Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 595−606
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.6051

Analisis Perbandingan Klasifikasi Intent Chatbot Menggunakan Deep Learning BERT, RoBERTa, dan IndoBERT

Aswin Dwiyono*, Abdiansah, Muhammad Fachrurrozi

Program Studi Magister Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya, Palembang
Jalan Palembang-Prabumulih, KM 32 Inderalaya, Kabupaten Ogan Ilir, Sumatera Selatan, Indonesia
Email: [1,*]aswindwiyono@gmail.com, [2]abdiansah@unsri.ac.id, [3]mfachrz@unsri.ac.id
Email Penulis Korespondensi: aswindwiyono@gmail.com
Submitted: **10/10/2024;** Accepted: **29/10/2024;** Published: **31/10/2024**

Abstrak−Chatbot adalah aplikasi percakapan yang dirancang untuk menangani masukan dari pengguna dan menghasilkan balasan yang sesuai berdasarkan masukan tersebut, yang kemudian dikomunikasikan kembali kepada pengguna. Agar dapat memberikan balasan yang akurat, chatbot harus dapat memahami maksud dari pengguna dengan benar. Salah satu permasalahan dalam pengembangan chatbot adalah bagaimana cara melakukan klasifikasi intent dari pengguna secara akurat. Kesalahan dalam memahami maksud pengguna dapat menyebabkan balasan yang tidak relevan. Untuk melakukan percakapan dengan pengguna, keinginan dari pengguna perlu diklasifikasikan dengan baik. Klasifikasi dilakukan untuk menentukan maksud dari teks yang dimasukkan pengguna agar sistem chatbot dapat memberikan jawaban yang sesuai. Penelitian ini membandingkan tiga model berbasis transformer yang canggih yaitu BERT (Bidirectional Encoder Representations from Transformers), RoBERTa (Robustly Optimized BERT Pretraining Approach), dan IndoBERT (Indonesia Bidirectional Encoder Representations from Transformer) untuk melakukan klasifikasi intent dalam sistem chatbot. Penelitian ini menggunakan dataset University Chatbot yang berisikan 38 intent dan telah diterjemahkan ke dalam Bahasa Indonesia. Berbagai metrik performa, termasuk akurasi, F1-score, presisi, dan recall, dianalisis untuk mengevaluasi dan menentukan model yang menghasilkan kinerja paling efektif dengan menggunakan parameter dan dataset yang sama. Pada akhir penelitian, model BERT mencapai akurasi 0,89, model RoBERTa mencapai akurasi 0,84, sedangkan model IndoBERT mencapai akurasi 0,94. Performa IndoBERT lebih baik dibandingkan dengan BERT dan RoBERTa disebabkan oleh pelatihan yang lebih spesifik untuk bahasa Indonesia, pretraining yang lebih relevan, dan adaptasi yang lebih efektif terhadap konteks dan struktur bahasa Indonesia.

Kata Kunci: Klasifikasi Intent; Chatbot; BERT; RoBERTa; IndoBERT; Transformer

Abstract−A chatbot is a software application to designed handle user inputs and generate appropriate replies based on those inputs, which are then communicated back to the user. In able to provide accurate responses, the chatbot must be able to understand the intent of the user accurately. An issue in the development of chatbots is how to accurate classify user intent. Incorrectly understanding user intent can result in irrelevant responses. In order to have a conversation with the user, the intent of the user needs to be classified correctly. This paper compares three state-of-the-art transformer-based models BERT (Bidirectional Encoder Representations from Transformers), RoBERTa (Robustly Optimized BERT Pretraining Approach), and IndoBERT (Indonesia Bidirectional Encoder Representations from Transformer) for the task of intent classification in chatbot systems. Various performance metrics, including accuracy, F1-score, precision, and recall, were analyzed to determine which model performs more effectively in the same parameter conditions. Performance metrics like accuracy and F1-score were compared to assess model BERT, RoBERTa and IndoBERT performs better in a University Chatbot Dataset in Indonesian language. The BERT model achieved an accuracy of 0.89, RoBERTa model achieved 0.84 and IndoBERT model achieved an accuracy of 0.94. The better performance of IndoBERT compared to BERT and RoBERTa is caused by more language-specific training, more relevant pretraining, and more effective adaptation to Indonesian context and structure.

Keywords: Intent Classification; Chatbot; BERT; RoBERTa; IndoBERT; Transformer

1. PENDAHULUAN

Chatbot adalah program yang mampu memproses masukkan dari pengguna dan menghasilkan tanggapan atas jawaban yang sesuai dengan masukan dari pengguna yang akan dikirim kembali ke pengguna [1]. Tujuan utama dari chatbot adalah untuk berkomunikasi dengan manusia dan dengan bantuan itu membuat banyak pekerjaan yang berlebihan menjadi lebih mudah bagi manusia [2]. Chatbot tidak dibatasi waktu dan karena dapat melayani pengguna kapan saja sehingga dapat membuat pekerjaan menjadi efisien. Salah satu elemen kunci dalam mengembangkan chatbot yang efektif adalah kemampuan chatbot untuk memahami dan mengklasifikasikan intent pengguna secara akurat. Untuk melakukan percakapan dengan pengguna, keinginan dari pengguna perlu diklasifikasikan dengan baik. Klasifikasi intent merupakan langkah awal yang penting dalam menentukan alur percakapan dan memahami konteksnya guna memberikan respons yang tepat dan relevan.

Ada beberapa metode dan penelitian sebelumnya dalam melakukan klasifikasi intent dalam sistem chatbot. Pendekatan konvensional sering mengandalkan sistem berbasis aturan, namun metode ini sering kali tidak memadai ketika dihadapkan pada konteks baru atau perubahan yang tidak terduga oleh aturan yang sudah ada. Akibatnya, sistem ini memerlukan pembaruan terus-menerus agar dapat beradaptasi dengan skenario baru [3], [4]. Selain itu, sistem berbasis aturan tidak fleksibel [5], membutuhkan biaya pengembangan dan pemeliharaan yang tinggi karena perlunya pembaruan secara manual [6], dan kurang tangguh dalam menangani variasi input pengguna, seringkali gagal mengenali maksud seperti kasus kesalahan ejaan, bahasa gaul, atau singkatan [7].

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 595−606
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.6051

Selain itu terdapat penelitian intent detection yang menggunakan machine learning tradisional seperti Naive Bayes, Adaboost, Support Vector Machine (SVM) dan Logistic Regression namun tidak dapat memahami informasi dalam teks secara mendalam sehingga sulit untuk mengetahui intensi pengguna dengan akurat [8]. Penelitian aplikasi chatbot berbasis teks menggunakan metode Naive Bayes Classifier dengan dataset kumpulan pertanyaan yang sering timbul (FAQ) pada perusahaan GRABADS menghasilkan nilai akurasi sebesar 93,33% dan nilai kesalahan sebesar 6,66% [2]. Penggunakan metode Naive Bayes Classifier dikarenakan komputasi yang lebih sederhana namun pada implementasi chatbot tidak dapat mengenali persamaan kata dari pertanyaan yang diajukan sehingga mengakibatkan chatbot memberikan jawaban yang salah [2]. Selanjutnya, terdapat penelitian pengembangan chatbot menggunakan metode Natural Language Processing berbasis dialogflow yang menghasilkan akurasi sebesar 92.3% [9]. Metode dialogflow ini mampu membuat komputer mampu mengerti bahasa manusia dalam bentuk teks namun tidak dapat memproses pertanyaan yang kompleks sehingga perlu dilakukan pembaruan secara berkala pada data training phrases chatbot yang bertujuan untuk meningkatkan pemahaman pada sistem chatbot [9]. Kemudian terdapat juga penelitian deteksi intensi chatbot berbahasa Indonesia dari percakapan pelanggan PT. Kazee dengan menggunakan Metode Capsule Network atau CapsNet yang menghasilkan nilai akurasi 70% [10]. CapsNet adalah sebuah model yang terdiri dari beberapa kapsul yang terbentuk dari kumpulan neuron yang akan memproses informasi yang diberikan sehingga dapat diketahui intensi dari sebuah teks berdasarkan perhitungan yang telah dilakukan [10]. Hasil model CapsNet 6 intensi sudah cukup baik namun mengalami penurunan nilai akurasi untuk 18 intensi dan jika metode ini digunakan dalam chatbot maka chabot mampu merespon pengguna, akan tetapi respon yang diberikan masih banyak yang tidak sesuai [10].

Untuk mengatasi tantangan ini, maka muncul teknik deep learning, terutama model berbasis transformer seperti BERT (Bidirectional Encoder Representations from Transformers), RoBERTa (Robustly Optimized BERT Pretraining Approach) dan IndoBERT (Indonesia Bidirectional Encoder Representations from Transformer). Model-model ini menggunakan pembelajaran mendalam untuk menangkap hubungan semantik dalam teks, meningkatkan kemampuan chatbot dalam memahami intent pengguna. BERT telah diadopsi secara luas untuk klasifikasi intent karena kemampuannya untuk memahami konteks dengan memproses teks dua arah, menawarkan akurasi yang tinggi dalam tugas klasifikasi intent [11], [12], [13]. Namun, meskipun BERT memberikan kemampuan yang canggih, BERT bukannya tanpa keterbatasan, seperti bias yang dapat mendukung kelas tertentu di atas yang lain [14], dan masalah yang terkait dengan sub-tokenisasi, yang dapat mengakibatkan ketidakselarasan yang mempengaruhi keakuratan prediksi label [15].

Meskipun BERT efektif, pengembangan RoBERTa oleh Facebook AI menunjukkan peningkatan signifikan dalam prapelatihan dan pemrosesan data BERT. Tidak seperti BERT, RoBERTa menghilangkan tujuan Prediksi Kalimat Berikutnya (NSP), menyederhanakan proses pelatihan dan memungkinkan model untuk lebih fokus pada pemahaman konteks dalam kalimat tunggal, yang mengarah pada peningkatan akurasi dalam klasifikasi intent [16], [17]. Arsitektur RoBERTa juga memungkinkannya untuk memahami pola dan hubungan yang detail dalam data, yang sangat penting untuk pengenalan intent yang efektif, terutama dalam dataset multibahasa[18]. Namun dalam implementasi menggunakan bahasa Indonesia, munculnya IndoBERT telah menjadi terobosan yang lebih signifikan. IndoBERT adalah varian dari BERT yang secara khusus dilatih pada korpus bahasa Indonesia yang dikumpulkan dari sumber-sumber umum seperti media sosial teks, blog, berita, dan situs web sehingga mampu menangkap nuansa bahasa yang unik, termasuk struktur morfologis dan sintaksis yang khas [19].

Dalam membandingkan kinerja BERT, RoBERTa dan IndoBERT untuk klasifikasi intent chatbot, ketiga model tersebut memiliki kekuatan yang unik. BERT, dengan pendekatan dua arah dan pemodelan kontekstual yang baik, memberikan hasil yang memuaskan dalam tugas-tugas bahasa alami. RoBERTa, dengan proses prapelatihan yang telah dioptimalkan, membuktikan bahwa pengembangan lebih lanjut dapat meningkatkan akurasi dan efisiensi klasifikasi intent. Sementara itu, IndoBERT yang dikembangkan secara khusus untuk bahasa Indonesia menawarkan kemampuan pemodelan yang lebih relevan dengan konteks linguistik lokal. Dengan menggunakan IndoBERT, diharapkan dapat mengetahui performa model dalam memahami nuansa bahasa Indonesia terutama dalam tugas-tugas klasifikasi intent yang melibatkan percakapan alami.

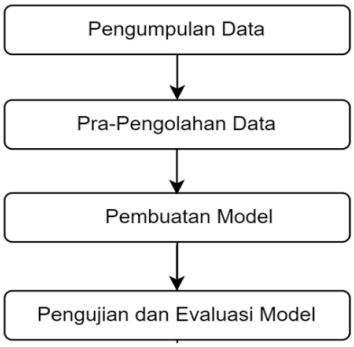
Berdasarkan penelitian-penelitian yang telah dilakukan, klasifikasi intent chatbot sangat penting dilakukan dalam menentukan alur percakapan dan memahami konteksnya guna memberikan respon yang tepat dan relevan. Penelitian ini bertujuan untuk menganalisis kinerja BERT, RoBERTa dan IndoBERT dalam melakukan klasifikasi intent chatbot menggunakan dataset dalam bahasa Indonesia. Hasilnya diharapkan dapat memberikan kontribusi wawasan yang bermanfaat dalam penerapan model berbasis deep learning untuk sistem percakapan otomatis atau chatbot yang lebih cerdas dan akurat.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Bagian ini menjelaskan metode yang dilakukan pada penelitian ini. Metode ini mempunyai beberapa tahapan, yaitu Pengumpulan Data, Pra-Pengolahan Data, Pembuatan Model dengan menggunakan BERT, RoBERTa, dan IndoBERT, kemudian dilakukan Pengujian dan Evaluasi Model. Berikut tahapan dari alur penelitian yang dapat dilihat pada Gambar 1.

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 595−606
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.6051



Gambar 1. Tahapan Penelitian

Uraian tahapan penelitian yang digambarkan pada Gambar 1 dapat dijelaskan sebagai berikut:

1. Pengumpulan Data
   Pada tahapan ini akan dilakukan pengumpulan data sekunder berupa dataset dari Kaggle yaitu University Chatbot Dataset [20]. Dataset berupa file JSON yang nantinya akan dikelompokkan berdasarkan intent. Dataset ini kemudian diterjemahkan ke dalam Bahasa Indonesia.
2. Pra-Pengolahan Data
   Data yang sudah dipersiapkan selanjutnya akan melalui tahapan pra-pengolahan data. Proses pra-pengolahan data mencakup tokenisasi teks, menghapus tanda baca, mengatasi kata-kata yang salah eja, dan langkah-langkah lainnya untuk membersihkan data. Ini bertujuan agar data tersebut lebih mudah untuk diproses dan meningkatkan tingkat keberhasilan klasifikasi.
3. Pembuatan Model BERT, RoBERTa dan IndoBERT
   Lalu proses selanjutnya adalah melakukan membuat model deep learning dengan menggunakan metode BERT, RoBERTa dan IndoBERT.
4. Pengujian dan Evaluasi Model
   Data yang telah melalui tahapan proses klasifikasi selanjutnya akan dihitung dengan menggunakan confusion matrix. Dari hasil kedua perhitungan tersebut akan didapatkan nilai akurasi, recall, presisi, dan F1-Score. Dimana nilai-nilai tersebut akan menjadi acuan untuk mengukur performa yang dihasilkan dari model yang digunakan yaitu BERT, RoBERTa, dan IndoBERT.

2.2 Dataset

Data yang dipakai untuk penelitian ini bersumber dari Kaggle yang membahas mengenai University Chatbot Dataset [20]. Dataset berupa file JSON yang berisikan teks kalimat-kalimat percakapan dan label intent yang sesuai. Dalam konteks chatbot, dataset intent adalah kumpulan data yang terdiri dari pasangan input (kalimat pengguna) dan output (intent atau maksud pengguna). Intent ini merepresentasikan maksud atau aksi yang ingin dicapai pengguna saat berinteraksi dengan chatbot. Dataset ini berisi 38 intent, yang juga disebut sebagai tag. Dataset ini dapat digunakan untuk melatih dan mengevaluasi model chatbot. Tabel 1 menunjukkan contoh dataset intent dalam bahasa Inggris.

Tabel 1. Dataset Chatbot dalam Bahasa Inggris

Tag	Pattern	Response
Fees	23	1
Location	14	1
Document	13	1

Selanjutnya, dataset dalam penelitian ini diterjemahkan ke dalam Bahasa Indonesia menggunakan DeepL [21] dan dengan bantuan manusia. Selain itu, terdapat beberapa isi pattern yang menggunakan bahasa Inggris namun tidak dapat diterjemahkan langsung melalui DeepL. Oleh karena itu, pola tersebut dihapus atau disesuaikan ke dalam Bahasa Indonesia dengan bantuan manusia, misalnya, 'ttyl' diterjemahkan menjadi 'berbicara dengan Anda nanti,' dan 'gtg' menjadi 'aku harus pergi.'. Contoh hasil terjemahannya dapat ditunjukkan pada Tabel 2.

Tabel 2. Dataset Chatbot dalam Bahasa Indonesia

Tag	Pattern	Response
Biaya	23	1
Lokasi	14	1
Dokumen	13	1

Tabel 2 menunjukan dataset chatbot dalam bahasa Indonesia yang terdiri dari tiga kolom yaitu tag, pattern, dan response. Tag adalah kategori atau topik yang terkait dengan maksud atau intent pengguna. Pattern merupakan jumlah pola untuk setiap intent dan response adalah jumlah respon untuk setiap intent.

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 595−606
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.6051

2.3 BERT, RoBERTa, dan IndoBERT

Bidirectional Encoder Representations from Transformers (BERT) merupakan teknik pembelajaran mesin yang dikembangkan ole Google berbasis Transformer untuk pra-pelatihan pemrosesan bahasa alami (Natural Language Processing). BERT merupakan model representasi kata kontekstual yang dilatih sebelumnya berdasarkan MLM (Masked Language Model), menggunakan transformers dua arah. Arsitektur model BERT adalah struktur encoder-decorder transformator dua arah multi-layer. Transformer mengikuti keseluruhan arsitektur ini menggunakan stacked self-attention dan point-wise, terhubung sepenuhnya untuk encoder dan decoder [22]. BERT membuat representasi kontekstual dari masukan teks dengan menggunakan encoder. Proses dimulai dengan tokenisasi teks, yang membentuk token kecil seperti kata, sub-kata, atau karakter. Selanjutnya, proses embedding digunakan untuk mengubah tiap token menjadi vektor kata. Proses ini menggunakan embedding yang telah dilatih sebelumnya oleh model. Kemudian, encoder BERT, yang terdiri dari berbagai layer transformer, bekerja pada token-token tersebut. Untuk membuat representasi kontekstual, setiap lapisan melakukan operasi multi-head self-attention dan full-connected feedforward networks. BERT unik karena dapat memahami konteks secara bidireksional dan menangkap hubungan kontekstual yang lebih kuat dengan memperhatikan token kiri dan kanan. Proses ini berulang pada setiap layer encoder yang ditumpuk, yang menghasilkan pemahaman teks input yang semakin hierarkis [23].

Robustly Optimized BERT Pretraining Approach (RoBERTa) merupakan modifikasi dari arsitektur BERT yang telah dikembangkan dengan total jumlah dataset sebesar 160GB berupa English-language corpora, sehingga dapat cocok atau melebihi kinerja semua metode post-BERT [16]. RoBERTa memiliki dua model pra-pelatihan yang disebut RoBERTa large dan RoBERTa base. Analisis dilakukan dengan menerapkan proses pre-training dan fine-tuning. Pre-training bertujuan untuk mendapatkan model RoBERTa yang sudah terlatih. Sebelum melakukan training, hal yang perlu dilakukan adalah melakukan konfigurasi terlebih dahulu dengan memasukkan hyperparameter yang digunakan, token yang didapatkan, dan data training yang telah dibuat. Fine-tuning didapatkan dari proses pre-training sebelumnya dan model yang sudah terlatih digunakan untuk melakukan proses masking. Setelah melakukan pra-pengolahan data yang bertujuan untuk mendapatkan data untuk training, testing, dan validasi. Kemudian dilakukan konfigurasi dengan menggunakan hyperparameter yang digunakan untuk proses fine-tuning [24].

IndoBERT adalah model berbasis BERT yang telah dilatih terlebih dahulu dengan menggunakan korpus besar bahasa Indonesia (korpus Indo4B) yang dikumpulkan dari sumber-sumber umum seperti media sosial teks, blog, berita, dan situs web dan dilatih menggunakan framework Huggingface yang mengikuti konfigurasi default BERT-Base (uncased). [19], [25]. IndoBERT memiliki dua varian arsitektur: IndoBERT-base dan IndoBERT-large. Keduanya berbeda dalam jumlah layer transformer, attention heads, dan parameter. IndoBERT-base memiliki 12 layer transformer, 12 attention heads, dan 110 juta parameter. IndoBERT-larget memiliki 24 layer transformer, 16 attention heads, dan 340 juta parameter. Dibandingkan IndoBERT-base, IndoBERT-large umumnya memberikan hasil presisi yang lebih tinggi namun membutuhkan waktu yang lebih lama dalam pelatihan data dibandingkan IndoBERT-base [26].

2.4 Pengaturan Eksperimen

Konfigurasi parameter model merupakan langkah penting dalam menentukan performa model BERT, RoBERTa, dan IndoBERT. Setiap model memiliki struktur dan karakteristik yang berbeda, sehingga pengaturan parameter perlu disesuaikan untuk mencapai performa terbaik pada tugas tertentu. Melalui penyesuaian konfigurasi ini, setiap model diharapkan dapat menunjukkan performa terbaiknya, baik dalam memahami konteks data maupun dalam akurasi prediksi.

Tabel 3. Parameter yang digunakan pada eksperimen

Parameter	Nilai
Num_train_epoch	100
Per_device_train_batch_size	32
Per_device_eval_batch_size	16
Warmup_steps	100
Weight_decay	0.05
Logging_steps	50
Evaluation_strategy	step
Eval_steps	50

Tabel 3 menampilkan berbagai konfigurasi parameter model yang digunakan dalam eksperimen ini yaitu pada model BERT, RoBERTa dan IndoBERT. Tabel ini terdiri dari dua kolom: (1) Parameter, yang berisi parameter yang digunakan dalam eksperimen; dan (2) Nilai, yang merupakan nilai parameter yang digunakan dalam eksperimen. Num_train_epoch mewakili jumlah epoch yang menunjukkan berapa kali model akan memproses seluruh dataset selama pelatihan. Dalam hal ini, model akan melalui dataset sebanyak 100 kali selama proses pelatihan. Per_device_train_batch_size mengacu pada ukuran batch untuk pelatihan yang menunjukkan

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 595−606
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.6051

jumlah sampel yang diproses model dalam satu iterasi pada setiap perangkat. Pada eksperimen ini model akan memproses 32 sampel sekaligus per iterasi pelatihan. Per_device_eval_batch_size mengacu pada ukuran batch yang digunakan selama evaluasi ketika model memvalidasi kinerjanya pada data validasi. Nilai 16 berarti 16 sampel dievaluasi dalam satu waktu. Warmup_steps mengacu pada jumlah langkah 'pemanasan' selama pelatihan. Langkah-langkah ini membantu mencegah perubahan mendadak pada awal pelatihan, yang dapat menyulitkan model untuk belajar secara efektif. Weight_decay adalah parameter regularisasi yang membantu mencegah overfitting dengan secara bertahap mengurangi nilai bobot dari waktu ke waktu. Nilai 0.05 mengindikasikan bahwa bobot akan dikurangi sebesar 5% pada setiap pembaruan untuk mencegah bobot menjadi terlalu besar. Logging_steps menentukan frekuensi pencatatan selama pelatihan. Setiap 50 langkah pelatihan, proses akan menghasilkan log, memberikan informasi seperti kerugian atau akurasi saat ini. Evaluation_strategy mendefinisikan strategi yang digunakan untuk evaluasi selama pelatihan. Ketika diatur ke 'step', model akan dievaluasi setiap beberapa langkah, bukan hanya pada akhir setiap epoch. Eval_steps menentukan seberapa sering evaluasi dilakukan selama pelatihan. Ketika diatur ke 50, evaluasi akan dijalankan setiap 50 langkah pelatihan.

2.5 Metrik Evaluasi

Dalam mengevaluasi performa klasifikasi model, kami memilih empat metrik yang umum digunakan dalam tugas klasifikasi: Precision (P), Recall (R), F1-score (F1), dan Accuracy (Acc). Nilai yang lebih tinggi menunjukkan kinerja klasifikasi yang lebih baik. Perhitungannya ditunjukkan pada persamaan (1)-(4). TP mewakili jumlah sampel yang diprediksi dengan benar, FP mewakili jumlah sampel yang diprediksi dengan salah, FN adalah jumlah sampel yang salah diklasifikasikan ke dalam kategori lain, dan TN adalah jumlah sampel yang diklasifikasikan dengan benar ke dalam kategori lain.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1-Score} = \frac{2 * (\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}} \quad (4)$$

3. HASIL DAN PEMBAHASAN

3.1 Dataset Hasil Terjemahan Bahasa Indonesia

Hasil dataset penelitian yang telah diterjemahkan ke dalam bahasa Indonesia ini terdiri dari 38 jumlah intent, 405 jumlah pola intent dan 47 jumlah total response. Adapun dataset intent yang akan digunakan pada penelitian ini dapat dilihat pada Tabel 4 berikut:

Tabel 4. Dataset Intent yang diterjemahkan ke dalam Bahasa Indonesia

Intent	Count Pattern	Count Response
Salam	10	3
Perpisahan	12	4
Pencipta	16	1
Nama	13	4
Jam	17	1
Nomor	15	1
Jurusan	27	1
Biaya	23	1
Lokasi	14	1
Asrama	22	1
Acara	11	1
Dokumen	13	1
Lantai	7	1
Silabus	7	1
Perpustakaan	14	1
Infrastruktur	3	1
Kantin	11	1
Menu	7	1
Karir	9	1
Kepala Jurusan	4	1
Kepala Prodi	4	1

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 595−606
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.6051

Intent	Count Pattern	Count Response
Sekretaris Jurusan	4	1
Rektor	7	1
Semester	11	1
Penerimaan	6	1
Beasiswa	26	1
Fasilitas	5	1
PMB	9	1
Seragam	9	1
Komite	6	1
Acak	3	1
Umpat	9	1
Liburan	12	1
Olahraga	7	1
Salut	13	1
Tugas	6	2
Pelonco	10	1
Dekan	3	1

Tabel 4 ini terdiri dari tiga kolom: (1) Intent, yang berisi 38 kategori intent yang berhubungan dengan chatbot universitas; (2) Count Pattern, yang menunjukkan jumlah pola untuk setiap intent; dan (3) Count Response, yang menunjukkan jumlah respon untuk setiap intent. Selanjutnya, sampel data pada dataset University Chatbot yang telah diterjemahkan dalam bahasa Indonesia dan digunakan pada penelitian dapat dilihat pada Gambar 2 berikut:

Gambar 2. Sampel Dataset

Gambar 2 menunjukan contoh sampel dataset yang digunakan pada penelitian ini. Dataset ini terdiri dari berbagai pola query dan response yang disimpan pada JSON dan digunakan untuk interaksi antara pengguna dan sistem chatbot. Tag adalah kategori yang terkait dengan maksud atau intent pengguna. Patterns adalah jumlah pola atau variasi pertanyaan yang terkait dengan setiap tag sedangkan responses adalah jumlah respon yang tersedia untuk setiap tag. Dataset ini membantu sistem chatbot dalam memahami jenis pertanyaan pengguna dan memberikan jawaban yang relevan sesuai tag yang dikenali.

3.2 Hasil Model BERT

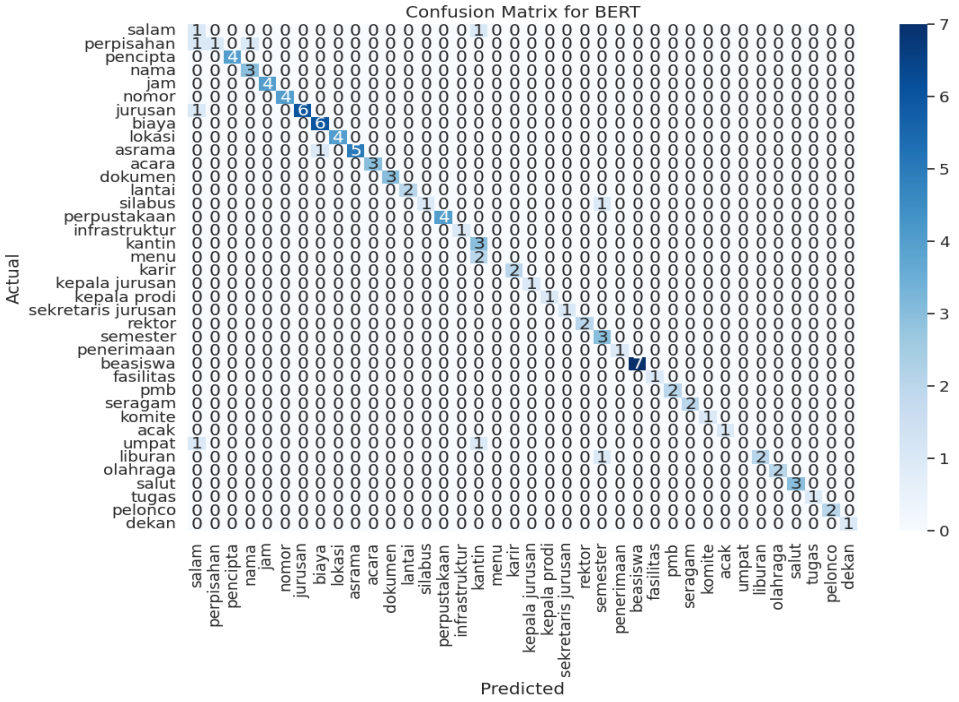
Setelah pelatihan, model dievaluasi dengan menggunakan dataset uji. Metrik evaluasi meliputi Accuracy (Acc.), F1-score (F1), Precision (Prec.) dan Recall. Hasil pelatihan dan pengujian model BERT ditunjukkan pada Tabel 5.

Tabel 5. Hasil Pelatihan dan Pengujian Model BERT

	Loss	Acc.	F1	Prec.	Recall
Train	0.015	0.997	0.998	0.998	0.999
Test	0.604	0.892	0.875	0.892	0.886

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 595−606
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.6051

Seperti yang ditunjukkan pada Tabel 5, model mencapai akurasi yang tinggi pada data pelatihan (0.99), tetapi akurasi pada data pengujian lebih rendah (0.89). Hal ini mengindikasikan adanya sedikit overfitting, artinya model menghafal terlalu banyak pola pada data pelatihan sehingga kurang mampu menggeneralisasi data baru. Demikian pula, F1-score didapatkan sebesar 0.99 pada data pelatihan tetapi turun menjadi 0.87 pada data pengujian.



Gambar 3. Hasil Confusion Matrix dari Model BERT

Gambar 3 menunjukkan confusion matrix dari model BERT yang diterapkan pada klasifikasi intent chatbot. Sumbu X mewakili kelas yang diprediksi, sedangkan sumbu Y mewakili kelas yang sebenarnya. Nilai di setiap sel menunjukkan jumlah sampel yang diklasifikasikan ke dalam kelas tertentu. Dari Gambar 3, terlihat bahwa model BERT memiliki performa yang baik dalam mengklasifikasikan intent asrama dan beasiswa. Namun, model ini sering salah mengklasifikasikan intent kantin sebagai menu, yang mengindikasikan bahwa model ini kesulitan untuk membedakan kedua intent tersebut.

	precision	recall	f1-score	support
salam	0.25	0.50	0.33	2
perpisahan	1.00	0.33	0.50	3
pencipta	1.00	1.00	1.00	4
nama	0.75	1.00	0.86	3
jam	1.00	1.00	1.00	4
nomor	1.00	1.00	1.00	4
jurusan	1.00	0.86	0.92	7
biaya	0.86	1.00	0.92	6
lokasi	1.00	1.00	1.00	4
asrama	1.00	0.83	0.91	6
acara	1.00	1.00	1.00	3
dokumen	1.00	1.00	1.00	3
lantai	1.00	1.00	1.00	2
silabus	1.00	0.50	0.67	2
perpustakaan	1.00	1.00	1.00	4
infrastruktur	1.00	1.00	1.00	1
kantin	0.43	1.00	0.60	3
menu	0.00	0.00	0.00	2
karir	1.00	1.00	1.00	2
kepala jurusan	1.00	1.00	1.00	1
kepala prodi	1.00	1.00	1.00	1
sekretaris jurusan	1.00	1.00	1.00	1
rektor	1.00	1.00	1.00	2
semester	0.60	1.00	0.75	3
penerimaan	1.00	1.00	1.00	1
beasiswa	1.00	1.00	1.00	7
fasilitas	1.00	1.00	1.00	1
pmb	1.00	1.00	1.00	2
seragam	1.00	1.00	1.00	2
komite	1.00	1.00	1.00	1
acak	1.00	1.00	1.00	1
umpat	0.00	0.00	0.00	2
liburan	1.00	0.67	0.80	3
olahraga	1.00	1.00	1.00	2
salut	1.00	1.00	1.00	3
tugas	1.00	1.00	1.00	1
pelonco	1.00	1.00	1.00	2
dekan	1.00	1.00	1.00	1
accuracy			0.89	102
macro avg	0.89	0.89	0.88	102
weighted avg	0.90	0.89	0.88	102

Gambar 4. Pengukuran Metrik Setiap Intent Model BERT

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 595−606
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.6051

Selanjutnya, kita dapat menghitung metrik evaluasi seperti accuracy, precision, recall, dan F1-score untuk setiap intent. Metrik-metrik ini memberikan gambaran yang lebih kuantitatif tentang kinerja model yang dapat dilihat pada Gambar 4. Sebagian besar intent memiliki precision, recall, dan F1-score = 1.00, hal ini mengindikasikan klasifikasi yang sangat baik untuk kategori-kategori tersebut, seperti pencipta, jam, nomor, lokasi, acara, dokumen, lantai, perpustakaan, infrastruktur, karir, kepala jurusan, dan kepala prodi, kepala prodi, sekretaris jurusan, rektor, penerimaan, beasiswa, fasilitas, PMB, seragam, komite, acak, olahraga, salut, tugas, pelonco, dekan. Hasil eksperimen untuk mengevaluasi model klasifikasi intent chatbot menggunakan metode BERT menunjukkan akurasi sebesar 0.89, seperti yang ditunjukkan pada Gambar 4.

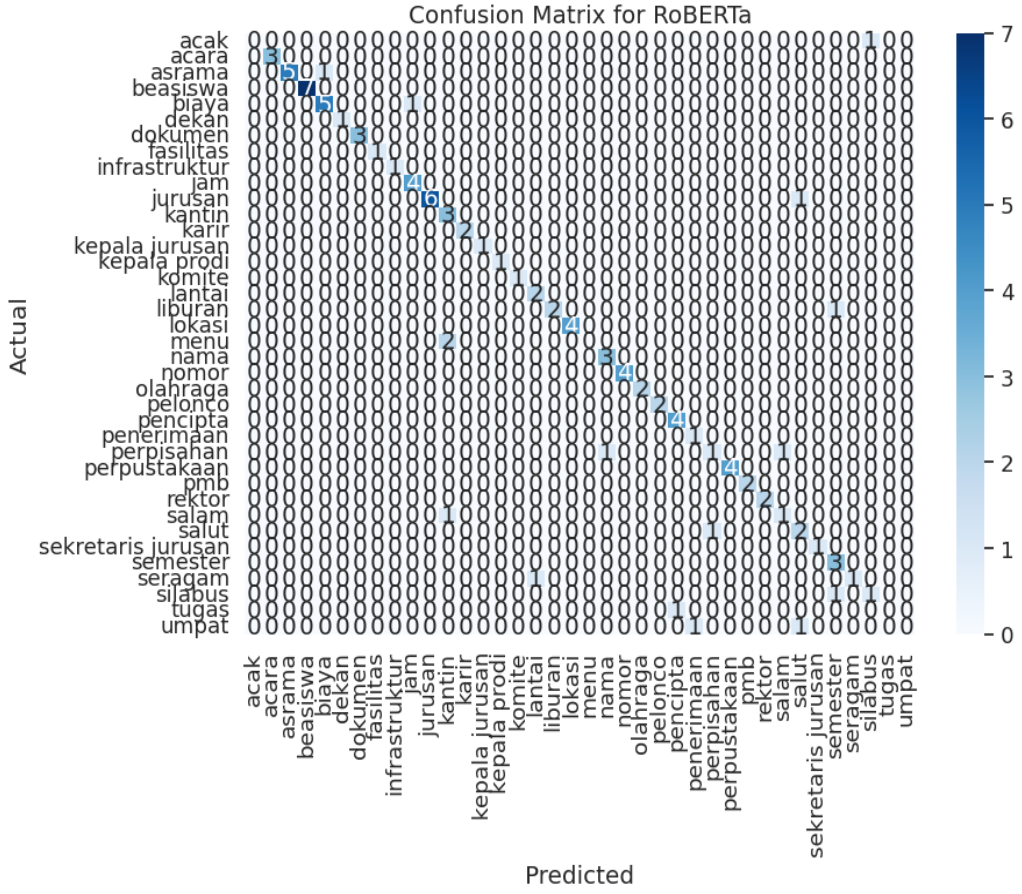
3.3 Hasil Model RoBERTa

Setelah pelatihan, model dievaluasi dengan menggunakan dataset uji. Metrik evaluasi meliputi Accuracy (Acc.), F1-score (F1), Precision (Prec.) dan Recall. Hasil dari pelatihan dan pengujian model RoBERTa ditunjukkan pada Tabel 6.

Tabel 6. Hasil Pelatihan dan Pengujian Model RoBERTa

	Loss	Acc.	F1	Prec.	Recall
Train	0.009	0.997	0.998	0.999	0.998
Test	0.765	0.843	0.779	0.775	0.808

Dapat dilihat pada Tabel 6 bahwa model mampu mencapai akurasi yang tinggi pada data training (0.99), namun akurasi pada data testing lebih rendah (0.84). Hal ini mengindikasikan adanya sedikit overfitting, artinya model menghafal terlalu banyak pola pada data training sehingga tidak mampu menggeneralisasi pada data yang baru. Selain itu, nilai F1-Score mendapatkan nilai 0.99 pada data training sedangkan pada data testing mendapatkan nilai 0.77.



Gambar 5. Hasil Confusion Matrix dari Model RoBERTa

Gambar 5 menampilkan confusion matrix dari model RoBERTa dalam klasifikasi intent chatbot. Sumbu X menunjukkan kelas yang diprediksi oleh model, sedangkan sumbu Y menunjukkan kelas yang sebenarnya. Nilai di setiap sel menunjukkan jumlah sampel yang diklasifikasikan ke dalam kelas tertentu. Dari Gambar 5 tersebut, terlihat bahwa model RoBERTa memiliki performa yang cukup baik dalam mengklasifikasikan intent beasiswa dan asrama. Namun, model masih sering salah dalam mengklasifikasikan intent kantin sebagai menu. Hal ini mengindikasikan bahwa model mengalami kesulitan dalam membedakan kedua intent tersebut.

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 595−606
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.6051

	precision	recall	f1-score	support
acak	0.00	0.00	0.00	1
acara	1.00	1.00	1.00	3
asrama	1.00	0.83	0.91	6
beasiswa	1.00	1.00	1.00	7
biaya	0.83	0.83	0.83	6
dekan	1.00	1.00	1.00	1
dokumen	1.00	1.00	1.00	3
fasilitas	1.00	1.00	1.00	1
infrastruktur	1.00	1.00	1.00	1
jam	0.80	1.00	0.89	4
jurusan	1.00	0.86	0.92	7
kantin	0.50	1.00	0.67	3
karir	1.00	1.00	1.00	2
kepala jurusan	1.00	1.00	1.00	1
kepala prodi	1.00	1.00	1.00	1
komite	1.00	1.00	1.00	1
lantai	0.67	1.00	0.80	2
liburan	1.00	0.67	0.80	3
lokasi	1.00	1.00	1.00	4
menu	0.00	0.00	0.00	2
nama	0.75	1.00	0.86	3
nomor	1.00	1.00	1.00	4
olahraga	1.00	1.00	1.00	2
pelonco	1.00	1.00	1.00	2
pencipta	0.80	1.00	0.89	4
penerimaan	0.50	1.00	0.67	1
perpisahan	0.50	0.33	0.40	3
perpustakaan	1.00	1.00	1.00	4
pmb	1.00	1.00	1.00	2
rektor	1.00	1.00	1.00	2
salam	0.50	0.50	0.50	2
salut	0.50	0.67	0.57	3
sekretaris jurusan	1.00	1.00	1.00	1
semester	0.60	1.00	0.75	3
seragam	1.00	0.50	0.67	2
silabus	0.50	0.50	0.50	2
tugas	0.00	0.00	0.00	1
umpat	0.00	0.00	0.00	2
accuracy			0.84	102
macro avg	0.78	0.81	0.78	102
weighted avg	0.82	0.84	0.82	102

Gambar 6. Pengukuran Metrik Setiap Intent Model RoBERTa

Selanjutnya, kita dapat menghitung metrik evaluasi seperti accuracy, precision, recall, dan F1-score untuk setiap intent dengan menggunakan RoBERTa. Metrik-metrik ini memberikan gambaran yang lebih kuantitatif tentang kinerja model yang dapat dilihat pada Gambar 6. Sebagian besar intent memiliki precision, recall, dan F1-score = 1.00, hal ini mengindikasikan klasifikasi yang sangat baik untuk kategori-kategori tersebut, seperti acara, beasiswa, dekan, dokumen, fasilitas, infrastruktur, karir, kepala jurusan, dan kepala prodi, kepala prodi, komite, lokasi, nomor, olahraga, pelonco, perpustakaan, PMB, rektor, dan sekretaris jurusan. Hasil eksperimen untuk mengevaluasi model klasifikasi intent chatbot menggunakan metode RoBERTa menunjukkan akurasi sebesar 0.84, seperti yang ditunjukkan pada Gambar 6.

3.4 Hasil Model IndoBERT

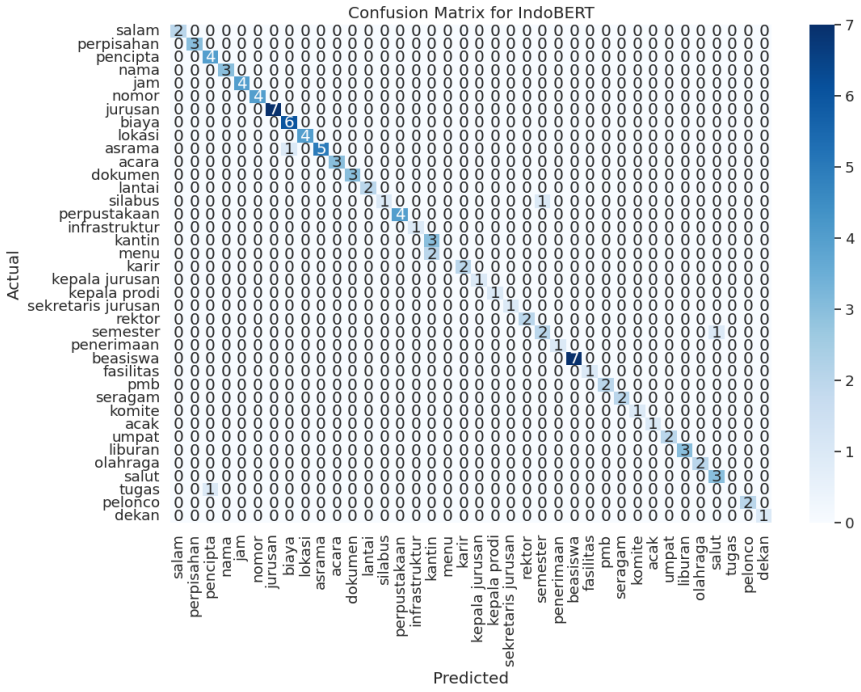
Setelah pelatihan, model dievaluasi dengan menggunakan dataset uji. Metrik evaluasi meliputi Accuracy (Acc.), F1-score (F1), Precision (Prec.) dan Recall. Hasil dari pelatihan dan pengujian model IndoBERT ditunjukkan pada Tabel 7.

Tabel 7. Hasil Pelatihan dan Pengujian Model IndoBERT

	Loss	Acc.	F1	Prec.	Recall
Train	0.005	0.997	0.998	0.998	0.998
Test	0.265	0.941	0.912	0.912	0.921

Dapat dilihat pada Tabel 7 bahwa model mampu mencapai akurasi yang tinggi pada data training (0.99), namun akurasi pada data testing lebih rendah (0.94). Meskipun sedikit menurun dari training, akurasi pada data uji tetap tinggi dan menunjukkan performa yang kuat. Selain itu, nilai F1-Score mendapatkan nilai 0.99 pada data training sedangkan pada data testing sebesar 0.91. Walaupun menghasilkan nilai yang sedikit lebih rendah, namun masih menunjukkan performa yang solid dalam mempertahankan keseimbangan antara precision dan recall pada data testing. Secara keseluruhan, hasil pelatihan dan pengujian model IndoBERT menunjukkan bahwa model memiliki performa yang sangat baik pada data training dan masih memberikan hasil yang solid pada data testing, meskipun terdapat penurunan yang wajar. Performa pada fase testing dengan akurasi sebesar 0.94 dan F1-score sebesar 0.91, menunjukkan bahwa IndoBERT mampu melakukan klasifikasi intent dengan sangat baik pada data bahasa Indonesia.

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 595−606
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.6051



Gambar 7. Hasil Confusion Matrix dari Model IndoBERT

Gambar 7 menampilkan confusion matrix dari model IndoBERT dalam klasifikasi intent chatbot. Sumbu X menunjukkan kelas yang diprediksi oleh model, sedangkan sumbu Y menunjukkan kelas yang sebenarnya. Nilai di setiap sel menunjukkan jumlah sampel yang diklasifikasikan ke dalam kelas tertentu. Dari gambar 7 di atas, terlihat bahwa model IndoBERT memiliki performa yang baik dalam mengklasifikasikan intent jurusan dan beasiswa. Namun, model masih sering salah dalam mengklasifikasikan intent salut sebagai semester. Hal ini mengindikasikan bahwa model mengalami kesulitan dalam membedakan kedua intent tersebut.

	precision	recall	f1-score	support
salam	1.00	1.00	1.00	2
perpisahan	1.00	1.00	1.00	3
pencipta	0.80	1.00	0.89	4
nama	1.00	1.00	1.00	3
jam	1.00	1.00	1.00	4
nomor	1.00	1.00	1.00	4
jurusan	1.00	1.00	1.00	7
biaya	0.86	1.00	0.92	6
lokasi	1.00	1.00	1.00	4
asrama	1.00	0.83	0.91	6
acara	1.00	1.00	1.00	3
dokumen	1.00	1.00	1.00	3
lantai	1.00	1.00	1.00	2
silabus	1.00	0.50	0.67	2
perpustakaan	1.00	1.00	1.00	4
infrastruktur	1.00	1.00	1.00	1
kantin	0.60	1.00	0.75	3
menu	0.00	0.00	0.00	2
karir	1.00	1.00	1.00	2
kepala jurusan	1.00	1.00	1.00	1
kepala prodi	1.00	1.00	1.00	1
sekretaris jurusan	1.00	1.00	1.00	1
rektor	1.00	1.00	1.00	2
semester	0.67	0.67	0.67	3
penerimaan	1.00	1.00	1.00	1
beasiswa	1.00	1.00	1.00	7
fasilitas	1.00	1.00	1.00	1
pmb	1.00	1.00	1.00	2
seragam	1.00	1.00	1.00	2
komite	1.00	1.00	1.00	1
acak	1.00	1.00	1.00	1
umpat	1.00	1.00	1.00	2
liburan	1.00	1.00	1.00	3
olahraga	1.00	1.00	1.00	2
salut	0.75	1.00	0.86	3
tugas	0.00	0.00	0.00	1
pelonco	1.00	1.00	1.00	2
dekan	1.00	1.00	1.00	1
accuracy			0.94	102
macro avg	0.91	0.92	0.91	102
weighted avg	0.93	0.94	0.93	102

Gambar 8. Pengukuran Metrik Setiap Intent Model IndoBERT

Selanjutnya, kita dapat menghitung metrik evaluasi seperti accuracy, precision, recall, dan F1-score untuk setiap intent dengan menggunakan IndoBERT. Metrik-metrik ini memberikan gambaran yang lebih kuantitatif tentang kinerja model dan hasil eksperimen untuk mengevaluasi model klasifikasi intent chatbot menggunakan metode IndoBERT menunjukkan akurasi sebesar 0.94, seperti yang ditunjukkan pada Gambar 8.

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 595−606
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.6051

3.5 Perbandingan Hasil Model BERT, RoBERTa dan IndoBERT

Dalam eksperimen ini, kita melakukan analisis perbandingan hasil antara model BERT, RoBERTa, dan IndoBERT untuk mengevaluasi performa masing-masing model pada tugas klasifikasi intent chatbot. Setiap model memiliki karakteristik unik dalam menangani data dan memahami konteks, yang tercermin dari arsitektur dan data pra-pelatihan yang digunakan.

Tabel 8. Hasil Perbandingan Model BERT, RoBERTa, dan IndoBERT

Komponen	BERT	RoBERTa	IndoBERT
Accuracy	0.89	0.84	0.94
Precision	0.89	0.77	0.91
Recall	0.88	0.80	0.92
F1-Score	0.87	0.77	0.91

Tabel 8 menunjukkan bahwa model BERT mencapai akurasi sebesar 0.89, model RoBERTa memperoleh akurasi 0.84, dan model IndoBERT mencapai akurasi tertinggi sebesar 0,94. Hasil ini menunjukkan bahwa model IndoBERT memberikan akurasi yang lebih baik dibandingkan model BERT dan RoBERTa ketika menggunakan hyperparameter dan dataset yang sama untuk bahasa Indonesia. Akurasi yang tinggi penting dalam klasifikasi intent karena semakin tinggi akurasi, semakin besar kemungkinan chatbot memberikan respon yang tepat sesuai dengan kebutuhan pengguna.

Ada beberapa faktor yang dapat menyebabkan kinerja IndoBERT memiliki akurasi lebih baik dibandingkan dengan BERT dan RoBERTa dalam tugas klasifikasi intent chatbot untuk bahasa Indonesia, meskipun menggunakan hyperparameter dan dataset yang sama. Pertama, IndoBERT dibangun khusus untuk bahasa Indonesia sehingga dilatih pada dataset yang diambil sepenuhnya dari teks dalam bahasa Indonesia. Ini memberikan IndoBERT keuntungan dalam memahami struktur, sintaks, dan kekhasan bahasa Indonesia, dibandingkan dengan BERT dan RoBERTa, yang pada dasarnya dilatih pada korpus multibahasa atau bahkan hanya bahasa Inggris. Kedua, dataset yang digunakan dalam pre-training IndoBERT diambil dari teks bahasa Indonesia, termasuk artikel berita, blog, media sosial, dan sumber-sumber lain yang relevan dengan konteks percakapan di Indonesia. Hal ini memungkinkan IndoBERT lebih akurat dalam menangkap intent pengguna yang berbicara dalam bahasa Indonesia. Meskipun IndoBERT dan BERT memiliki arsitektur yang serupa, arsitektur model IndoBERT mungkin telah disesuaikan atau dioptimalkan untuk menangani masalah-masalah linguistik tertentu dalam bahasa Indonesia. Sedangkan beberapa faktor yang dapat menyebabkan kinerja BERT lebih unggul dibandingkan model RoBERTa. Pertama, karakteristik dan ukuran dataset yang relatif kecil mungkin membuatnya lebih cocok untuk model BERT sehingga mengurangi risiko overfitting. Kedua, penyesuaian hyperparameter memainkan peran penting, karena konfigurasi hyperparameter yang berbeda dapat menghasilkan kinerja yang lebih baik untuk model BERT khusus pada dataset ini.

4. KESIMPULAN

Perbandingan hasil eksperimen untuk klasifikasi intent chatbot menggunakan model BERT, RoBERTa dan IndoBERT menghasilkan kesimpulan bahwa model BERT, RoBERTa dan IndoBERT dapat digunakan secara efektif untuk klasifikasi intent chatbot berbahasa Indonesia. Setelah menerjemahkan dataset chatbot universitas ke dalam Bahasa Indonesia, model IndoBERT mencapai akurasi 0.94, mengungguli model BERT yang mencapai 0.89 dan model RoBERTa yang mencapai 0.84. Hal ini menunjukkan bahwa performa IndoBERT lebih baik dalam klasifikasi intent chatbot dibandingkan BERT dan RoBERTa ketika menggunakan dataset Bahasa Indonesia dan hyperparameter yang sama. Model IndoBERT memberi keunggulan dalam memahami teks berbahasa Indonesia dikarenakan model IndoBERT dilatih dengan korpus yang spesifik dan relevan dengan bahasa Indonesia sedangkan model BERT dan RoBERTa dilatih dengan fokus utama pada bahasa Inggris atau data multibahasa. Secara keseluruhan, performa IndoBERT lebih baik dibandingkan dengan BERT dan RoBERTa disebabkan oleh pelatihan yang lebih spesifik untuk bahasa Indonesia, pretraining yang lebih relevan, serta adaptasi yang lebih efektif terhadap konteks dan struktur bahasa Indonesia. Inilah yang menjadikan IndoBERT sebagai model yang unggul dalam klasifikasi intent chatbot berbahasa Indonesia.

UCAPAN TERIMAKASIH

Penelitian ini didanai oleh Direktorat Riset, Teknologi, dan Pengabdian kepada Masyarakat di bawah Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, sesuai dengan sesuai dengan kontrak pelaksanaan Program Bantuan Operasional Perguruan Tinggi Negeri (Program Penelitian) Tahun Anggaran 2024, Nomor Kontrak: 090/E5/PG.02.00.PL/2024.

REFERENCES

Journal of Information System Research (JOSH)
Volume 6, No. 1, Oktober 2024, pp 595−606
ISSN 2686-228X (media online)
https://ejurnal.seminar-id.com/index.php/josh/
DOI 10.47065/josh.v6i1.6051

[1] Rohim and Zuliarso, "Penerapan Algoritma Deep Learning Untuk Pengembangan Chatbot Yang Digunakan Untuk Konsultasi Dan Pengenalan Tentang Virus Covid-19," PIXEL, vol. 15, no. 2, pp. 267–278, Dec. 2022, doi: 10.51903/pixel.v15i2.777.
[2] R. C. Hutama, F. Fauziah, and R. T. Komalasari, "Aplikasi Chatbot Berbasis Teks Menggunakan Algoritma Naive Bayes Classifier FAQ GrabAds," STRING, vol. 6, no. 1, p. 90, Aug. 2021, doi: 10.30998/string.v6i1.9919.
[3] N. Shahin and L. Ismail, "From Rule-Based Models to Deep Learning Transformers Architectures for Natural Language Processing and Sign Language Translation Systems: Survey, Taxonomy and Performance Evaluation," 2024, arXiv. doi: 10.48550/ARXIV.2408.14825.
[4] L. Villa, D. Carneros-Prado, A. Sánchez-Miguel, C. C. Dobrescu, and R. Hervás, "Conversational Agent Development Through Large Language Models: Approach with GPT," in Proceedings of the 15th International Conference on Ubiquitous Computing & Ambient Intelligence (UCAmI 2023), vol. 835, J. Bravo and G. Urzáiz, Eds., in Lecture Notes in Networks and Systems, vol. 835. , Cham: Springer Nature Switzerland, 2023, pp. 286–297. doi: 10.1007/978-3-031-48306-6_29.
[5] D. Griol, Z. Callejas, J. M. Molina, and A. Sanchis, "Adaptive dialogue management using intent clustering and fuzzy rules," Expert Systems, vol. 38, no. 1, p. e12630, Jan. 2021, doi: 10.1111/exsy.12630.
[6] W. Maeng and J. Lee, "Designing a Chatbot for Survivors of Sexual Violence: Exploratory Study for Hybrid Approach Combining Rule-based Chatbot and ML-based Chatbot," in Asian CHI Symposium 2021, Yokohama Japan: ACM, May 2021, pp. 160–166. doi: 10.1145/3429360.3468203.
[7] A. Birim and M. Erden, "Robustness to Spelling Errors for Intent Detection," in 2022 30th Signal Processing and Communications Applications Conference (SIU), Safranbolu, Turkey: IEEE, May 2022, pp. 1–4. doi: 10.1109/SIU55565.2022.9864722.
[8] J. Liu, Y. Li, and M. Lin, "Review of Intent Detection Methods in the Human-Machine Dialogue System," J. Phys.: Conf. Ser., vol. 1267, no. 1, p. 012059, Jul. 2019, doi: 10.1088/1742-6596/1267/1/012059.
[9] R. A. Sanjaya and E. Winarno, "Pengembangan Chatbot Informasi Pariwisata di Kabupaten Pati Menggunakan Metode Natural Language Processing Berbasis Dialogflow," Jutisi J. Tek. Sis. Info, vol. 13, no. 1, p. 368, Apr. 2024, doi: 10.35889/jutisi.v13i1.1828.
[10] F. Fatharani, K. P. Kania, J. Hutahaean, and S. R. Wulan, "Deteksi Intensi Chatbot Berbahasa Indonesia dengan Menggunakan Metode Capsule Network," josh, vol. 3, no. 4, pp. 590–596, Jul. 2022, doi: 10.47065/josh.v3i4.1821.
[11] N. Boudjani, V. Colas, and A. Fotouhi, "Intent Classification: French Recruitment Chatbot Use Case," in 2023 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA: IEEE, Dec. 2023, pp. 681–685. doi: 10.1109/CSCI62032.2023.00117.
[12] J.-H. Lee, E. H.-K. Wu, Y.-Y. Ou, Y.-C. Lee, C.-H. Lee, and C.-R. Chung, "Anti-Drugs Chatbot: Chinese BERT-Based Cognitive Intent Analysis," IEEE Trans. Comput. Soc. Syst., vol. 11, no. 1, pp. 514–521, Feb. 2023, doi: 10.1109/TCSS.2023.3238477.
[13] F. Roma, G. Sansonetti, G. D'Aniello, and A. Micarelli, "A BERT-Based Approach to Intent Recognition," in IEEE EUROCON 2023 - 20th International Conference on Smart Technologies, Torino, Italy: IEEE, Jul. 2023, pp. 568–572. doi: 10.1109/EUROCON56442.2023.10198959.
[14] S. Sayenju et al., "Quantification and Mitigation of Directional Pairwise Class Confusion Bias in a Chatbot Intent Classification Model," Int. J. Semantic Computing, vol. 16, no. 04, pp. 497–520, Dec. 2022, doi: 10.1142/S1793351X22500040.
[15] Y. Guo et al., "ESIE-BERT: Enriching Sub-words Information Explicitly with BERT for Joint Intent Classification and SlotFilling," Feb. 02, 2023, arXiv: arXiv:2211.14829. Accessed: Sep. 02, 2024. [Online]. Available: http://arxiv.org/abs/2211.14829
[16] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," Jul. 26, 2019, arXiv: arXiv:1907.11692. Accessed: Mar. 13, 2024. [Online]. Available: http://arxiv.org/abs/1907.11692
[17] A. Souha, C. Ouaddi, L. Benaddi, and A. Jakimi, "Pre-Trained Models for Intent Classification in Chatbot: Comparative Study and Critical Analysis," in 2023 6th International Conference on Advanced Communication Technologies and Networking (CommNet), Rabat, Morocco: IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/CommNet60167.2023.10365312.
[18] K. K. Jayanth, G. Bharathi Mohan, R. P. Kumar, and M. Rithani, "Intent Recognition Leveraging XLM-RoBERTa for Effective NLU," in 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India: IEEE, Jun. 2024, pp. 877–882. doi: 10.1109/ICAAIC60222.2024.10575275.
[19] B. Wilie et al., "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," Oct. 08, 2020, arXiv: arXiv:2009.05387. Accessed: Oct. 07, 2024. [Online]. Available: http://arxiv.org/abs/2009.05387
[20] Nirali Vaghani, "Chatbot dataset." Kaggle, 2024. doi: 10.34740/KAGGLE/DSV/5024271.
[21] DeepL, "DeepL Translator." [Online]. Available: https://www.deepl.com/
[22] J. H. Tandijaya and I. Sugiarto, "Klasifikasi dalam Pembuatan Portal Berita Online dengan Menggunakan Metode BERT," vol. Vol 9, No 2 (2021), 2021.
[23] A. Aljabar, "Mengungkap Opini Publik: Pendekatan BERT-based- caused untuk Analisis Sentimen pada Komentar Film," vol. 5, no. 1, 2024.
[24] R. Khusuma, W. Maharani, and P. H. Gani, "Personality Detection On Twitter User With RoBERTa," Jurnal Media Informatika Budidarma, vol. 7, 2023.
[25] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," Nov. 01, 2020, arXiv: arXiv:2011.00677. Accessed: Oct. 09, 2024. [Online]. Available: http://arxiv.org/abs/2011.00677
[26] L. Geni, E. Yulianti, and D. I. Sensuse, "Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using IndoBERT Language Models," vol. 9, no. 3, 2023.