

Analisis Sentimen Aplikasi Inarisk Personal di Google Play Store Menggunakan Algoritma NLP

Anugrah Putra Ridha Utama¹, Wirdayanti², Syahrullah³, Anita Ahmad Kasim⁴

aputraridha728@gmail.com¹, wirda_arbie@untad.ac.id², syahroellah.ms@gmail.com³,

nita.kasim@gmail.com⁴

^{1,2,3,4} Universitas Tadulako

Informasi Artikel

Diterima : 13 Mei 2024

Direview : 16 Jun 2024

Disetujui : 30 Jun 2024

Kata Kunci

Analisis Sentimen,
Aplikasi, Algoritma,
Inarisk, NLP

Abstrak

Dalam perspektif geografis, Indonesia diakui sebagai negara yang rawan terhadap bencana alam, BNPB sebagai salah satu lembaga penanggulangan bencana yang ada di Indonesia memiliki aplikasi inarisk sebagai aplikasi yang memberikan informasi tentang kebencanaan. Untuk meningkatkan efektivitas layanan kepada pengguna analisis sentimen dilakukan untuk mengetahui review pengguna pada aplikasi inarisk. Tujuan penelitian ini adalah untuk menjelajahi pendapat pengguna terhadap aplikasi InaRisk di platform Google Play Store. Peneliti menggunakan total 597 ulasan sebagai data yang akan dianalisis untuk mengetahui polaritas dengan menggunakan model NLP. Hasil penelitian menunjukkan bahwa sebanyak 34,7% ulasan memiliki sentimen positif, 52,6% ulasan memiliki sentimen netral, dan 12,7% ulasan memiliki sentimen negatif. Ini menunjukkan meskipun aplikasi sudah cukup baik namun masih butuh beberapa perbaikan dalam aplikasinya. Dalam pengklasifikasian prediksi dilakukan menggunakan 6% data sebagai data uji dan 94% sebagai data latih dimana model regresi digunakan dan berhasil menghasilkan akurasi sebesar 69,81% ini mengindikasikan bahwa model regresi logistik yang digunakan untuk melakukan pengklasifikasian prediksi cukup baik sehingga dapat digunakan untuk analisis sentimen penelitian lainnya.

Keywords

Sentiment Analysis,
Applications, Algorithms,
Inarisk, NLP

Abstract

From a geographical perspective, Indonesia is recognized as a country that is prone to natural disasters, BNPB as one of the disaster management institutions in Indonesia has the Inarisk application as an application that provides information about disasters. To improve the effectiveness of services to users, sentiment analysis is carried out to find out user reviews on the InaRisk application. The purpose of this study is to explore user opinions on the InaRisk application on the Google Play Store platform. Researchers used a total of 597 reviews as data to be analyzed to determine polarity using NLP models. The results showed that as many as 34.7% of reviews had a positive sentiment, 52.6% of reviews had a neutral sentiment, and 12.7% of reviews had a negative sentiment. This shows that even though the application is good enough, it still needs some improvements in the application. In classifying predictions carried out using 6% of data as test data and 94% as training data where regression models were used and succeeded in producing an accuracy of 69.81%, this indicates that the logistic regression model used to classify predictions is good enough so that it can be used for sentiment analysis of other studies.

A. Pendahuluan

Indonesia dikenal sebagai salah satu negara yang rentan terhadap kejadian bencana[1], Indonesia dikenal sebagai salah satu negara dengan tingkat kejadian bencana alam tertinggi di dunia[2]. pada tahun 2023 jumlah kejadian bencana di Indonesia berjumlah 5.400 bencana[3]. BNPB memiliki fungsi penting dalam upaya penanggulangan bencana. Sebagai sebuah lembaga nasional non-pemerintah, BNPB dibentuk untuk merespons kebutuhan akan pengembangan strategi penanganan bencana alam[4]. Dalam rangka meningkatkan efektivitas penanggulangan bencana, BNPB telah mengembangkan aplikasi bernama InaRisk. InaRisk adalah sebuah aplikasi yang menyajikan informasi tentang tingkat risiko di suatu wilayah dan menyediakan rekomendasi tindakan yang dapat diambil secara partisipatif untuk mengantisipasi risiko tersebut. Aplikasi ini merupakan produk dari kerja sama antara pemerintah dan mitra-mitra lain yang memiliki pengalaman dalam bidang edukasi kebencanaan di Indonesia[5].

Berdasarkan data di situs Google Play Store, aplikasi InaRisk telah di unduh oleh 100 ribu lebih pengguna dengan rating 4,4 dari 1,9 ribu ulasan. Dengan rating yang bisa dibilang cukup baik namun aplikasi ini masih memiliki kekurangan contohnya yaitu respon aplikasi yang masih lambat dimana dari salah satu kekurangan tersebut itu saja bisa menjadi penghambat BNPB dalam menyajikan pelayanan terbaik untuk penggunanya. Jika ini terus diabaikan maka akan dapat mengakibatkan keterlambatan informasi ke pengguna secara realtime yang bisa saja menjadi kerugian bagi pengguna dalam evakuasi ketika bencana sedang terjadi dan keterlambatan penerimaan informasi potensi bahaya di sekitar pengguna serta juga ketidaknyamanan pengguna dalam pemakaian bahkan dapat mendorong pengguna untuk beralih ke platform digital lainnya. sehingga perbaikan aplikasi menjadi hal yang penting untuk diperhatikan oleh BNPB untuk meningkatkan kualitas layanan.

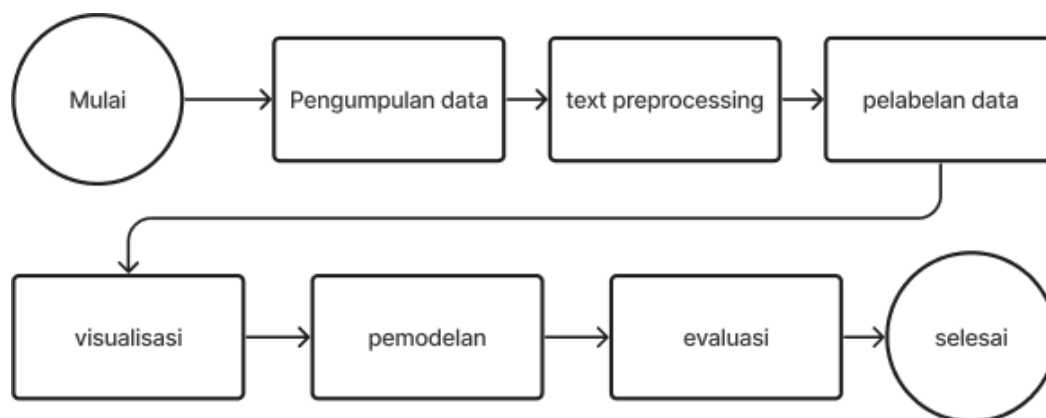
BNPB sebagai lembaga pemerintah pengelola aplikasi InaRisk perlu mengevaluasi layanan, salah satu metode yang dapat diterapkan adalah analisis sentimen terhadap ulasan pengguna. Analisis sentimen adalah sebuah proses di mana dokumen teks dikelompokkan ke dalam kategori-kategori tertentu, seperti sentimen positif dan negatif. Hal ini memungkinkan kita untuk mengevaluasi seberapa besar pengaruh dan manfaat dari analisis sentimen tersebut[6]. Analisis sentimen memungkinkan pengidentifikasian masalah atau kekurangan yang dihadapi pengguna dalam penggunaan aplikasi. Selain itu, ini juga membantu BNPB memahami kebutuhan dan preferensi pengguna secara lebih mendalam, sehingga dapat dilakukan penyesuaian atau penambahan fitur yang sesuai dengan harapan pengguna.

Natural Language Processing (NLP) adalah salah satu bidang dalam machine learning yang memungkinkan untuk melakukan analisis sentimen. NLP (Natural Language Processing) adalah teknologi yang digunakan untuk memproses data dalam bentuk bahasa alami sehingga dapat dipahami atau dimengerti oleh mesin[7]. TextBlob merupakan salah satu perpustakaan (library) NLP yang berguna untuk memproses data tekstual, seperti melakukan analisis sentimen[8]. Proses ini akan mempermudah dalam pengklasifikasian ulasan dengan lebih baik dan akurat.

Berdasarkan uraian diatas penulis bertujuan untuk melakukan analisis sentimen pengguna aplikasi Inarisk dengan menggunakan machine learning Natural Language Processing (NLP). Hasil penelitian ini dapat digunakan sebagai evaluasi oleh BNPB untuk meningkatkan layanan mereka dalam memberikan edukasi dan informasi kepada pengguna.

B. Metode Penelitian

Tujuan dari penelitian ini adalah untuk menganalisis sentimen pengguna terhadap InaRisk dengan tujuan memahami sentimen pengguna saat memberikan ulasan terhadap aplikasi tersebut, dalam penelitian ini Google play store digunakan dalam pencarian dataset. Analisis sentimen dilakukan menggunakan sistem cerdas NLP(natural language processing) dengan library Textblob. Dalam penelitian ini dibutuhkan tahapan penelitian yang digunakan untuk mempermudah dalam pengerjaan. Metode pada penelitian ini melalui 6 tahapan. Garis besar penelitian tergambar dengan jelas dalam Gambar 1.



Gambar 1. Metode Penelitian Tahapan

Tahapan yang terlihat dalam Gambar 1 dapat dijelaskan sebagai berikut:

1. Pengumpulan data

Data dikumpulkan dari Google Play Store menggunakan teknik pengambilan data (scraping) dengan bahasa pemrograman Python melalui platform Google Colab. Hasil scraping berupa dataset yang nantinya akan digunakan untuk melakukan klasifikasi.

2. Text preprocessing

Preprocessing menjadi langkah pertama yang sangat krusial dalam analisis sentimen karena memiliki dampak besar terhadap performa klasifikasi [9]. Beberapa proses preprocessing antara lain adalah cleaning, case folding, stemming, tokenizing

3. Pelabelan data

Pada tahap ini, dilakukan proses penentuan polaritas teks dengan menggunakan library TextBlob. TextBlob adalah sebuah perpustakaan (library) Python yang berguna untuk memproses data tekstual. Dengan menyediakan API, TextBlob memungkinkan untuk menjalankan berbagai tugas dalam pemrosesan bahasa alami (NLP), termasuk ekstraksi frasa kata

benda, analisis sentimen, klasifikasi, penerjemahan, dan fungsi-fungsi lainnya.[10].Dimana polaritas ini nantinya berguna untuk pemberian label positif, netral, ataupun negatif pada sebuah ulasan.

4. Visualisasi

Tahap Visualisasi dilakukan menggunakan library matplotlib dengan menampilkan diagram pie, dan word cloud agar hasil dari analisis dan pelabelan data dapat dengan mudah dilihat dan dipahami.

5. Pemodelan

Pemodelan dilakukan pada data yang ada dengan menggunakan machine learning regresi logistik untuk memprediksi sentimen dari ulasan yang diberikan. Dengan melatih model, model dapat belajar untuk mengenali pola dan fitur dalam ulasan yang berkaitan dengan masing masing kategori sentimen sehingga nantinya model dapat digunakan untuk memprediksi sentimen dari ulasan baru yang belum terlihat sebelumnya.

6. Evaluasi

Tahap evaluasi dilakukan berdasarkan hasil pemodelan yang dilakukan sebelumnya. Evaluasi bekerja dengan metrik evaluasi yang berguna untuk menganalisis kinerja model klasifikasi regresi logistik dalam melakukan pemodelan.

C. Hasil dan Pembahasan

1. Pengumpulan Data

Dalam penelitian ini, data dikumpulkan dengan melakukan teknik pengambilan data atau scraping dari Google Play Store menggunakan library google-play-scraper di lingkungan Google Colaboratory. Dari keseluruhan 1.900 ulasan pengguna yang dapat ditemukan di Google Play Store, peneliti berhasil mengambil 597 ulasan sesuai dengan batas yang diterapkan oleh library yang digunakan untuk pengambilan data. Data yang didapatkan antara lain yaitu username, rating, waktu, dan ulasan yang diberikan. Dataset kemudian disimpan dalam format CSV. Gambar 2 menampilkan data mentah yang berhasil diambil.

	userName	score	at	text
0	hermawan tanambel	3	2024-03-31 20:12:13	Masih perlu banyak perbaikan, salah satunya so...
1	koko irenk	1	2023-11-15 05:15:08	mau isi evaluasi SPAB kok lola, mosok ya ram 8...
2	ma met	5	2024-01-03 14:41:29	Aplikasi fungsinya bagus. Cuma masih banyak pe...
3	Patar J.H Manalu (Mr.Jansen)	5	2023-12-15 11:45:13	bagus sudah berusaha ,kita aprisiasi usahanya ...
4	Aldo Afandi	3	2020-08-13 11:38:34	Yah aplikasinya masih perlu banyak pemutakhira...

Gambar 2. Hasil scraping data google play store

2. Text Preprocessing

Pada text preprocessing data text harus dibersihkan dari elemen-elemen yang tidak diperlukan, tahapan awal melibatkan penyesuaian teks dengan merubah semua hurufnya menjadi huruf kecil. Langkah ini dimaksudkan untuk memastikan konsistensi dalam penulisan, sehingga perbedaan huruf tidak memengaruhi interpretasi kata. Selanjutnya, elemen-elemen yang tidak digunakan seperti emotikon akan dihapus. Langkah selanjutnya adalah menghapus karakter non-huruf dan non-spasi dari teks untuk membersihkannya dari karakter khusus seperti tanda baca. Tambahan dari itu, dilakukan pula tokenisasi pada teks untuk mengubahnya dari rangkaian huruf menjadi unit yang lebih kecil, seperti kata-kata atau frasa. Tahap ini dilakukan agar text dapat lebih mudah di klasifikasi. Data yang telah melalui tahap pembersihan atau pra-pemrosesan dapat dilihat pada Gambar 3.

	userName	score	at	text
0	hermawan tanambel	3	2024-03-31 20:12:13	perbaikan salah satunya proses respon nya yg l...
1	koko irenk	1	2023-11-15 05:15:08	isi evaluasi spab lola mosok ya ram 8 plus and...
2	ma met	5	2024-01-03 14:41:29	aplikasi fungsinya bagus pembenahan sinkronisa...
3	Patar J.H Manalu (Mr.Jansen)	5	2023-12-15 11:45:13	bagus berusaha aprisiasi usahanya 5 bintang na...
4	Aldo Afandi	3	2020-08-13 11:38:34	yah aplikasinya pemutakhiran tugas kkn mahasis...

Gambar 3. Hasil dari pra-pemrosesan data teks

3. Pelabelan data

Pada tahap ini, dilakukan proses pengelompokan kumpulan data ke dalam kelompok-kelompok tertentu. Pembobotan nilai sentimen text menggunakan library TextBlob dengan membuat sebuah fungsi untuk melabelkan text dengan menghitung polarity dimana jika polarity kurang dari 0 maka akan terlabel negatif, jika polarity 0 maka akan terlabel netral, dan jika polarity lebih dari 0 maka akan mendapatkan label positif. Gambar 4 menunjukkan data yang telah melalui proses pelabelan.

	userName	score	at	text	polarity_results	polarity	untranslated_texts	SplittedText
0	hermawan tanambel	3	2024-03-31 20:12:13	perbaikan salah satunya proses respon nya yg l...	(-0.07500000000000002, [])	-0.075	[]	[p, e, r, b, a, i, k, a, n, s, a, l, a, h, ...]
1	koko irenk	1	2023-11-15 05:15:08	isi evaluasi spab lola mosok ya ram 8 plus and...	(0.2, [])	0.200	[]	[i, s, i, e, v, a, l, u, a, s, i, s, p, ...]
2	ma met	5	2024-01-03 14:41:29	aplikasi fungsinya bagus pembenahan sinkronisa...	(0.7, [])	0.700	[]	[a, p, l, i, k, a, s, i, f, u, n, g, s, i, ...]
3	Patar J.H Manalu (Mr.Jansen)	5	2023-12-15 11:45:13	bagus berusaha aprisiasi usahanya 5 bintang na...	(0.19999999999999996, [])	0.200	[]	[b, a, g, u, s, b, e, r, u, s, a, h, a, ...]
4	Aldo Afandi	3	2020-08-13 11:38:34	yah aplikasinya pemutakhiran tugas kkn mahasis...	(0.0, [])	0.000	[]	[y, a, h, a, p, l, i, k, a, s, i, n, y, a, ...]

Gambar 4. Hasil pencarian polarity

Setelah berhasil menghitung dan mendapatkan polarity kemudian data akan diberikan label sesuai dengan polarity yang didapatkan. Gambar 5 menampilkan hasil pelabelan.

positif antara lain “bagus”, “bermanfaat”, “mudah”. Kemudian untuk kata yang banyak digunakan pada sentimen negatif antara lain “lemot”, “lambat”, “error”, “bingung”. Ini memperlihatkan bagaimana pengguna menilai aplikasi Inarisk.

5. Pemodelan

Pada tahap pemodelan dataset dibuatkan variabel X dan Y terlebih dahulu dimana variabel X diambil pada kolom ‘text’ dan Y diambil pada kolom ‘sentiment’. Selanjutnya, Dataset dibagi menjadi subset data latih dan data uji. Dalam penelitian ini, 6% dari keseluruhan dataset digunakan sebagai data uji. Pemodelan dilakukan menggunakan machine learning logistik regresi dimana data latih variabel X dan Y akan dilatih. Akurasi dihitung dan mendapatkan nilai akurasi 0.6754716981132075 yang jika di persentasekan menjadi 67.55%. Kemudian peningkatan akurasi dilakukan menggunakan model SVC dan berhasil meningkatkan akurasi menjadi 69.81%.

6. Evaluasi

Hasil evaluasi dilakukan menggunakan library sklearn dengan menampilkan accuracy, precision, recall, dan skor F1 dengan 6% dari keseluruhan dataset digunakan sebagai data uji, sementara 94% digunakan sebagai data latih. Informasi ini dapat dilihat pada Gambar 7, didapatkan nilai accuracy sebesar 70% , precision 53%, recall 67%, f1_score 53%. Yang menunjukkan kemampuan yang layak dalam mengklasifikasikan data. Meskipun masih ada ruang untuk peningkatan, secara keseluruhan, performa model ini dapat dikatakan cukup baik dalam melakukan analisis sentimen.

	precision	recall	f1-score	support
Negative	0.08	0.50	0.13	6
Neutral	0.97	0.67	0.79	206
Positive	0.53	0.85	0.65	53
accuracy			0.70	265
macro avg	0.53	0.67	0.53	265
weighted avg	0.86	0.70	0.75	265

Gambar 8. Hasil evaluasi

D. Simpulan

Hasil penelitian menunjukkan bahwa :

1. Dari total 597 ulasan yang didapatkan, data dibagi menjadi 3 kelompok berdasarkan sentimennya yaitu data dengan sentimen positif yang berjumlah 207 ulasan atau 34,7% yang mana bisa dikatakan aplikasi ini mendapat tanggapan yang cukup positif dari pengguna, data dengan setimen netral berjumlah 315 ulasan atau 52,6%, dan data dengan sentimen negatif berjumlah 75 ulasan atau 12,7% ini menunjukkan meskipun aplikasi ini sudah mendapatkan respon yang cukup baik namun butuh beberapa sedikit perbaikan untuk meningkatkan efektivitas layanan. Pelabelan sentimen dilakukan secara otomatis menggunakan NLP dengan library textblob. Dari sentimen positif dan negatif terdapat beberapa kata yang sering muncul, pada setimen positif terdapat kata “bagus”,

- “bermanfaat”, “mudah” yang mengindikasikan kepuasan pengguna atas aplikasi inarisk yang bagus, bermanfaat, dan mudah digunakan. Sedangkan dari sentimen negatif terdapat kata “lemot”, “lambat”, “error”, “bingung” mengindikasikan masih adanya ketidakpuasan pengguna menggunakan aplikasi inarisk dimana pengguna menilai aplikasi masih lambat dalam merespon, sering terjadi error, dan ada juga pengguna yang masih merasa bingung menggunakan inarisk. Dari hasil ini BNPB dapat menggunakan sebagai bahan evaluasi untuk meningkatkan kualitas dan efektivitas aplikasi inarisk dalam upaya penanggulangan bencana secara keseluruhan.
2. Dalam proses klasifikasi yang dilakukan menggunakan machine learning regresi logistik model berhasil mendapatkan akurasi 69,81%. Dalam analisis sentimen akurasi model ini dapat dikatakan cukup baik sehingga model ini dapat digunakan untuk manajemen dan mitigasi resiko bencana seperti evakuasi darurat dan pemberitahuan dini dengan memprediksi sentimen dari ulasan baru yang belum terlihat sebelumnya.

E. Ucapan Terima Kasih

Terima kasih kepada Universitas Tadulako dan semua yang telah mendukung penelitian ini.

F. Referensi

- [1] I. Santoso, S. I. Oktora, S. Muchlisoh, and E. Pasaribu, “Social Network Analysis untuk Identifikasi Pengguna Twitter Berpengaruh pada Topik Bencana Gempa dan Tsunami di Indonesia,” *J. Edukasi dan Penelit. Inform.*, vol. 9, no. 1, p. 115, 2023, doi: 10.26418/jp.v9i1.62211.
- [2] H. Firdaus and A. Sofro, “Analisa Cluster Menggunakan K-Means Dan Fuzzy C-Means Dalam Pengelompokan Provinsi Menurut Data Intesitas Bencana Alam Di Indonesia Tahun 2017-2021,” *MATHunesa J. Ilm. Mat.*, vol. 10, no. 1, pp. 50–60, 2022, doi: 10.26740/mathunesa.v10n1.p50-60.
- [3] BNPB, “Geoportal data bencana indonesia,” BNPB. Accessed: May 10, 2024. [Online]. Available: <https://gis.bnpb.go.id>
- [4] K. K. Aprilia, “Analisis Potensi Bahaya Pada Gudang Logistik Dan Peralatan Dengan Job Safety Analysis (Jsa) (Studi Kasus Di BNPB),” 2021, [Online]. Available: <https://library.universitaspertamina.ac.id//xmlui/handle/123456789/5774>
- [5] inaRISK, “InaRISK Personal,” BNPB. Accessed: May 10, 2024. [Online]. Available: <https://inarisk.bnpb.go.id/inariskapps>
- [6] Y. Cahyono and S. Saprudin, “Analisis Sentiment Tweets Berbahasa Sunda Menggunakan Naive Bayes Classifier dengan Seleksi Feature Chi Squared Statistic,” *J. Inform. Univ. Pamulang*, vol. 4, no. 3, p. 87, 2019, doi: 10.32493/informatika.v4i3.3186.
- [7] N. Munasatya and S. Novianto, “Natural Language Processing untuk Sentimen Analisis Presiden Jokowi Menggunakan Multi Layer Perceptron,” *Techno.Com*, vol. 19, no. 3, pp. 237–244, 2020, doi: 10.33633/tc.v19i3.3630.
- [8] D. Hernikawati Balai Pengembangan Sumber Daya Manusia dan Penelitian Kominfo Jakarta, J. Pegangsaan Timur No, and J. Pusat, “Kecenderungan

- Tanggapan Masyarakat Terhadap Vaksin Sinovac Berdasarkan Lexicon Based Sentiment Analysis The Trend of Public Response to Sinovac Vaccine Based on Lexicon Based Sentiment Analysis," *J. Ilmu Pengetah. dan Teknol. Komun.*, vol. 23, no. 1, pp. 21–31, 2021, [Online]. Available: <http://dx.doi.org/10.33169/iptekkom.23.1.2021.21-31>
- [9] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.
- [10] R. Azhar, A. Surahman, and C. Juliane, "Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naïve Bayes," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 1, pp. 267–281, 2022.