JURNAL FASILKOM
Volume 13 No. 2 | Agustus 2023: 326-333

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Analisis sentimen terhadap pelayanan Kesehatan berdasarkan ulasan Google Maps menggunakan BERT

Ardiansyah[1], Adika Sri Widagdo[2], Krisna Nuresa Qodri[3], Fachruddin Edi Nugroho Saputro[4], Nisrina Akbar Rizky P[5]

[1,2,3,4,5]Teknologi Informasi, Fakultas Kesehatan Dan Teknologi, Universitas Muhammadiyah Klaten

[1]ardiansyah@umkla.ac.id *, [2] adikasw@umkla.ac.id, [3]krisna@umkla.ac.id, [4] fachruddinedi@umkla.ac.id, [5]nisrinaakbar@umkla.ac.id

Abstract

The utilization of technology has developed in various scientific fields, without exception in health. Hospitals, health centers, and clinics are part of the health sector. Thus, it must evolve according to health service standards and patient measures or service user satisfaction that needs to be measured using sentiment analysis. The Media to give opinions to Health service providers is Google Maps. However, the anomaly is that the reviews and the given text are sometimes not correlated. Thus, The utilization of sentiment analysis using the scientific branch of artificial intelligence, namely Natural Language Processing (NLP), is an effective way to infer opinions. The research concluded that the BERT indobenchmark/indobert-base-p1 model has good performance to use of Indonesian text classification with a dataset of 4228 data after preprocessing, which at the beginning of the collection process obtained data as much as 4748 data. Split datasets into 3 data, namely training, validation, and test data, with a ratio of 70:30:30. The experimental results, The researchers found that the model allows the use of the model with other Indonesian texts. The results are 85% for accuracy and weighted avg, and macro avg 75% on the validation data training process. While the testing data training process is 86% for accuracy and weighted avg, the macro avg 73%. In addition, researchers found that services are the most frequent topic in Health Services. Even though health services have improved, positive sentiment is the highest compared to other sentiment classes.

Keywords: BERT, Sentiment Analysis, Healthcare, Opinion Mining, Google Maps

Abstrak

Pemanfaatan teknologi telah berkembang diberbagai bidang keilmuan tanpa terkecuali bidang Kesehatan. Rumah sakit, balai atau klinik Kesehatan merupakan bagian dari bidang Kesehatan yang dipaksa harus terus berkembang menyesuaikan standar dari pelayanan Kesehatan. Dengan adanya standart pelayanan Kesehatan, pengukur terkait kepuasan pasien atau pengguna layanan menjadi sesuatu yang perlu diukur menggunakan sentiment analisis. Salah satu media yang menjadi wadah pemberian opini kepada penyedia layanan Kesehatan adalah Google Maps. Namun, yang menjadi anomaly adalah antara review dan teks diberikan terkadang tidak berkorelasi sehingga, pemanfaatan analisis sentimen menggunakan cabang keilmuan kecerdasan buatan ialah Natural Language Processing (NLP) menjadi cara yang efektif dalam menyimpulkan opini. Penelitian yang dilakukan oleh peneliti penyimpulkan bahwa model BERT indobenchmark/indobert-base-p1 memiliki kinerja yang baik dalam penggunaan klasifikasi teks berbahasa Indonesia dengan dataset sebanyak 4228 data setelah proses preprocessing yang pada awal proses pengumpulan mendapatkan data sebanyak 4748 data. Dataset dibagi menjadi 3 data yaitu data latih, data validasi, dan data test dengan ratio 70:30:30. Dari hasil percobaan peneliti mendapatkan model dapat di gunakan dengan teks berbahasa Indonesia lainnya dapat dilakukan dengan baik. Selanjutnya, hasil yang didapatkan 85% untuk akurasi dan weighted avg, macro avg 75% pada proses latih data validasi. Sedangkan proses latih data testing 86% untuk akurasi dan weighted avg, macro avg 73%. Selain itu, peneliti mendapatkan topik pelayanan menjadi topik yang paling sering muncul terhadap penggunaan layanan Kesehatan. Walaupun begitu, pelayanan kesehatan telah mengalami peningkatan hal tersebut di tunjukan dengan sentiment positive menjadi sentimen tertinggi dibandingkan dengan kelas sentimen lainnya.

Kata kunci: BERT, Analisis Sentimen, Pelayanan Kesehatan, Sentimen Analisis, Google Maps

.

1. Pendahuluan

Pemanfaatan teknologi informasi telah mengalami perkembangan yang sangat signifikan dalam waktu yang singkat diberbagai keilmuan seperti Pendidikan, Sosial dan masih banyak bidang keilmuan yang telah menggunakan teknologi informasi tanpa kecuali pada bidang keilmuan Kesehatan [1]. Pemanfaatan teknologi yang digunakan pada bidang kesehatan seperti pencatatan medis secara elektronik [2], layanan kesehatan berupa aplikasi smartphone yang telah terhubung dengan badan jaminan sosial [1], serta penggunaan kecerdasan buatan dalam memperkuat layanan Kesehatan [3].

Penyedia atau fasilitas layanan kesehatan seperti Rumah Sakit, balai / klinik Kesehatan [4] merupakan pihak yang dipaksa terus berbenah serta berinovasi agar

P-ISSN : 2089-3353
E-ISSN : 2808-9162

mencapai kualitas pelayanan yang sesuai standar dalam memenuhi kepentingan masyarakat [5]. Dengan adanya standart dalam pelayanan kesehatan sehingga diperlukan pengukuran terkait kepuasan pasien atau pengguna layanan tersebut [6]. Dilain sisi, seringkali pelanggan tidak memberikan penilaian yang mengekspresikan secara keseluruhan terhadap pelayanan yang diberikan [7]. Hal ini dikarenakan opini yang diberikan biasanya dalam bentuk teks dan tidak relevan dengan bintang pada fitur penilaian [8].

Pelayanan pelanggan telah banyak diberikan oleh pelanggan atau pengguna layanan secara sukarela melalui berbagai media yang tersedia seperti sosial media, Google, serta media online lainnya [9]. Pemanfaatan analisis sentimen menjadi sesuatu yang akan membantu dalam representasi opini dari pengguna layanan [10]. Analisis sentimen merupakan bagian dari bidang keilmuan *Natural Language Processing* (NLP) [11]. Analisis sentiment sendiri bertujuan untuk mendapatkan *Point Of View* (POV) yang lebih baik terhadap produk ataupun layanan yang digunakan [12]. Sentiment analisis dapat digunakan sebagai alat bantu menganalisis kepuasan pasien pada bagian yang perlu diperbaiki atau ditingkatkan [13].

Penelitian sebelumnya yang telah dilakukan menggunakan Twitter sebagai pengambilan data untuk mendapatkan opini publik terkait program kesehatan masyarakat yang telah berjalan dalam rentan waktu 2015 sampai 2019 [14]. Penelitian yang dilakukan bertujuan mengetahui sentimen masyarakat terhadap imunisasi, asuransi Kesehatan, stunting, gizi buruk, pelayanan Kesehatan serta jaminan Kesehatan yang telah berjalan. Dataset yang didapatkan dari Twitter sebanyak 6000 pesan tweet yang tidak terstruktur sehingga perlunya preprocessing dataset. Jumlah pesan yang didapatkan terbagi menjadi Gizi Buruk 450 tweet, Jaminan Kesehatan 148 tweet, Asuransi Kesehatan 465 tweet, Imunisasi 1000 tweet, Stunting 1000 tweet, serta pelayanan kesehatan 248 tweet. Selain itu, proses klasifikasi menggunakan metode Lexicon yang mengkategorikan setiap tweet berdasarkan sifat dari tweet tersebut kedalam 3 kategori Negatif, Netral, dan Positif.

Penelitian lainnya dengan pemanfaatan sosial media sebagai media pengumpulan data serta pengolahan data menggunakan *Bidirectional Encoder From Transformers* (BERT) menyimpulkan bahwasannya sentiment analisis berdasarkan pengumpulan data dari sosial media dapat dilakukan [15]. Namun penulis menyarankan agar peneliti selanjutnya melakukan preprosessing yang lebih detail agar mendapatkan hasil yang lebih maksimal dari penelitian yang dilakukan. Pengambilan data dari sosial media Instagram berjumlah 1.052 data terkait Covid-19 varian omicron yang dibagi menjadi 663, 338, 1 (negatif, netral, positif).

Penelitian selanjutnya terhadap pelayanan Rumah Sakit Umum Daerah menggunakan Support Vector Machine (SVM) dan Term Frequency-Inverse Document Frequency (TF-IDF) mendapatkan hasil berdasarkan proses Cross Validation yaitu akurasi 88%, recall 87,5%, presisi 90%, dan f-measure 87,5% [5]. Data yang digunakan merupakan data primary yang didapatkan langsung dari RSUD dan telah di labeli dari pihak RSUD.

Penelitian lainnya terkait opini pasien terhadap pelayanan kesehatan yang dilakukan RamyaSri dkk., menggunakan metode Lexicon-Based. Pengumpulan data berasal dari sosial media yang memanfaatkan library BeautifulSoup untuk mengambil data. Alat yang digunakan pada penelitian tersebut Vader dan TextBlob sebagai klasifikasi sentiment analisis. Sedangkan Confusion Matrix sebagai alat pengukur akurasi dari algoritma yang digunakan. TextBlob Lexicon-Based didapatkan lebih unggul 1.1% daripada Vader Lexicon-based. Namun jika dikomparasi secara keselurahan presisi, recall, dan F1-Score menampilkan Vader Lexicon-Based lebih unggul. Akurasi 71%, Presisi 42.2%, Recall 38.1% dan F1-Score 40% hasil dari Vader Lexicon-Based. Sedangkan hasil dari Textblob 73%, 36.5%, 34.3%, 35.36% [3].

Penelitian tentang kepuasan pelayanan Rumah Sakit Mitra Husada menggunakan Logistic Regression serta data berdasarkan hasil survey dari google form yang diisikan langsung oleh pengguna layanan [6]. Validasi dan pengujian data menggunakan confusion matrix dan cross validation 10kali. Dari penelitian yang dilakukan adalah akurasi 88%, presisi 88%, Recall 88%. Selain itu kesimpulan yang diambil Logistic Regression dapat digunakan untuk merepresentasikan pelayanan terhadap Rumah Sakit Mitra Husada.

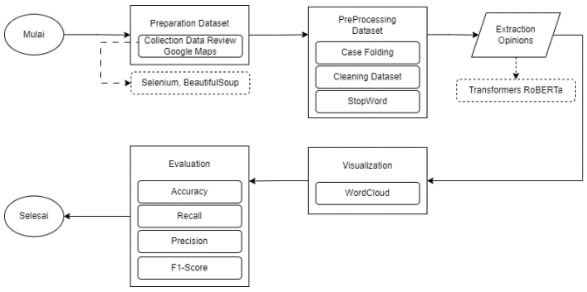
Penelitian lain yang menggunakan *review* Google *map* sebagai tempat pengambilan data adalah Arianto & Budi serta Mathayomchan dan Sripanidkulchai [16], [17]. Pada kedua penelitian yang dilakukan dengan mempertimbangkan antara rating dan teks yang diberikan pelanggan pada ulasan di google maps. Selain itu pada keduanya memberikan skema skema uji coba dan memvalidasinya menggunakan cross validation. Terkhusus penelitian Mathayomchan dan Sripanidkulchai menampilkan wordcloud sebagai representasi apa topik yang sering dibahas.

Pada penelitian yang dilakukan bertujuan untuk menguji penggunaan BERT berbahasa Indonesia terhadap data yang didapatkan dari ulasan Google *maps*. Selain itu, penelitian mencari topik yang menjadi perhatian pelanggan terhadap pelayanan Kesehatan yang tersedia tanpa menyebutkan instansi terkait.

2. Metode Penelitian

Metode penelitian yang digunakan terlihat pada gambar 1 sebagai alur penelitian.

P-ISSN : 2089-3353
E-ISSN : 2808-9162



Gambar 1. Alur Penelitian

2.1. *Preparation* Data

Collection data yang dilakukan menggunakan bahasa pemrograman Python, selenium, library BeautifulSoup dengan parameter layanan kesehatan, fasilitas Kesehatan untuk pengumpulan data pada Google Maps. Yang dimana platform tersebut merupakan layanan yang terkenal diseluruh dunia. Proses klasifikasi menjadi lebih mudah di bandingkan dengan sosial media dikarenakan pada Google Maps tidak hanya memberikan pengguna layanan untuk berkomentar tetapi memberikan bintang sebagai parameter opini pengguna [17]. Selain itu, Proses penghapusan data yang duplikat untuk menjaga dataset menjadi konsisten sebelum masuk pada tahap selanjutnya. Walaupun begitu, seringkali ditemukan ulasan dengan bintang yang baik tetapi memiliki ulasan yang buruk.

2.2. *Pre-Processing Dataset*

Case Folding
Proses perubahan teks pada kalimat menjadi huruf kecil meskipun nama kota, dan yang lainnya. Dikarenakan pada ulasan yang di dapatkan dari hasil koleksi data tidak memiliki format yang sama dengan yang lainnya [18].

Data Cleaning
Pembersihan data adalah proses penghapusan teks yang mengandung Prefix, HTML tag, Punctuation, maupun angka, menggunakan regex.

Stop Word Removal
Proses penghapusan kata yang tidak berkaitan pada kalimat berdasarkan *Stop Word* lis dalam bentuk bahasa Inggris atau bahasa Indonesia [19].

2.3. *Extraction Opinions*

Ekstrasi opini pada penelitian yang dilakukan menggunakan algoritma *Robustly optimized BERT approach* (RoBERTa) Transformers yang dimana merupakan algoritma lanjutan dari BERT yang dilaunching Google [20] pada tahun 2018, sedangkan RoBERTa pada tahun 2019 [21]. Pemanfaatan Tranformers RoBERTa menggunakan model berbahasa Indonesia untuk memberikan label pada setiap data yang dimiliki. Labeling menggunakan Transformers RoBERTa dapat mempersingkat waktu penelitian ataupun kebutuhan lainnya dikarenakan RoBERTa sendiri telah dilakukan pelatihan dataset menggunakan *corpus* yang sangat besar. Sehingga teks mudah di klasifikasikan kedalam kelas kelas tertentu [22].

2.4. *Output Word Cloud*

WordCloud merupakan representasi kalimat yang dominan, semakin besar kalimat yang tampil maka kalimat tersebut adalah kalimat yang paling sering digunakan pada hasil analisis [19].

2.5. Evaluasi

Evaluasi metriks digunakan untuk mengukur akurasi, recall, presisi, dan f1-score dari hasil klasifikasi atau perbandingan metode yang digunakan, berikut persamaannya [23], [24]:

Akurasi merupakan pengukuran prediksi yang benar secara keseluruhan saat pelatihan dilakukan. Berikut persamaan 1 menghitung akurasi.

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

Persamaan 2 merupakan *Recall* yang dimana merupakan pengukuran kemungkinan dari kelas positive atau benar yang bernilai benar juga. Selain itu persamaan 3 dan 4 merupakan perhitungan *macro-average* dan *weighted*

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$Macro\ Avg\ Recall = \frac{\sum_{i=1}^{l} \frac{TP_i}{TP_i+FN_i}}{l} \quad (3)$$

$$Weighted\ Avg\ Recall = \frac{\sum_{i=1}^{l} \frac{TP_i}{TP_i+FN_i} * n_i}{l} \quad (4)$$

Presisi merupakan pengukuran prediksi terhadap nilai benar meskipun data sebenarnya telah masuk kedalam kelas positive ataupun benar. Berikut persamaan 5,6,7 menghitung presisi, *macro* dan *weighted*.

$$Presisi = \frac{TP}{TP+FP} \quad (5)$$

$$Macro\ Avg\ Precision = \frac{\sum_{i=1}^{l} \frac{TP_i}{TP_i+FP_i}}{l} \quad (6)$$

$$Weighted\ Avg\ Precision = \frac{\sum_{i=1}^{l} \frac{TP_i}{TP_i+FP_i} * n_i}{l} \quad (7)$$

F1-Score merupakan pengukuran yang dihasilkan dari hasil pemrosesan atau perhitungan presisi dan recall dengan mempertimbangkan kelas hasil prediksi dengan kelas sesungguhnya pada data menggunakan persamaan 8 dibawah ini.

$$F1-Score = \frac{2*Recall*Presisi}{Presisi+Recall} \quad (8)$$

Dengan TP adalah *True Positive*, TF adalah *True False*, FP adalah *False Positive*, FN adalah *False Negative*, l merupakan total kelas pada data yang dilatih, sedangkan n_i merupakan total data yang dilatih dari setiap kelas [23], [24].

P-ISSN : 2089-3353
E-ISSN : 2808-9162

3. Hasil dan Pembahasan

3.1. *Collection Dataset*

Pada penelitian yang telah dilakukan proses pengumpulan dataset, peneliti mendapatkan data dari ulasan di Google *Maps* puskesmas dan rumah sakit swasta ataupun negri yang berada pada daerah X. Penggunaan selenium, beautifulsoup menjadi tools yang digunakan dalam pengumpulan data perlu dilakukan dikarenakan ulasan pada Google *maps* menggunakan *lazyload*. Yang dimana perlu dilakukan scrolling untuk menampilkan ulasan secara keseluruhan dan perlu Tindakan lebih lanjut untuk menekan ulasan secara keselurahan ulasan yang panjang.

Selain itu, pengumpulan ulasan dilakukan 1 juli 2023 dengan batasan tahun mulai dari tahun 2019 sampai tahun 2023. Dari hasil pengumpulan data didapatkan sebanyak 4748 data dengan *users unique* sebanyak 4342 penguna. Berikut tabel 1 merupakan contoh hasil pengumpulan data.

Tabel 1.Contoh Hasil Pengumpulan Data

Waktu	Ulasan	Rating
setahun lalu	Satpamnya tidak ramah pada saat saya berkunjung kesana, mending ke RS lain aja deh.	1
sebulan lalu	pelayanan obat antre lama sekali. padahal ada loket , tapi kenapa loket no tidak dioptimalkan? kasihan sama pasien yang nunggu obat.	1
8 bulan lalu	Petugas labnya tidak ramah, senyum aja kyknya enggak. Nggak responsive diajak bicara. Semuanya serba lama.. sepertinya lebih milih rs lain kalo begini pelayanannya	2
setahun lalu	Pelayanan memuaskan . tertib rapi lancar	2
3 tahun lalu	Pendaftaran pasien yg cukup lama sekali,mendahulukan pasien yg gawat bukan karena urutan kedatangan tp tidak seperti yg tertera harus menunggu sesuai panggilan.	3
seminggu lalu	fasilitas kesehatan yang baik layanan nya	3
5 tahun lalu	Pelayanannya okee 👍 …	4
sebulan lalu	Alur pelayanan mudah di pahami , petugas pelayanan ramah😍 .	4
7 jam lalu	Pegawai nya emosian aowkaowk	5
3 bulan lalu	Rumah sakitnya bersih, megah seperti istana.	5

Dari data yang didapatkan peneliti melakukan penghapusan duplikat data dengan parameter username, rating dan komentar yang sama untuk menjaga dataset konsisten dari spam ulasan.

3.2. *Pre-Processing Dataset*

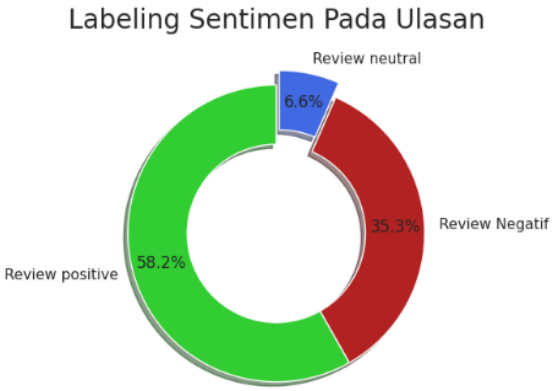
Tahapan pre-processing peneliti melakukan beberapa langkah yaitu: mengubah semua teks berbentuk huruf kecil secara keseluruhan, penghapusan prefix, Punctuation, dan mengubah bentuk kata yang menggunakan bahasa lokal menjadi bahasa Indonesia baku. Berikut tabel 2 contoh perubahan teks pada dataset.

Tabel 2.Contoh Pre-Processing Dataset

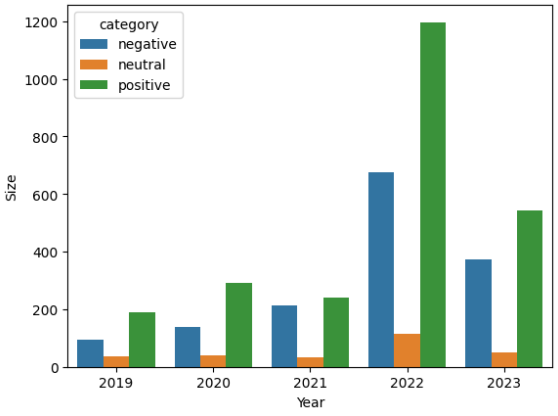
Data Asli	Security ne Jan Ra enek Adab blasss . Tolong segera dirubah management SDM nya . Apa perlu di tembus sampai atas suara suara ini ???
Bahasa lokal	Security nya Jan tidak ada Adab sekali . Tolong segera di rubah management SDM nya . Apa perlu di tembus sampai atas suara suara ini ?
Hasil	security nya adab sekali . tolong rubah management sdm nya . perlu tembus atas suara suara ?

3.3. Ekstrasi Opini

Labeling dataset yang dilakukan menggunakan transformers hungging face dengan model RoBERTa berbahasa Indonesia memperoleh presentasi sentimen positive sebagai sentiment tertinggi dengan 2460 data atau 58.2% dari pada kelas negative ataupun netral. Berikut gambar 2 hasil labeling dataset serta gambar 3 memperlihatkan persebaran sentimen setiap tahun.



Gambar 2. Persebaran Sentimen



Gambar 3. Grafik Progres Sentimen

Selain itu, pada penelitian didapatkan topik atau penggunaan kata yang paling sering digunakan pada ulasan terlihat pada gambar 4 visualisasi word cloud untuk semua kelas.

P-ISSN : 2089-3353
E-ISSN : 2808-9162

Gambar 4. Wordcloud kata pada semua kelas

Penggunaan kata 6 teratas untuk semua kelas terlihat pada tabel 3 dibawah ini.

Tabel 3.Tabel Frekuensi Kemunculan Kata

Kelas	Kata	Frekuensi
Positive	Pelayanan	1624
	Ramah	1161
	Cepat	396
	Dokter	348
	Pasien	323
	Perawat	311
Negative	Pelayanan	813
	Pasien	624
	Jam	590
	Dokter	463
	Ramah	365
	Menunggu	282
Netral	Pelayanan	42
	Jam	37
	Pasien	33
	Dokter	27
	Poli	24

Dari proses pre-processing yang telah dilakukan peneliti dataset yang awalnya 4748 data berubah menjadi 4228 data. Hal ini dikarenakan adanya ulasan yang hanya menggunakan *emoticon* ataupun simbol sehingga ketika proses *cleaning dataset* data tersebut akan dihapus dari daftar dataset yang akan di latih.

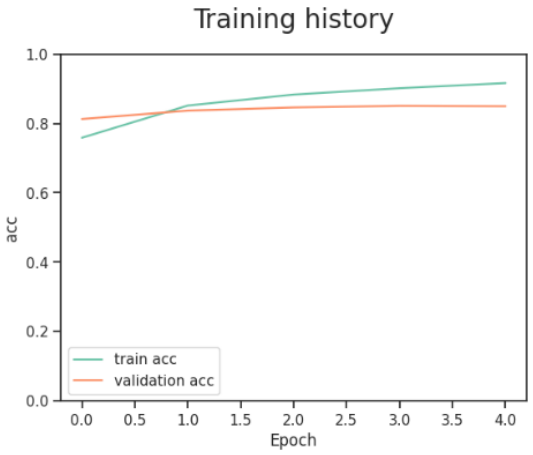
3.3. Modeling

Sebelum dilakukan modeling peneliti melakukan pembagian dataset menjadi 3 yaitu: data latih, validasi, serta test. Pemecahan data dilakukan dengan ratio 70% dari data secara keseluruhan menjadi data latih, 30% menjadi data validasi. Sedangkan data test diambil 30% dari data validasi. Berikut tabel 4 pembagian dataset.

Tabel 4. Pembagian Dataset

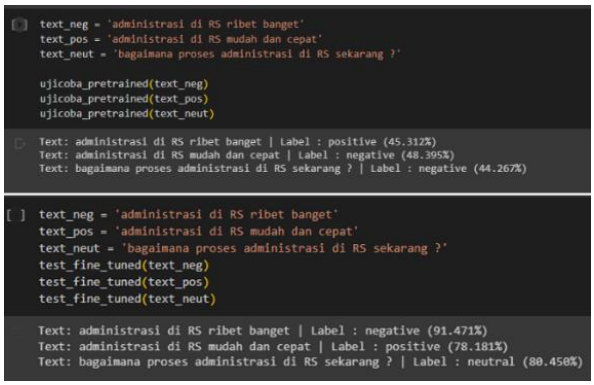
Ratio	Data Latih	Data Validasi	Data Testing
70:30:30	2900	870	374

Dalam proses melatih data peneliti memilih model indobenchmark/indobert-base-p1 [25] hugging face menggunakan GPU Tesla T4 di Google Colab. Selain itu pada percobaan yang dilakukan peneliti juga menggunakan 5 epochs, Adam optimizer, parameter *learningrate* 3e-6, *batchsize* 32, *num_workers* 16, dan maksimal Panjang kalimat 512. Berikut gambar 5 grafik hasil latih dataset pada penelitian.



Gambar 5. Grafik *Train Dataset*

Sebelum dan setelah melatih dataset untuk model belajar, peneliti melakukan ujicoba terhadap 3 jenis kalimat dengan berikut gambar 6 hasil ujicoba terhadap model.



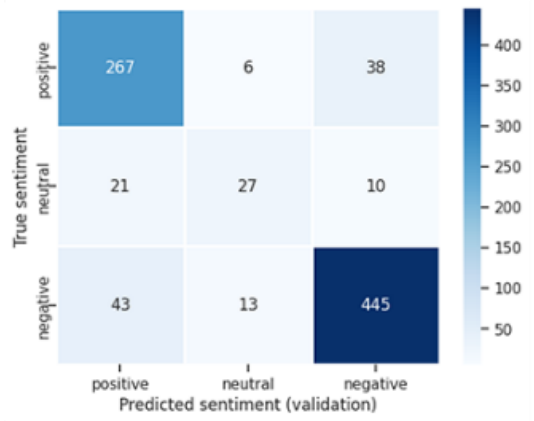
Gambar 6. Uji Coba Model

Gambar 6 terlihat adanya peningkatan performa secara signifikan setelah dilakukan pelatihan data yang dimiliki.

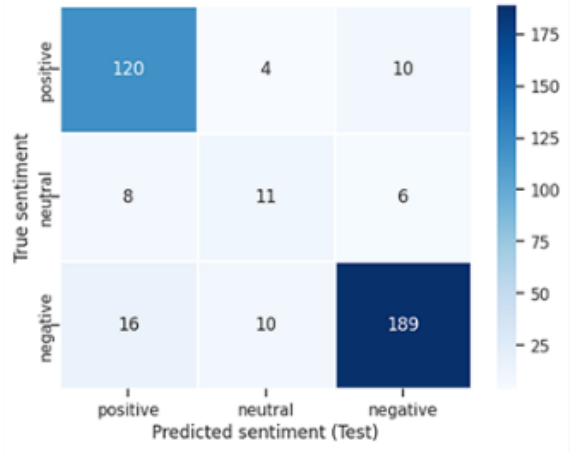
3.4. Evaluasi

Hasil dari melatih data yang telah dilakukan peneliti mendapatkan hasil *confusion matrix* untuk semua kelasnya seperti terlihat pada gambar 7 *confusion matrix* saat latih menggunakan data validasi dan gambar 8 saat melatih menggunakan data test.

P-ISSN : 2089-3353
E-ISSN : 2808-9162



Gambar 7. *Confusion Matrix Result* Data Validasi



Gambar 8. *Confusion Matrix Result* Data Testing

Dari hasil matrix yang di dapatkan akan di hitung untuk mendapatkan akurasi, *recall*, presisi, dan skor f1 berdasarkan persamaan 1,2,3,4 pada *section* sebelumnya. Berikut dibawah ini perhitungan untuk mendapatkan akurasi, *recall*, presisi, dan skor f1 pada data validasi dengan total data yang telah disebutkan pada tabel 4.

$$Accuracy = \frac{267 + 27 + 445}{870} = \frac{739}{870} = 0.85$$

$$Precision\ (positive) = \frac{267}{267 + (21 + 43)} = \frac{267}{64} = 0.81$$

$$Precision\ (neutral) = \frac{27}{27 + (6 + 13)} = \frac{27}{46} = 0.59$$

$$Precision\ (negative) = \frac{445}{445 + (38 + 10)} = \frac{445}{493} = 0.90$$

$$Macro\ Avg\ Precision = \frac{0.81 + 0.59 + 0.90}{3} = \frac{2.3}{3} = 0.77$$

$$Weighted\ Avg\ Precision = \frac{(0.81 * 311) + (0.59 * 58) + (0.90 * 501)}{(311 + 58 + 501)} = \frac{(251.99) + (34.22) + (450.90)}{870} = \frac{737.03}{870} = 0.85$$

$$Recall\ (positive) = \frac{267}{267 + (6 + 38)} = \frac{267}{311} = 0.86$$

$$Recall\ (neutral) = \frac{27}{27 + (21 + 10)} = \frac{27}{58} = 0.47$$

$$Recall\ (negative) = \frac{445}{445 + (43 + 13)} = \frac{445}{501} = 0.89$$

$$Macro\ Avg\ Recall = \frac{0.86 + 0.47 + 0.89}{3} = \frac{2.22}{3} = 0.74$$

$$Weighted\ Avg\ Recall = \frac{(0.86 * 311) + (0.47 * 58) + (0.89 * 501)}{(311 + 58 + 501)} = \frac{(267) + (27) + (445)}{870} = \frac{739}{870} = 0.85$$

$$F1 - Score\ (positive) = \frac{2 * 0.81 * 0.86}{0.81 + 0.86} = \frac{1.39}{1.67} = 0.83$$

$$F1 - Score\ (neutral) = \frac{2 * 0.59 * 0.47}{0.59 + 0.47} = \frac{0.55}{1.05} = 0.52$$

$$F1 - Score\ (negative) = \frac{2 * 0.90 * 0.89}{0.90 + 0.89} = \frac{1.60}{1.79} = 0.90$$

Dari perhitungan tersebut didapatkan hasil terlihat pada gambar 9 untuk semua kelas dari pelatihan data validasi. Selain itu, perhitungan yang dilakukan menggunakan fungsi *ROUNDUP* atau pembulatan ke atas.

VALIDASI

	precision	recall	f1-score	support
positive	0.81	0.86	0.83	311
neutral	0.59	0.47	0.52	58
negative	0.90	0.89	0.90	501
accuracy			0.85	870
macro avg	0.77	0.74	0.75	870
weighted avg	0.85	0.85	0.85	870

Gambar 9. Hasil Evaluasi Data Validasi

Berikut perhitungan akurasi, presisi, recall dan skor f1 untuk data test.

$$Accuracy = \frac{120 + 11 + 189}{374} = \frac{320}{374} = 0.86$$

$$Precision\ (positive) = \frac{120}{120 + (8 + 16)} = \frac{120}{144} = 0.83$$

$$Precision\ (neutral) = \frac{11}{11 + (4 + 10)} = \frac{11}{25} = 0.44$$

$$Precision\ (negative) = \frac{189}{189 + (10 + 6)} = \frac{189}{205} = 0.92$$

$$Macro\ Avg\ Precision = \frac{0.83 + 0.44 + 0.92}{3} = \frac{2.19}{3} = 0.73$$

$$Weighted\ Avg\ Precision = \frac{(0.83 * 134) + (0.59 * 25) + (0.90 * 215)}{(311 + 58 + 501)} = \frac{(111.22) + (11) + (197.8)}{374} = \frac{320.02}{374} = 0.86$$

$$Recall\ (positive) = \frac{120}{120 + (4 + 10)} = \frac{120}{134} = 0.90$$

$$Recall\ (neutral) = \frac{11}{11 + (8 + 6)} = \frac{11}{25} = 0.44$$

$$Recall\ (negative) = \frac{189}{189 + (16 + 10)} = \frac{189}{215} = 0.88$$

$$Macro\ Avg\ Recall = \frac{0.90 + 0.44 + 0.88}{3} = \frac{2.22}{3} = 0.74$$

$$Weighted\ Avg\ Precision = \frac{(0.90 * 134) + (0.44 * 25) + (0.88 * 215)}{(311 + 58 + 501)} = \frac{(120) + (11) + (189)}{374} = \frac{320}{374} = 0.86$$

Setelah dilakukan perhitungan menggunakan persamaan di dapatkan hasil presisi, *recall*, f1-score, akurasi, *macro average*, dan *weighted average* seperti terlihat pada gambar 10.

P-ISSN : 2089-3353
E-ISSN : 2808-9162

TESTING

	precision	recall	f1-score	support
positive	0.83	0.90	0.86	134
neutral	0.44	0.44	0.44	25
negative	0.92	0.88	0.90	215
accuracy			0.86	374
macro avg	0.73	0.74	0.73	374
weighted avg	0.86	0.86	0.86	374

Gambar 10. Hasil Evaluasi Data Test

Pada gambar 9 dan 10 terlihat adanya perbedaan hasil skor f1, presisi, recall pada kelas netral memiliki nilai paling rendah di antara kelas yang lainnya. Hal ini dikerenakan jumlah data di kelas tersebut sangat minim. Berbeda dengan 2 kelas lainnya yang memiliki kecukupan data hal ini biasa dikatakan tidak keseimbangan data pada kelas.

Selain itu, dari penelitian yang telah dilakukan mendapatkan peningkatan klasifikasi setelah *fine-tuning* pada model pada gambar 6 walapun dataset yang digunakan pada melatih model merupakan *imbalance* dataset yang di labeling menggunakan tranformers RoBERTa. Selanjutnya topik yang menjadi fokus pada masyarakat adalah topik pelayanan pada semua kelas telihat pada table 3. Selanjutnya, pada pelatihan untuk model validasi didapatkan 85% akurasi, 75% macro avg, 85% weighted avg. Sedangkan pada pelatihan model testing 86% akurasi, 73% macro avg, 86% weighted avg.

4. Kesimpulan

Pada penelitian yang dilakukan peneliti penyimpulkan bahwasan nya model indobenchmark/indobert-base-p1 memiliki kinerja yang baik dalam penggunaan klasifikasi teks berbahasa Indonesia serta model tersebut dapat dilakukan dengan menggunakan teks berbahasa Indonesia lainnya. Namun, pengumpulan dataset sangat perlu diperhatikan untuk mendapatkan hasil yang lebih baik lagi pada setiap kelasnya.

Selanjutnya, dari hasil penelitian yang dilakukan penyedia layanan Kesehatan telah meningkat dari waktu ke waktu. Namun, masih ada masalah yang sama muncul di setiap kelasnya seperti pelayanan yang masih perlu di tinjau ulang. Dikarenakan hal tersebut merupakan fokus utama dari masyarakat terhadap pelayanan Kesehatan.

Ucapan Terimakasih

Terima kasih untuk Universitas Muhammadiyah Klaten yang telah mendukung terselenggarakan kegiatan penelitian ini.

Daftar Rujukan

[1] A. Hendra, "Analisis Sentimen Review Halodoc Menggunakan Nai¨ve Bayes Classifier," MEI, 2021.

[2] A. M. Abirami and A. Askarunisa, "Sentiment analysis model to emphasize the impact of online reviews in healthcare industry," *Online Information Review*, vol. 41, no. 4, pp. 471–486, 2017.

[3] V. I. S. RamyaSri, C. Niharika, K. Maneesh, and M. Ismail, "Sentiment Analysis of Patients' Opinions in Healthcare using Lexicon-based Method," *Int J Eng Adv Technol*, vol. 9, no. 1, pp. 6977–6981, 2019.

[4] F. Nurfitriani and F. Sembiring, "Sistem Pendukung Keputusan Pemilihan Rumah Sakit Menggunakan Metode Simple Additive Weight (Saw)," in *Seminar Nasional Sistem Informasi dan Manajemen Informatika Universitas Nusa Putra*, 2021, pp. 98–106.

[5] J. D. C. Aruan, B. Rahayudi, and A. Ridok, "Analisis Sentimen Opini Masyarakat terhadap Pelayanan Rumah Sakit Umum Daerah menggunakan Metode Support Vector Machine dan Term Frequency–Inverse Document Frequency," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, vol. 2548, p. 964X, 2022.

[6] A. R. C. Adi, "ANALISIS KEPUASAN PELAYANAN RUMAH SAKIR MITRA HUSADA PRINGSEWI MENGGUNAKAAN METODE LOGISTIC REGRESSION," *Jurnal Ilmu Data*, vol. 2, no. 12, 2022.

[7] E. H. Muktafin, P. Pramono, and K. Kusrini, "Sentiments analysis of customer satisfaction in public services using K-nearest neighbors algorithm and natural language processing approach," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 19, no. 1, pp. 146–154, 2021.

[8] S. Kushwah and S. Das, "Sentiment analysis of big-data in healthcare: issue and challenges," in *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, IEEE, 2020, pp. 658–663.

[9] J. Godara, R. Aron, and M. Shabaz, "Sentiment analysis and sarcasm detection from social network to train health-care professionals," *World Journal of Engineering*, vol. 19, no. 1, pp. 124–133, 2022.

[10] E. Saad *et al.*, "Determining the efficiency of drugs under special conditions from users' reviews on healthcare web forums," *IEEE Access*, vol. 9, pp. 85721–85737, 2021.

[11] D. Georgiou, A. MacFarlane, and T. Russell-Rose, "Extracting sentiment from healthcare survey data: An evaluation of sentiment analysis tools," in *2015 Science and Information Conference (SAI)*, IEEE, 2015, pp. 352–361.

[12] G. Saranya, G. Geetha, K. Meenakshi, and S. Karpagaselvi, "Sentiment analysis of healthcare Tweets using SVM Classifier," in *2020 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)*, IEEE, 2020, pp. 1–3.

[13] K. Jindal and R. Aron, "A systematic study of sentiment analysis for social media data," *Mater Today Proc*, 2021.

[14] A. S. Aribowo, "Analisis Sentimen Publik pada Program Kesehatan Masyarakat menggunakan Twitter Opinion Mining," in *Seminar Nasional Informatika Medis (SNIMed)*, 2018, pp. 17–23.

[15] D. Pradipta, K. Kusrini, and H. Al Fatta, "Sentiment Analysis Comments Covid-19 Variant Omicron on Social Media Instagram with Bidirectional Encoder from Transformers (BERT)," *JTECS: Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem dan Komputer*, vol. 3, no. 1, pp. 51–58, 2023.

[16] D. Arianto and I. Budi, "Aspect-based Sentiment Analysis on Indonesia's Tourism Destinations Based on Google Maps User Code-Mixed Reviews (Study Case: Borobudur and Prambanan Temples)," in *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation*, 2020, pp. 359–367.

[17] B. Mathayomchan and K. Sripanidkulchai, "Utilizing Google translated Reviews from Google maps in sentiment analysis for Phuket tourist attractions," in *2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, IEEE, 2019, pp. 260–265.

[18] D. Alita and A. R. Isnain, "Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier," *Jurnal Komputasi*, vol. 8, no. 2, pp. 50–58, 2020.

P-ISSN : 2089-3353
E-ISSN : 2808-9162

[19] T. Mustaqim, K. Umam, and M. A. Muslim, "Twitter text mining for sentiment analysis on government's response to forest fires with vader lexicon polarity detection and k-nearest neighbor algorithm," in *Journal of Physics: Conference Series*, IOP Publishing, 2020, p. 032024.

[20] A. PATEL, P. OZA, and S. AGRAWAL, "Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model," *Procedia Comput Sci*, vol. 218, pp. 2459–2467, 2023.

[21] J. A. García-Díaz, F. García-Sánchez, and R. Valencia-García, "Smart Analysis of Economics Sentiment in Spanish based on Linguistic Features and Transformers," *IEEE Access*, 2023.

[22] Y. Liu *et al.*, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[23] V. Shanmuganathan, V. H. C. de Albuquerque, P. C. S. Barbosa, M. C. dos Reis, G. Dhiman, and M. A. Shah, "Software based sentiment analysis of clinical data for healthcare sector," *IET Software*, 2023.

[24] A. H. Oliaee, S. Das, J. Liu, and M. A. Rahman, "Using Bidirectional Encoder Representations from Transformers (BERT) to classify traffic crash severity types," *Natural Language Processing Journal*, vol. 3, p. 100007, Jun. 2023, doi: 10.1016/j.nlp.2023.100007.

[25] B. Wilie *et al.*, "IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding," *arXiv preprint arXiv:2009.05387*, 2020.