sejumlah 20 sampel acak komentar berbasis teks dari media sosial *corporate university*, seperti Instagram, Twitter, dan Youtube.

3. Hasil dan Pembahasan

Bagian ini dilakukan analisis hasil secara menyeluruh rangkaian pekerjaan berdasarkan Gambar 1. Hasil dan pembahasan diuraikan dalam beberapa bagian, yaitu hasil *fine-tuning* dan komparasi antara model IndoRoBERTa dan beberapa varian IndoBERT. Setelah itu, dilanjutkan hasil evaluasi dari pelatihan model yang paling optimal dari komparasi hasil *fine-tuning*, yaitu laporan klasifikasi, *confusion matrix*, sensitivitas, spesifisitas, MCC, indeks Kappa Cohen dan kurva ROC. Setelah penjelasan evaluasi model, diakhiri dengan pembahasan hasil uji prediksi sentimen berdasarkan sampel komentar dari media sosial *corporate university*.

3.1. Hasil *Fine-Tuning* dan Komparasi IndoRoBERTa dan Varian IndoBERT

Berdasarkan hasil *fine-tuning* model, didapatkan nilai *F1-Score* dan akurasi tes dari model IndoRoBERTa dan varian IndoBERT, yang ditunjukkan pada Tabel 1.

Tabel 1. Hasil *Fine-Tuning* Model IndoRoBERTa dan Varian IndoBERT

Model	NEG	NEU	POS	<i>F1-Score</i>	Test Acc.
w11wo/indonesian-roberta-base-sentiment-classifier	0.95	0.93	0.98	0.952	0.962
indobenchmark/indobert-base-p1	0.90	0.84	0.95	0.899	0.925
indobenchmark/indobert-base-p2	0.90	0.86	0.95	0.902	0.925
indolem/indobert-base-uncased	0.89	0.83	0.95	0.889	0.918
indolem/indoberttweet-base-uncased	0.89	0.85	0.95	0.894	0.920

Berdasarkan Tabel 1, dilihat secara komparatif berbagai model untuk klasifikasi sentimen bahasa Indonesia, dan model IndoRoBERTa yang diusulkan yaitu *w11wo/indonesian-roberta-base-sentiment-classifier* menunjukkan performa terbaik di antara semua model. Dengan *F1-Score* sebesar 0.952 (95.2%) dan akurasi tes sebesar 0.962 (96.2%), model ini menunjukkan kemampuan yang sangat baik dalam mengklasifikasikan sentimen negatif (NEG), netral (NEU), dan positif (POS) dengan skor masing-masing sebesar 0.95, 0.93, dan 0.98. Performa ini menjadikannya pilihan unggul untuk tugas klasifikasi sentimen.

Varian model IndoNLU yang diusulkan adalah *indobenchmark/indobert-base-p1* dan *indobenchmark/indobert-base-p2* menunjukkan performa yang sangat mirip, dengan *F1-Score* masing-masing sebesar 0.899 (89.9%) dan 0.902 (90.2%), serta akurasi tes yang sama yaitu 0.925 (92.5%). Kedua model ini juga mampu mengklasifikasikan sentimen dengan baik, meskipun sedikit di bawah performa model IndoRoBERTa. Skor sentimen negatif dan positif untuk kedua model ini adalah sama, yaitu 0.90 dan 0.95, sementara sentimen netral sedikit berbeda dengan skor 0.84 untuk *indobert-base-p1* dan 0.86 untuk *indobert-base-p2*.

Varian model IndoLEM yang diusulkan yaitu *indolem/indobert-base-uncased* dan *indolem/indoberttweet-base-uncased* menunjukkan performa yang sedikit lebih rendah dibandingkan model-model sebelumnya. *Indoberttweet-base-uncased* memiliki *F1-Score* sebesar 0.894 (89.4%) dan akurasi tes sebesar 0.920 (92.0%), sementara *indobert-base-uncased* memiliki *F1-Score* sebesar 0.889 (88.9%) dan akurasi tes sebesar 0.918 (91.8%). Kedua model ini memiliki skor sentimen yang sama untuk sentimen negatif dan positif, yaitu masing-masing 0.89 dan 0.95, dengan perbedaan kecil pada sentimen netral.

Secara keseluruhan, model IndoRoBERTa yang paling unggul dalam tugas klasifikasi sentimen bahasa Indonesia, diikuti oleh varian IndoBERT dari IndoNLU dengan performa yang hampir setara. Varian IndoLEM sedikit tertinggal, namun tetap menunjukkan performa yang baik untuk tugas sentimen. Setelah diketahui bahwa model IndoRoBERTa lebih unggul dibandingkan varian IndoBERT lainnya, maka dilakukan evaluasi lebih lanjut terhadap model IndoRoBERTa.

3.2. Evaluasi Laporan Klasifikasi

Laporan klasifikasi menampilkan nilai *recall*, *precision* dan *F1-Score* untuk model. Laporan klasifikasi menyediakan rangkuman detail dari kinerja model untuk setiap kelas, yang lebih informatif dibandingkan hanya menggunakan metrik tunggal seperti akurasi (Géron, 2019). Adapun contoh fungsi perhitungan untuk laporan klasifikasi ditunjukkan pada Algoritma 1. Untuk hasil evaluasi laporan klasifikasi ditunjukkan pada Tabel 2.

Algoritma 1. Fungsi Perhitungan: Laporan Klasifikasi

```

Procedure CalculateMetrics(true_labels, predicted_labels):
// Menghitung jumlah kelas
total_classes = count(unique(true_labels))
// Inisialisasi dictionary untuk menyimpan metrik untuk setiap kelas
metrics = {}
// Inisialisasi variabel untuk menghitung total akurasi dan F1-Score
total_accuracy = 0
total_f1_score = 0
// Loop melalui setiap kelas
For each class in total_classes:
// Memilih sampel yang sesuai dengan kelas
true_samples = select samples where true_labels == class
predicted_samples = select samples where predicted_labels == class
// Menghitung positif aktual (TP), negatif aktual (TN), positif salah (FP) dan negatif salah (FN)
TP = count(true_samples where predicted_samples == class)
FP = count(predicted_samples where true_samples != class)
TN = count(true_samples where predicted_samples != class)
FN = count(predicted_samples where true_samples == class)
// Menghitung presisi (precision), recall, dan F1-Score untuk kelas tertentu
precision = TP / (TP + FP)
recall = TP / (TP + FN)
f1_score = 2 * (precision * recall) / (precision + recall)
// Menyimpan metrik untuk kelas tertentu
metrics[class] = {
  "Precision": precision,
  "Recall": recall,
  "F1-Score ": f1_score}
// Menambahkan kontribusi kelas terhadap akurasi dan F1-Score
total_accuracy += (TP + TN) / (TP + TN + FP + FN)
total_f1_score += f1_score
// Menghitung akurasi dan F1-Score
average_accuracy = total_accuracy / total_classes
average_f1_score = total_f1_score / total_classes
End Procedure

```

Tabel 2.Evaluasi Laporan Klasifikasi Sentimen Model IndoRoBERTa

Kelas	Laporan Klasifikasi				
	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Test Acc.</i>	<i>Avg. F1-Score</i>
NEGATIF	0.95	0.95	0.95	0.962	0.952
NETRAL	0.91	0.95	0.93		
POSITIF	0.98	0.97	0.98		

Berdasarkan Tabel 2, *precision*, *recall*, dan *F1-Score* yang tinggi menunjukkan bahwa model dapat melakukan tugas klasifikasi dengan baik di semua kelas, dengan kelas positif memiliki kinerja terbaik dan kelas netral memiliki kinerja paling rendah. Kemungkinan ini dapat terjadi karena ketidakseimbangan data pada masing-masing kelas, yang mana sebaran distribusi kelas positif memiliki jumlah data tertinggi dibandingkan kelas lainnya (Gambar 1). Secara keseluruhan, nilai akurasi tes adalah 0.962 (96.2%) dan *F1-Score* rata-rata adalah 0.952 (95.2%), menunjukkan kinerja yang sangat baik dari model klasifikasi sentimen. Namun, akurasi bukan merupakan faktor penentu satu-satunya untuk performa kinerja model. Oleh karena itu, diperlukan matriks lain seperti *confusion matrix* untuk membantu mengevaluasi performa model lebih lanjut, ditambah dengan kondisi data yang tidak seimbang (Kulkarni et al., 2020).

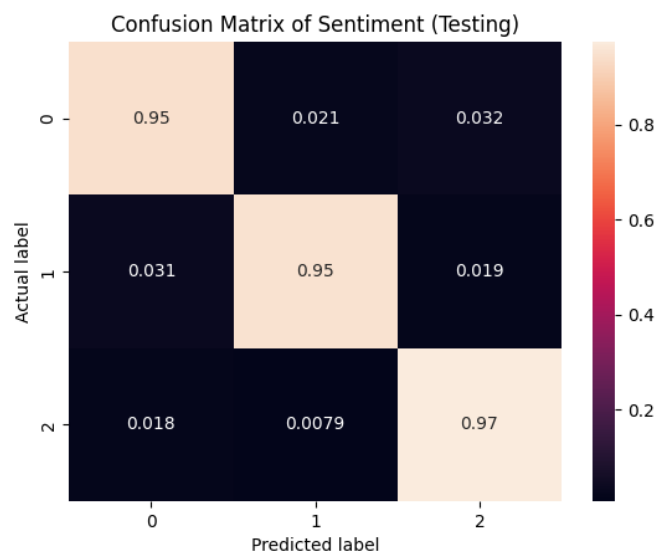
3.3. Evaluasi *Confusion Matrix*

Confusion matrix dapat digunakan untuk memecahkan tugas klasifikasi. Ini berlaku untuk klasifikasi biner dan klasifikasi multikelas. Kulkarni et al. (2020) menjelaskan bahwa *confusion matrix* mendeskripsikan total nilai prediksi beserta nilai aktual. Nilai aktual dan nilai prediksi menghasilkan 4 jenis hasil: positif aktual (TP), negatif aktual (TN), positif salah (FP) dan negatif salah (FN). Jumlah contoh positif yang diklasifikasikan secara aktual dan akurat ditunjukkan oleh TP, dan jumlah contoh negatif yang diklasifikasikan secara aktual dan akurat ditunjukkan oleh TN. Jumlah contoh negatif aktual yang diklasifikasikan sebagai positif ditunjukkan oleh FP, dan jumlah contoh positif aktual yang diklasifikasikan sebagai negatif ditunjukkan oleh FN.

Matriks ini memberikan gambaran tentang prediksi model dibandingkan dengan nilai sebenarnya, yang memungkinkan analisis lebih rinci terhadap kesalahan klasifikasi yang dilakukan oleh model (Chicco & Jurman, 2020; Géron, 2019). Contoh fungsi perhitungan *confusion matrix* dapat dilihat pada Algoritma 2. Hasil evaluasi *confusion matrix* sentimen ditunjukkan pada Gambar 4.

Algoritma 2. Fungsi Perhitungan: *Confusion Matrix*

```
Procedure ConfusionMatrix(true_labels, predicted_labels):
// Menghitung jumlah kelas
total_classes = count(unique(true_labels))
// Inisialisasi matriks kebingungan (confusion matrix)
confusion_matrix = matrix(total_classes, total_classes, filled_with_zeros)
// Loop melalui setiap pasangan true label dan predicted label
For i = 1 to length(true_labels):
    true_class = true_labels[i]
    predicted_class = predicted_labels[i]
// Memperbarui entri matriks kebingungan untuk pasangan kelas
confusion_matrix[true_class][predicted_class] += 1
// Menampilkan confusion matrix
DisplayMatrix(confusion_matrix)
End Procedure
```



Gambar 4. Evaluasi *Confusion Matrix* Sentimen Untuk Model IndoRoBERTa

Berdasarkan Gambar 4, diketahui untuk kelas 0 adalah negatif, kelas 1 adalah netral, dan kelas 2 adalah positif. Untuk kelas sentimen negatif, model ini berhasil mengklasifikasikan 0.95 (95%) dari contoh yang sebenarnya negatif dengan benar, meskipun ada 0.021 (2.1%) yang diklasifikasikan secara salah sebagai netral dan 0.032 (3.2%) sebagai positif. Untuk kelas sentimen netral, 0.95 (95%) dari contoh diklasifikasikan dengan benar, sementara 0.031 (3.1%) diklasifikasikan salah sebagai negatif dan 0.019 (1.9%) sebagai positif. Pada kelas sentimen positif, model juga menunjukkan performa tinggi dengan 0.97 (97%) dari contoh yang diklasifikasikan

dengan benar, meskipun ada 0.018 (1.8%) yang diklasifikasikan secara salah sebagai negatif dan 0.0079 (0.79%) sebagai netral.

Secara keseluruhan, matriks ini menunjukkan bahwa model memiliki akurasi tinggi dalam memprediksi kelas sentimen yang benar, dengan sebagian besar prediksi berada pada diagonal matriks, yang mencerminkan prediksi yang benar. Kesalahan klasifikasi terutama terjadi antara sentimen negatif dan positif, tetapi persentasenya tetap kecil, menunjukkan kemampuan model yang kuat dalam tugas klasifikasi sentimen.

3.4. Evaluasi Sensitivitas, Spesifisitas, MCC, dan Indeks Kappa Cohen

Tharwat (2020) menjelaskan, sensitivitas yang tinggi menunjukkan bahwa model mampu mendeteksi sebagian besar contoh dari setiap kelas dengan tepat (*True Positives Rate*), sedangkan spesifisitas adalah proporsi sebenarnya dari negatif yang diidentifikasi dengan benar oleh model (*True Negatives Rate*). Spesifisitas yang tinggi menunjukkan bahwa model mampu mengenali sebagian besar contoh yang bukan dari kelas tertentu, yang berarti model tidak memberikan banyak *False Positives* untuk kelas tersebut.

MCC (*Matthews Correlation Coefficient*) adalah ukuran kualitas klasifikasi yang mempertimbangkan positif aktual (TP), negatif aktual (TN), positif salah (FP) dan negatif salah (FN), yang dianggap sebagai metrik yang lebih andal daripada akurasi dalam situasi ketidakseimbangan kelas. MCC dapat digunakan dalam kasus multikelas. MCC memperhitungkan semua elemen dalam *confusion matrix* dan memberikan satu nilai yang mencerminkan performa keseluruhan model klasifikasi (Chicco & Jurman, 2020).

Indeks Kappa Cohen digunakan untuk mengukur tingkat kesepakatan antar penilai dalam klasifikasi kategori, dengan memperhitungkan kesepakatan yang terjadi secara kebetulan, sehingga memberikan gambaran yang lebih akurat tentang konsistensi klasifikasi dalam kasus multikelas (Warrens, 2015; Pérez et al., 2020). Untuk seluruh metrik evaluasi, nilai yang mendekati 1 dapat dikatakan bahwa model memiliki kinerja yang sangat baik. Algoritma 3 menunjukkan fungsi perhitungan sensitivitas, spesifisitas, MCC, dan indeks Kappa Cohen. Nilai metrik evaluasi yang berkaitan dengan kinerja model ditunjukkan dalam Tabel 3.

Algoritma 3. Fungsi Perhitungan: Sentivitas, Spesifisitas, MCC, dan Indeks Kappa Cohen

```

Procedure CalculateMetrics(true_labels, predicted_labels):
// Menghitung jumlah kelas
total_classes = count(unique(true_labels))
// Inisialisasi variabel untuk menyimpan jumlah TP, TN, FP, dan FN
TP = 0
TN = 0
FP = 0
FN = 0
// Loop melalui setiap pasangan true label dan predicted label
For i = 1 to length(true_labels):
    true_class = true_labels[i]
    predicted_class = predicted_labels[i]
// Menghitung jumlah TP, TN, FP, dan FN
    if true_class == predicted_class:
        if true_class == positive_class:
            TP += 1
        else:
            TN += 1
    else:
        if true_class == positive_class:
            FN += 1
        else:
            FP += 1
// Menghitung sensitivity dan specificity
sensitivity = TP / (TP + FN)
specificity = TN / (TN + FP)
// Menghitung MCC
MCC = (TP * TN - FP * FN) / sqrt((TP + FP) * (TP + FN) * (TN + FP) * (TN + FN))
// Menghitung indeks Kappa Cohen
po = (TP + TN) / (TP + TN + FP + FN)
pe = ((TP + FP) * (TP + FN) + (FP + TN) * (FN + TN)) / ((TP + TN + FP + FN) * (TP + TN + FP + FN))
kappa_cohen = (po - pe) / (1 - pe)
End Procedure

```

Tabel 3.Evaluasi Sentivitas, Spesifisitas, MCC, dan Indeks Kappa Cohen dari Model IndoRoBERTa

Sensitivitas	Spesifisitas	MCC	Indeks Kappa Cohen
0.9569	0.9784	0.9340	0.9339

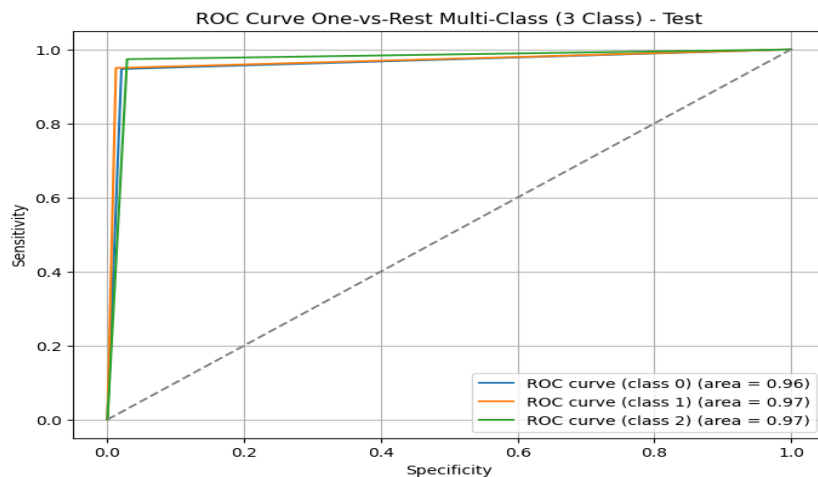
Berdasarkan Tabel 3, sensitivitas dikategorikan tinggi dengan nilai 0.9569, menunjukkan model sangat baik dalam mendeteksi contoh dari setiap kelas (negatif, netral, positif). Spesifisitas dikategorikan sangat tinggi dengan nilai 0.9784, menunjukkan model sangat baik dalam mengenali contoh yang bukan dari kelas tertentu (meminimalkan prediksi *False Positives*). MCC dikategorikan sangat tinggi dengan nilai 0.9340, menunjukkan korelasi yang sangat baik antara prediksi dan label sebenarnya dalam konteks multikelas. Indeks Kappa Cohen dikategorikan sangat tinggi dengan nilai 0.9339, menunjukkan kesepakatan yang baik antara prediksi model dan label sebenarnya dalam konteks multikelas. Secara keseluruhan, model memiliki performa yang sangat baik dalam klasifikasi multikelas untuk sentimen (negatif, netral, positif), dengan keseimbangan yang juga sangat baik antara mendeteksi contoh dari setiap kelas dan mengenali contoh yang bukan dari kelas tertentu, serta kesepakatan yang tinggi dengan label sebenarnya.

3.5. Evaluasi Kurva ROC

Receiver Operating Characteristic Curve atau disebut Kurva ROC, biasanya ditujukan untuk tugas klasifikasi kelas biner. Namun, saat ini dibuat pendekatan multikelas menggunakan Kurva ROC *One-vs-Rest* (OvR), yaitu teknik evaluasi kinerja model multikelas dengan mengubah masalah multikelas menjadi serangkaian masalah klasifikasi biner. ROC multikelas cocok digunakan pada dataset yang tidak seimbang, dengan spesifikasi bobot kesalahan klasifikasi dan interval *confidence* untuk mengakomodasi berbagai kriteria evaluasi pada masing-masing kelas (Wang & Carvalho, 2024). Kurva ROC terdiri dari grafik *True Positive Rate* (TPR) atau tingkat positif benar terhadap *False Positive Rate* (FPR) atau tingkat positif salah, untuk berbagai keputusan model klasifikasi. Kurva ini menunjukkan hubungan antara TPR atau sensitivitas dan FPR atau spesifisitas. Serupa dalam penelitian Zhang (2021) yang juga mengungkapkan, bahwa dalam masalah klasifikasi multikelas dapat diatasi dengan dengan cara membagi klasifikasi multikelas menjadi beberapa klasifikasi biner untuk memperoleh kurva ROC. Kurva ROC digunakan untuk menghitung skor AUC (*Area Under the Curve*). Nilai skor AUC berkisar antara 0 sampai dengan 1. Semakin nilai AUC mendekati 1, semakin baik modelnya. Contoh fungsi perhitungan kurva ROC dapat dilihat pada Algoritma 4, sedangkan hasil evaluasi dari kurva ROC ditunjukkan melalui Gambar 5.

Algoritma 4. Fungsi Perhitungan: Kurva ROC

```
Procedure KurvaROC-OvR
Input: X (matriks fitur), y (label kelas), classifier (pengklasifikasi biner)
//Langkah 1: Inisialisasi variabel yang diperlukan
kelas_unik = ["negatif", "netral", "positif"] // daftar kelas unik
kurva_roc = [] // daftar untuk menyimpan data ROC untuk setiap kelas
//Langkah 2: Loop untuk setiap kelas
for kelas in kelas_unik:
    //Buat label biner untuk kelas saat ini
    y_biner = [1 jika label == kelas else 0 untuk setiap label dalam y]
    //Latih pengklasifikasi biner untuk kelas saat ini
    classifier.fit(X, y_biner)
    //Dapatkan skor probabilitas untuk kelas positif
    y_skor = classifier.predict_proba(X)[:, 1]
    //Hitung true positive rates (TPR) dan false positive rates (FPR)
    fpr, tpr, _ = roc_curve(y_biner, y_skor)
    //Simpan data ROC untuk kelas saat ini
    kurva_roc.append((kelas, fpr, tpr))
//Langkah 3: Plot Kurva ROC untuk setiap kelas
//untuk setiap (kelas, fpr, tpr) dalam kurva_roc:
    plot(fpr, tpr, label=kelas)
Output: Kurva ROC untuk setiap kelas
End Procedure
```



Gambar 5. Evaluasi Kurva ROC Sentimen Untuk Model IndoRoBERTa

Kurva biru menunjukkan kelas negatif (kelas 0), dengan nilai AUC 0.96, menunjukkan performa yang sangat baik, artinya model mampu membedakan kelas negatif dari kelas lainnya dengan tingkat akurasi yang tinggi. Selanjutnya, kurva oranye menunjukkan kelas netral (kelas 1), dengan nilai AUC 0.97, menunjukkan performa yang sangat baik, berarti model mampu membedakan kelas netral dari kelas lainnya dengan tingkat akurasi yang tinggi. Terakhir, kurva hijau menunjukkan kelas positif (kelas 2), dengan nilai AUC 0.97, menunjukkan performa terbaik di antara ketiga kelas, artinya model mampu membedakan kelas positif dari kelas lainnya dengan tingkat akurasi yang sangat tinggi. Dari sisi akurasi klasifikasi, model menunjukkan performa yang sangat baik dalam membedakan setiap kelas dari kelas lainnya, dengan nilai AUC lebih dari 0.95 untuk ketiga kelas. Dari sisi keseimbangan kinerja, nilai AUC yang tinggi untuk tiap kelas, secara berurutan untuk negatif, netral, dan positif (0.96, 0.97, dan 0.97) menunjukkan bahwa model memiliki kinerja yang seimbang dan efektif dalam klasifikasi multikelas. Dari sisi keandalan model, AUC yang mendekati 1 untuk ketiga kelas mengindikasikan bahwa model sangat andal dalam klasifikasi sentimen (negatif, netral, positif), dan memiliki kemampuan diskriminatif yang tinggi.

3.6. Prediksi Sentimen Pada Sampel Komentar Media Sosial *Corporate University*

Berdasarkan hasil evaluasi kinerja model, secara keseluruhan model yang dilatih menggunakan IndoRoBERTa memberikan performa yang baik dalam tugas klasifikasi sentimen multikelas. Selanjutnya, dilakukan uji prediksi dari model. Sebelum uji prediksi, diambil 20 sampel acak berdasarkan komentar publik dari berbagai postingan media sosial Instagram, Twitter dan YouTube berkaitan dengan konten *corporate university* (Gambar 6), antara lain yaitu Badiklat BPK RI, BPPK dan Ditjen Perimbangan Keuangan dari Kementerian Keuangan, PPSDM Kementerian ATR/BPN, BPSDM Kementerian Hukum dan HAM, BPSDM Kementerian Perindustrian, P3K Bangkom ASN dan Pusbang Kader dari LAN RI, Kementerian Kesehatan, Pusdiklat Mahkamah Agung, BPSDM Provinsi DKI Jakarta, dan Badiklat Provinsi DI Yogyakarta.



Gambar 6. Berbagai Konten Media Sosial *Corporate University* Sebagai Sampel Uji Prediksi

Setelah itu, dilakukan uji prediksi berdasarkan sampel yang telah dikumpulkan menggunakan model yang telah dilatih untuk tugas sentimen. Untuk label prediksi terdiri atas negatif, netral, atau positif dan nilai prediksi berkisar antara 0 - 100%. Nama akun publik yang tampil di komentar, akan disamarkan menjadi '[username]' untuk menjaga privasi. Begitu juga nama subjek yang disebut dalam komentar publik dan terprediksi negatif akan disamarkan menjadi '[xyz]'. Hasil prediksi sentimen media sosial *corporate university* ditunjukkan di Tabel 4.

Tabel 4. Hasil Prediksi Sentimen Terhadap Komentar Dari Konten Media Sosial *Corporate University*

Komentar Media Sosial	Prediksi Sentimen	Nilai Prediksi
Keren..semoga bs kerja sama lagi dg badan diklat BPK..tempatny menyenangkan dan widyasarunya bagus dan benar2 bimbing kami sampai faham 🙏	Positif	99.57%
Proses diklat yang kreative dan inovative, semoga bisa memberikan manfaat dan hasil guna yang terbaik 🙌🙌	Positif	99.51%
Bisa ga mulai dr Security, front desk smp Karyawan [xyz] di training ulang? Lihat2 di Gmaps ulasan 90% pada ngeluh semua tentang pelayanan di kantor tersebut, tolong dong di tindaklanjuti, pelayan masyarakat loh masa kalah sama pelayanan bank [xyz], belajar dr sanalah pendapatan anda juga dr rakyat kok	Negatif	99.35%
Sekalian dibuat resolusi masalah hasil konsolidasi tanah, buat petunjuk penyelesaiannya dalam bentuk aturan/edaran	Netral	98.45%
Bapak Ir.Razilu lebih cocok sebagai Assesor sebenarnya karna lebih cenderung memotivasi mengingat back round nya lulusan IPB (Institut Pertanian Bogor)bukan dari Alumni AKIP/ AIM dan sarjana Hukum.,ya mudah mudahan pak Ir.Razilu bisa beradaptasi dengan zona di BPSDM Hukum dan HAM sebagai kepala Badan dan membawa Insan BPSDM Hukum dan HAM lebih baik lagi ,disiplin,berkarakter dan Maju.	Positif	99.46%
Oalah pantesan [xyz] yang ngadain inkubasi tekstil hanya jakarta, kasian bgt kami pelaku industri tekstil di daerah luar jakarta ga dapat kesempatan	Negatif	99.36%
Min minta tlg bantuan untuk verifikasi seminar di plataran sehat.. Saya ngga bisa ikut workshop lainnya 🙏	Netral	72.11%
[username] saya sudah 100% pembelajaran di status berjalan tidak pindah ke selesai sudah 5 hari ini menghambat pembelajaran selanjutnya untuk mendapat kan skp selanjutnya mohon tidak lanjutnya sudah email, wa, DM, tp belum ada respons	Negatif	97.12%
Min sertifikatnya agenda 1 dan agenda 2 kpn sdh bisa diunduh ya? Terimakasih sebelumnya 😊	Netral	98.54%
Terima kasih [username].lanri sudah mengobati kangen kami kepada Mama Kota meski secara virtual. Sampai jumpa di lain kesempatan... 🙏	Positif	99.17%
Kenapa kalau diklat [xyz] untuk urusan penggantian uang jalan peserta selalu lama ya? Padahal kalau diklat teknis begitu selesai langsung diganti tuh	Negatif	99.33%
Wuih, kerennn, Prof RG yg mengisi, No Rocky No Party.. 😄😄👏👏	Positif	99.52%
Maaf min, mohon direvisi link pendaftarannya karena link tidak valid	Netral	98.51%
Min.. saya sdh dm berkali2 kok ga ada respon ya .. nnt klo viral malah disalahin , ditanya ga ada respon .. 🙏	Negatif	99.36%
Min.. dokumen kami sudah bisa di upload di sipka akreditasi ?	Netral	98.43%
Baru Selesai... Sangat Bermanfaat... 🙌	Positif	99.53%
Bagaimana sinkronisasi kompetensi SDM Aparatur yang dihasilkan Corpu dgn lembaga profesi ybs. Apalagi organisasi seperti RS yang SDM Aparaturnya masing-masing memiliki organisasi profesi. Tks	Negatif	98.33%

Komentar Media Sosial	Prediksi Sentimen	Nilai Prediksi
maju terus Badan DIKLAT DIY, lancar dan sukses Corpu Pemda DIY untuk meningkatkan kompetensi ASN DIY dalam melayani dan meningkatkan kesejahteraan masyarakat. 🙌 🙌	Positif	99.46%
Mantap sekali inovasinya ... Semoga Badan Diklat DIY semakin maju dan jaya. Sukses selalu utk Bandiklat DIY.	Positif	99.55%
[username] SDN Tebet Timur 11 hadir, terima kasih untuk pencerahannya...Semoga dengan pembelajaran JakCorpu ini ASN tambah maju kedepannya selalu mendapat informasi kompetensi terbaru.	Positif	99.39%
Nilai Prediksi Rata-Rata		97.70%

Dari Tabel 4, hasil prediksi label sentimen keseluruhan dari sampel diperoleh masing-masing berjumlah 9 untuk positif, 6 untuk negatif, dan 5 untuk netral. Nilai prediksi yang tinggi menunjukkan bahwa model klasifikasi sentimen dapat diandalkan, dengan rata-rata nilai keseluruhan dari sampel prediksi adalah 97.70%. Berdasarkan interpretasi makna kalimat, komentar dengan prediksi positif cenderung mengekspresikan rasa puas, terima kasih, atau penghargaan terhadap layanan, acara, atau individu. Komentar positif juga berisi pujian atau dukungan. Komentar dengan prediksi negatif umumnya mengandung kritik, keluhan, atau ketidakpuasan. Publik sering menyoroti masalah atau kekurangan dalam layanan atau situasi tertentu. Komentar dengan prediksi netral biasanya bersifat informatif atau permintaan bantuan tanpa menunjukkan perasaan positif atau negatif yang kuat. Publik lebih fokus pada pertanyaan atau permintaan yang spesifik.

Komentar-komentar ini mencerminkan berbagai reaksi dan kebutuhan publik terhadap layanan atau situasi yang mereka hadapi dalam lingkup media sosial *corporate university*. Mengenai interpretasi makna kalimat, perlu diuji lebih lanjut terkait validitas prediksi, seperti melibatkan ahli bahasa, linguistik, atau ahli terkait lainnya, sehingga hasil prediksi dari model tersebut dapat dipertanggungjawabkan. Secara keseluruhan, prediksi sentimen ini membantu memahami bagaimana suatu organisasi pemerintahan dengan *corporate university* merespon layanan atau produk yang disediakan serta area mana yang memerlukan perhatian lebih lanjut, khususnya dalam layanan pendidikan dan pengembangan SDM.

4. Kesimpulan & Saran

4.1. Kesimpulan

Dari hasil penelitian menunjukkan bahwa model IndoRoBERTa berhasil mengklasifikasikan tugas sentimen multikelas dengan baik, dan mengungguli varian IndoBERT lainnya. Ditunjukkan dari nilai akurasi tes 96.2% dan nilai *F1-Score* 95.2% berdasarkan laporan klasifikasi. Selain itu, dilihat dari evaluasi *confusion matrix*, model IndoRoBERTa menunjukkan performa yang sangat baik dengan prediksi akurat untuk semua label, baik negatif, netral, maupun positif. Begitu pun dengan evaluasi sensitivitas, spesifisitas, MCC, dan indeks Kappa Cohen, IndoRoBERTa menunjukkan performa yang sangat baik dilihat dari keseimbangan dalam mendeteksi contoh dari setiap kelas, mengenali contoh yang bukan dari kelas tertentu, serta kesepakatan yang tinggi dengan label sebenarnya. Diakhiri dengan evaluasi kurva ROC, yang juga menunjukkan kinerja model sangat baik dari sisi akurasi klasifikasi, keseimbangan kinerja, dan keandalan model dengan nilai dari masing-masing kelas adalah 0.96 (negatif), 0.97 (netral), dan 0.97 (positif).

Dari uji prediksi terhadap sampel komentar dari media sosial *corporate university*, menunjukkan bahwa model IndoRoBERTa dapat diandalkan dalam memprediksi sentimen, dengan rata-rata nilai prediksi keseluruhan adalah 97.70%. Dengan hasil uji prediksi ini, dapat diinterpretasikan terhadap makna komentar dari masing-masing kelas sentimen. Sentimen positif berisi pujian atau dukungan, sentimen negatif mengandung kritik, keluhan, atau ketidakpuasan, dan sentimen netral lebih bersifat informatif atau permintaan bantuan tanpa menunjukkan perasaan positif atau negatif yang kuat. Namun, perlu validasi lebih lanjut mengenai hasil interpretasi konteks kalimat dari hasil prediksi sentimen dengan melibatkan ahli bahasa, linguistik atau keilmuan terkait lainnya, untuk menguji konteks kalimat dari hasil prediksi agar dapat dipertanggungjawabkan. Secara keseluruhan, hasil prediksi ini membantu memahami bagaimana suatu organisasi pemerintahan dengan sistem *corporate university* merespon

layanan dan produk, khususnya terkait pendidikan dan pelatihan SDM yang disediakan serta area mana yang memerlukan perhatian lebih lanjut.

4.2. Saran

Saran untuk penelitian selanjutnya adalah permasalahan ketidakseimbangan data yang perlu menjadi perhatian. Dengan menggunakan teknik *resampling*, seperti *oversampling*, *undersampling*, atau teknik dokumentasi lainnya, dalam mengatasi permasalahan tidak seimbangnya dataset sebagai upaya memaksimalkan performa model. Penelitian lanjutan juga bisa dilakukan dengan menggunakan dataset yang lebih besar dan beragam.

Selanjutnya, penelitian dapat berfokus pada investigasi mengenai interpretabilitas model untuk memahami bagaimana model membuat prediksi, dalam hal ini prediksi sentimen. Sebagai contoh ekstraksi fitur, yang mana fitur bisa berupa kata-kata atau frasa tertentu dalam teks. Teknik visualisasi seperti *heatmaps* atau *attention maps* juga dapat difungsikan untuk menunjukkan bagian mana dari teks yang diperhatikan oleh model saat membuat prediksi. Metode yang bisa diterapkan pada berbagai jenis model untuk menjelaskan prediksi seperti LIME (*Local Interpretable Model-agnostic Explanations*) dan SHAP (*SHapley Additive exPlanations*), yang menghasilkan penjelasan tentang bagaimana fitur individu berkontribusi pada prediksi, atau berbagai metode lainnya berkaitan dengan interpretabilitas model.

Ucapan Terimakasih

Penulis mengucapkan terima kasih kepada Asosiasi Profesi Widyaiswara Indonesia (APWI) karena telah menyelenggarakan secara terbuka pembuatan karya tulis ilmiah ini, sehingga membuka peluang kepada seluruh ASN dan asosiasi profesi jabatan fungsional untuk ikut berkarya dalam bentuk tulisan, yang kiranya memberi manfaat dan makna positif, khususnya di bidang teknologi kecerdasan artifisial. Tidak lupa diucapkan terima kasih kepada unsur pimpinan dan rekan-rekan dari Badan Kepegawaian Negara, yang telah mendukung pembuatan karya tulis ini.

Daftar Referensi

- Poria, S., Hussain, A., & Cambria, E. (2018). Multimodal Sentiment Analysis. In *Socio-Affective Computing* (Vol. 8). Cham: Springer International Publishing.
- Das, D. (2023). *Computational Intelligence Applications For Text And Sentiment Data Analysis*. New York: Academic Press.
- Gasper, K., Danube, C. L., & Hu, D. (2021). Making room for neutral affect: Evidence indicating that neutral affect is independent of and co-occurs with eight affective states. *Motivation and Emotion*, 45(1), 103-121. <https://doi.org/10.1007/s11031-020-09848-6>
- Muthasima, R., Sumpeno, S., & Suprpto, Y. K. (2019). Twitter Sentiment Analysis of Juvenile Behaviour Deviations using LSA (Latent Semantic Analysis). *Journal of Physics. Conference Series*, 1201(1), 012026. <https://doi.org/10.1088/1742-6596/1201/1/012026>
- Olaoye, F., & Potter, K. (2024). *Natural Language Processing and Sentiment Analysis*. <https://easychair.org/publications/preprint/tXbX>
- Statista. (2024, February 14). *Natural language processing market growth worldwide from 2021-2030*. <https://www.statista.com/forecasts/1449874/world-nlp-market-size-growth>
- Kemp, S. (2024). *Digital 2024: Indonesia. Data Reportal*. <https://datareportal.com/reports/digital-2024-indonesia> (diakses pada tanggal 25 Mei 2024).
- Setiawati, R., Sutisna, A., & Saefudin, Y. S. (2023). Analisis Systematic Literatur Review : *Corporate University Sebagai Strategi dalam Meningkatkan Kompetensi Pegawai*. *ejournal.undiksha.ac.id*. <https://doi.org/10.23887/jjpe.v15i2.69980>
- Etter, M., Colleoni, E., Illia, L., Meggiorin, K., & D'Eugenio, A. (2018). Measuring Organizational Legitimacy in Social Media: Assessing Citizens' Judgments With Sentiment Analysis. *Business & Society*, 57(1), 60-97. <https://doi.org/10.1177/0007650316683926>
- Anderson, A., Brown, B., Carter, C., Davis, D., & Evans, E. (2021). The Impact Of Social Media On Corporate Reputation. *Journal Of Social Media Research*, 12(3), 234-256. <https://doi.org/10.1234/jsmr.2021.34567>
- Kapeš, J., Keča, K., Fugošić, N., & Tanković, A. Č. (2022). Management Response Strategies To A Negative Online Review: Influence On Potential Guests' Trust. *Tourism and Hospitality Management*, 28(1), 1–27. <https://doi.org/10.20867/thm.28.1.1>

- 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/1907.11692>
- Rajapaksha, P., Farahbakhsh, R., & Crespi, N. (2021). BERT, XLNet or RoBERTA: the best transfer learning model to detect clickbaits. *IEEE Access*, 9, 154704–154716. <https://doi.org/10.1109/access.2021.3128742>
- Naseer, M., Asvial, M., & Sari, R. F. (2021). An Empirical Comparison of BERT, RoBERTa, and Electra for Fact Verification. 2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC). <https://doi.org/10.1109/icaaic51459.2021.9415192>
- Fransiscus, Girsang, A. S. (2022). Sentiment Analysis of COVID-19 Public Activity Restriction (PPKM) Impact using BERT Method. *International Journal of Engineering Trends and Technology*, 70(12), 281–288. <https://doi.org/10.14445/22315381/ijett-v70i12p226>
- Sandra, L., Marcel, G., Gunarso, G., Fredicia, & Riruma, O. W. (2022). Are University Students Independent: Twitter Sentiment Analysis of Independent Learning in Independent Campus Using RoBERTa Base IndoLEM Sentiment Classifier Model. In 2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE) (pp. 249-253). Jakarta, Indonesia. doi: 10.1109/ISMODE53584.2022.9743110
- Recker, J. (2021). Scientific research in information systems. In *Progress in IS*. <https://doi.org/10.1007/978-3-030-85436-2>
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020, November 2). *IndOLEM and IndoBERT: a benchmark dataset and pre-trained language model for Indonesian NLP*. arXiv.org. <https://arxiv.org/abs/2011.00677>
- Koto, F., Lau, J. H., & Baldwin, T. (2021, September 10). IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization. arXiv.org. <https://arxiv.org/abs/2109.04607>
- Felix, E. A., & Lee, S. P. (2019). Systematic literature review of preprocessing techniques for imbalanced data. *IET Software*, 13(6), 479–496. <https://doi.org/10.1049/iet-sen.2018.5193>
- Sarica, S., & Luo, J. (2021b). Stopwords in technical language processing. *PloS One*, 16(8), e0254937. <https://doi.org/10.1371/journal.pone.0254937>
- Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018, October 11). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv.org. <https://arxiv.org/abs/1810.04805>
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020, September 11). INDONLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. arXiv.org. <https://arxiv.org/abs/2009.05387>
- Sebastian, D., Purnomo, H. D., & Sembiring, I. (2022). BERT for Natural Language Processing in Bahasa Indonesia. *BERT for Natural Language Processing in Bahasa Indonesia*. <https://doi.org/10.1109/icitya57421.2022.10038230>
- Wongso, W. Indonesian Roberta Base Sentiment Classifier. Hugging Face. (2023b, February 3). <https://huggingface.co/w11wo/indonesian-roberta-base-sentiment-classifier>. doi: 10.57967/hf/0644.
- Kulkarni, A., Chong, D., & Batarseh, F. A. (2020). Foundations of data imbalance and solutions for a data democracy. Elsevier eBooks, 83–106. <https://doi.org/10.1016/b978-0-12-818366-3.00005-8>
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1), 6. <https://doi.org/10.1186/s12864-019-6413-7>
- Tharwat, A. (2020). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168–192. <https://doi.org/10.1016/j.aci.2018.08.003>
- Warrens, M. J. (2015). Five ways to look at Cohen's kappa. *Journal of Psychology & Psychotherapy*, 5(4), 197–202. <https://doi.org/10.4172/2161-0487.1000197>
- Pérez, J., Díaz, J., Garcia-Martin, J., & Tabuenca, B. (2020). Systematic literature reviews in software engineering—enhancement of the study selection process using Cohen's Kappa statistic. *Journal of Systems and Software*, 168, 110657. <https://doi.org/10.1016/j.jss.2020.110657>
- Wang, L., & Carvalho, L. (2024, April 19). Multiclass ROC. arXiv.org. <https://doi.org/10.48550/arXiv.2404.13147>
- Zhang, M. (2021). Comparing ROC curves on multiclass classification for predicting quality of wine. *Digital WPI*. <https://digital.wpi.edu/concern/etds/ws859j70j?locale=en>