

Analisis Sentimen Aplikasi Zenius Menggunakan Metode Logistic Regression

I Made Juniandika^{a1}, Ida Bagus Made Mahendra^{a2}

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali Indonesia
¹juniandika7072@gmail.com
²ibm.mahendra@unud.ac.id

Abstract

E-Learning application is one of the media used to conduct adaptive learning. One of the e-learning applications that is widely used is Zenius. To know the success of an application, sentiment analysis technique is needed. In this research, the data used to perform sentiment analysis is Zenius application user review data taken from Google Play Store with a total of 19892 review data with a score of 1,2,3, and 5. The data is classified using the logistic regression method with two classes of labels, namely positive and negative. The accuracy result obtained using this method is 90% with the classification results of positive labeled data as much as 51.17% and negative labeled data as much as 48.83%.

Keywords: *E-Learning, Sentiment Analysis, Zenius, Classifier, Logistic Regression, Accuracy*

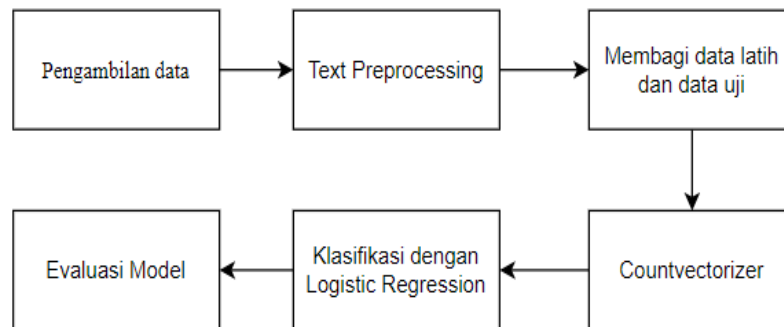
1. Pendahuluan

Kemajuan teknologi informasi memiliki dampak yang sangat besar bagi kehidupan masyarakat, salah satunya yaitu perkembangan pada bidang pendidikan. Perkembangan teknologi informasi menuntut dunia pendidikan untuk dapat mengikuti kemajuan teknologi dengan membuat metode pembelajaran yang adaptif dengan memanfaatkan berbagai media sehingga dapat meningkatkan kualitas hasil belajar. Aplikasi e-learning merupakan salah satu media yang dapat digunakan untuk melakukan proses pembelajaran secara elektronik dengan memanfaatkan internet. Ketersediaan aplikasi e-learning di Indonesia saat ini mengalami perkembangan yang pesat mulai dari yang berbayar hingga tidak berbayar, salah satunya adalah aplikasi Zenius. Zenius merupakan aplikasi yang menyediakan pembelajaran online berbayar yang diciptakan untuk memberikan kemudahan akses belajar bagi pelajar dari kalangan Sekolah Dasar sampai dengan Sekolah Menengah Atas. Zenius menawarkan layanan berupa akses video pembelajaran, forum diskusi, serta ujian online. Saat ini aplikasi Zenius sudah tersedia dalam bentuk web dan mobile sehingga dapat memudahkan siswa untuk mengakses materi pembelajaran di mana saja dan kapan saja. Untuk berbasis mobile, pengguna dapat mengunduh aplikasi secara gratis di google play store. Berdasarkan situs pada google play pada tanggal 23 April 2023, aplikasi Zenius telah diunduh kurang lebih sebanyak 5 juta unduhan, dengan rating 4.6, serta ulasan sebanyak 79.9 ribu ulasan [1]. Saat ini, Zenius bukanlah satu-satunya aplikasi yang menyediakan tawaran pembelajaran online yang tersedia di Indonesia. Terdapat aplikasi lainnya yang menawarkan layanan serupa dengan aplikasi Zenius seperti Ruangguru dan Quipper. [2] Dengan banyaknya aplikasi serupa yang tersedia membuat pengguna untuk lebih selektif dalam menentukan aplikasi yang baik dan cocok untuk digunakan. Ulasan aplikasi pada laman unduhan menjadi salah satu faktor yang memengaruhi pengguna dalam memilih aplikasi karena dijadikan sebagai bahan pertimbangan dalam menilai suatu aplikasi [3]. Ulasan-ulasan tersebut nantinya akan diolah dengan menggunakan teknik analisis sentimen. Teknik analisis sentimen ini akan sangat membantu dalam melakukan pengembangan aplikasi, hal ini dikarenakan analisis sentimen akan mengklasifikasi ulasan-ulasan positif, negatif, dan netral yang nantinya dapat dijadikan sebagai tolak ukur bagi developer dalam meningkatkan performa aplikasi. Pada tahun 2021, hal ini sudah pernah diteliti oleh Nina Ismaya Pangaribuan dkk dengan jurnal yang berjudul "Analisis Sentimen Aplikasi E-Learning Selama Pandemi Covid-19 Dengan Menggunakan Metode Support Vector

Machine dan Convolutional Neural Network” yang bertujuan untuk melakukan evaluasi pada peningkatan hal positif dan memperbaiki hal negatif dari aplikasi e-learning salah satunya yaitu aplikasi Zenius [4]. Penelitian tersebut mengambil ulasan data aplikasi Zenius sebanyak 17.350 [4]. Hasil akurasi yang diperoleh dari aplikasi Zenius dengan menggunakan metode SVM yaitu sebesar 76% sedangkan dengan metode CNN menghasilkan akurasi sebesar 84%. Perbedaan penelitian ini dengan penelitian sebelumnya yaitu metode yang digunakan. Pada penelitian ini, penulis melakukan analisis sentimen aplikasi Zenius menggunakan metode logistic regression. Hasil akhir dari penelitian ini yaitu tingkat persentase sentimen positif dan negatif ulasan komentar pengguna yang dihasilkan dengan metode logistic regression

2. Metode Penelitian

Metode penelitian yang digunakan pada penelitian ini menggunakan metode eksperimen dengan urutan langkah-langkah seperti gambar di bawah ini.



Gambar 1. Langkah-langkah Penelitian

2.1 Pengambilan data

Dataset yang digunakan dalam penelitian ini yaitu ulasan komentar pengguna aplikasi Zenius dari tanggal 03 Maret 2019 sampai dengan 26 April 2023 pada google play store yang diambil dengan teknik scrapping. Ulasan yang diambil adalah ulasan berbahasa Indonesia dengan bintang satu, dua, dan tiga yang mempresentasikan ulasan negatif dan bintang lima yang mempresentasikan ulasan positif. Hasil dari data yang diperoleh berjumlah 19892 data dengan label positif sebanyak 16356 data dan label negatif sebanyak 3536 data.

2.2 Preprocessing

Tahap selanjutnya setelah melakukan pengambilan data yaitu melakukan preprocessing data yang diimplementasikan dengan menggunakan bahasa pemrograman python. Adapun tahapan-tahapan dalam melakukan preprocessing data adalah sebagai berikut:

a. Missing value and duplicated removal

Pada tahapan ini dilakukan pembersihan data berupa penghapusan data duplikat dan menghilangkan missing value setelah proses ini selesai maka diperoleh data bersih sebanyak 16281 data dengan jumlah data berlabel positif sebanyak 12856 data dan berlabel negatif sebanyak 3425.

b. Case folding

Case folding merupakan proses pembersihan data teks dengan mengubah semua bentuk kata menjadi sama seperti mengubah ke dalam lowercase.

c. Punctuations, Special Character, and Emoticon removal

Penghilangan punctuation pada teks yang meliputi tanda baca serta simbol-simbol yang tidak penting seperti "#\$%&'()*+,-./:;<=>?@[\\]^_`{|}~", penghapusan special karakter seperti angka-angka yang terdapat dalam teks, serta penghapusan emoji.

d. Stopwords removal

Stopwords removal merupakan proses pembersihan kata-kata yang dianggap tidak relevan dengan dokumen sehingga dapat meningkatkan efektivitas pemrosesan data. Dalam tahapan ini, kata-kata yang tidak penting serta tidak memberikan kontribusi signifikan terhadap makna dokumen akan dihilangkan

e. Slang words

Slang words adalah tahapan untuk mengubah kata-kata yang mengalami kesalahan ejaan dalam penulisan ke dalam bentuk yang baku.

f. Stemming

Stemming adalah proses untuk mengubah kata ke dalam bentuk dasarnya sehingga dapat mengurangi variasi kata dalam dokumen. Tahapan ini akan menormalisasikan teks dengan menghilangkan imbuhan berupa awalan, akhiran dan sisipan dalam teks.

g. Oversampling

Oversampling adalah sebuah teknik untuk mengatasi ketidakseimbangan kelas label dalam suatu dataset dengan cara menambahkan sampel kepada kelas label yang jumlahnya sedikit. Pada penelitian ini, jumlah data kelas negatif hanya sebanyak 3425 data sehingga diperlukan teknik oversampling untuk kelas berlabel negatif sehingga jumlah data antara label positif dan negatif menjadi seimbang.

2.3 Pembagian data

Pada tahapan ini, dilakukan pembagian data menjadi data latih dan data uji dengan perbandingan 80 :20. Data latih nantinya digunakan untuk melatih model machine learning sedangkan data uji untuk menguji kinerja dari model. Pembagian proporsi yang tepat pada data latih dan data uji akan membantu dalam menciptakan model yang akurat dalam memprediksi kelas label pada data baru.

2.4 Countvectorizer

Countvectorizer adalah sebuah teknik untuk mengubah teks menjadi representasi numerik dengan menghitung frekuensi kemunculan kata dalam setiap dokumen tanpa memperhatikan urutan kata. Pada tahapan ini, setiap huruf dalam kata akan sudah mengalami proses tokenisasi.

2.5 Klasifikasi dengan Logistic Regression

Pada tahapan ini, peneliti menggunakan metode logistic regression untuk melakukan analisis sentimen yang diimplementasikan dengan bahasa pemrograman python. Tujuan digunakan metode ini yaitu untuk mendapatkan model sederhana dan terbaik untuk menjelaskan hubungan antara output dari variabel respon (Y) dengan variabel prediktornya (X) [5].

Rumus Logistic Regression:

$$\sum_{i=1}^m \frac{y_{\text{pred}}(i) - y_{\text{test}}(i)}{M}, \text{ dimana } m = \text{jumlah dataset} \quad (1)$$

2.6 Evaluasi

Pada tahapan ini, peneliti menggunakan confusion matrix untuk mengetahui performa klasifikasi model.



3. Hasil dan Diskusi

Proses pengambilan data dalam penelitian ini dilakukan dengan teknik scrapping dengan bahasa pemrograman python menggunakan library google-play-scraper dari ulasan aplikasi Zenius pada google play store.

	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt
0	591c5fb1-059e-456d-8c06-8887f6b6c8cf	Juwita Anggi Pratiwi	lh.googleusercontent.com/a-/ACB-R...	masih tahap awal belajar lewat zenius dan udh ...	5	6	2.8.9	2023-04-07 10:31:28	None	NaT
1	7a96c37b-53a0-4a75-b60a-d803577c3f2b	Oktanito Yeremia	lh.googleusercontent.com/a-/ACB-R...	kenapa size data appnya besar sekali? bukannya...	3	1	2.8.9	2023-04-16 20:08:02	Halo, terimakasih atas kritik dan saran dari k...	2023-04-17 12:15:20
2	6442a915-e7d3-4a7a-b760-b51e5289c033	Dany AkmallunN	lh.googleusercontent.com/a-/ACB-R...	Aplikasinya banyak bug jadi ga mood belajarnya...	1	60	2.8.7	2023-02-23 12:04:00	Hai, mohon maaf atas kendala yang kamu temui belajarnya...	2023-02-26 19:01:24

Gambar 3. Hasil Scrapping Data

Setelah proses scrapping data selesai, maka kita seleksi fitur score dengan mengambil score bernilai 1,2,3 yang nantinya mewakili ulasan negatif dan 5 mewakili ulasan positif lalu disimpan dalam file bernama "Zenius.csv".

content	score	label
masih tahap awal belajar lewat zenius dan udh ...	5	0
Aplikasinya banyak bug jadi ga mood belajarnya...	1	1
Aplikasi nya bagus, video nya mudah di mengert...	5	0
Dari sekian banyaknya apk belajar,nih aplikasi...	1	1
zenius masih banyak bug, kadang kadang keluar ...	1	1
...
	5	0
	5	0
Gud Gud Gud	5	0

Gambar 4. Seleksi Fitur Score

3.2 Text Preprocessing

Sebelum data digunakan dalam model, maka data harus mengalami preprocessing terlebih dahulu. Pada tahapan ini dilakukan case folding ke dalam lowercase

	content	label	content_lower
	masih tahap awal belajar lewat zenius dan udh ...	0.0	masih tahap awal belajar lewat zenius dan udh ...
	kenapa size data appnya besar sekali? bukannya...	1.0	kenapa size data appnya besar sekali? bukannya...
	Aplikasinya banyak bug jadi ga mood belajarnya...	1.0	aplikasinya banyak bug jadi ga mood belajarnya...
	Saat pernah mendapatkan soal terkait tabel dat...	1.0	saat pernah mendapatkan soal terkait tabel dat...
	Aplikasi nya bagus, video nya mudah di mengert...	0.0	aplikasi nya bagus, video nya mudah di mengert...

Gambar 5. Hasil Convert Lowercase

Selanjutnya melakukan penghapusan punctuation dengan memanfaatkan library string, special character, serta penghapusan emoticon.

	content	label	content_clean
0	masih tahap awal belajar lewat zenius dan udh ...	0.0	masih tahap awal belajar lewat zenius dan udh ...
1	kenapa size data appnya besar sekali bukannya ...	1.0	kenapa size data appnya besar sekali bukannya ...
2	aplikasinya banyak bug jadi ga mood belajarnya...	1.0	aplikasinya banyak bug jadi ga mood belajarnya...
3	saat pernah mendapatkan soal terkait tabel dat...	1.0	saat pernah mendapatkan soal terkait tabel dat...
4	aplikasi nya bagus video nya mudah di mengerti...	0.0	aplikasi nya bagus video nya mudah di mengerti...

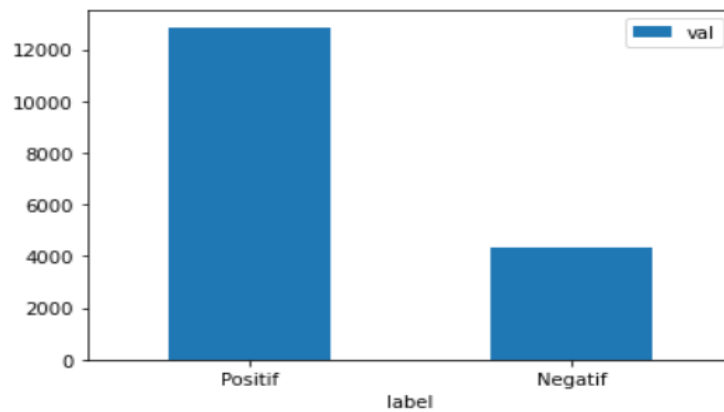
Gambar 6. Hasil Penghapusan Punctuation, Special Character, dan Emoticon

Tahap selanjutnya yaitu melakukan penghapusan stopwords untuk mengurangi kata-kata yang tidak relevan terhadap dokumen, contoh kata yang termasuk ke dalam stopwords bahas indonesia adalah “di”, “ke”, “dari”, “dimana”, “kemana”, dan lain sebagainya. Setelah melakukan penghapusan stopwords, tahap berikutnya yaitu melakukan slang word untuk mengubah kesalahan ejaan dalam penulisan kalimat seperti “7an : tujuan”, “abis : habis”, “aj : saja”, dan lain sebagainya. Dan yang terakhir adalah melakukan stemming untuk mengubah bentuk kata ke dalam bentuk dasarnya seperti “mengajar: ajar”, “memberikan: beri”, dan lain sebagainya.

	content	label
0	tahap ajar zenius udh langgan premium member n...	0.0
1	size data appnya platform bas streaming data v...	1.0
2	aplikasi bug ga mood ajar video kadang henti a...	1.0
3	kait tabel data foto tabel lengkap ngilangin p...	1.0
4	aplikasi nya bagus video nya mudah erti bahas ...	0.0

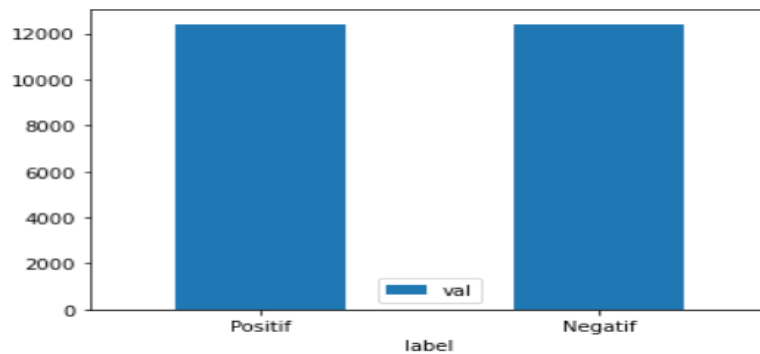
Gambar 7. Hasil Preprocessing Akhir

Setelah itu dilanjutkan dengan tahapan sampling data, karena jumlah data berlabel positif tidak seimbang dengan jumlah data yang berlabel negatif.



Gambar 8. Jumlah Data Tiap Label

Dari gambar di atas terlihat bahwa data pada kelas label tidak seimbang. Data dengan label positif berjumlah 12856 data sedangkan untuk yang berlabel hanya 5324 data sehingga perlu dilakukan oversampling pada data yang berlabel negatif sehingga jumlahnya menjadi sama dengan data berlabel positif



Gambar 9. Jumlah Data Setelah Sampling Data

3.3 Klasifikasi dengan Logistic Regression

Sebelum memasuki pemodelan dengan logistic regression, data harus dibagi menjadi data latih dan data uji. Perbandingan data uji dan data latih yang digunakan adalah 80:20. Tujuan dari data latih yaitu untuk melatih model sedangkan data uji nantinya digunakan sebagai data yang akan diprediksi. Setelah melakukan pembagian data, maka tahapan selanjutnya yaitu melakukan countvectorizer. yang nantinya di dalam tahapan ini setiap kata dalam kalimat akan mengalami proses tokenisasi kemudian akan dirubah ke numerik dalam bentuk vektor. Setelah mengalami proses countvectorizer maka masuk ke tahapan pemodelan dengan menggunakan metode logistic regression. Input yang dibutuhkan adalah hasil feature extraction X yang kemudian menghasilkan output prediksi y dengan nilai antar 0 atau 1. Jika nilai y prediksi lebih dari sama dengan 0.5 maka masuk kategori label positif (1) dan sebaliknya.

3.4 Evaluasi

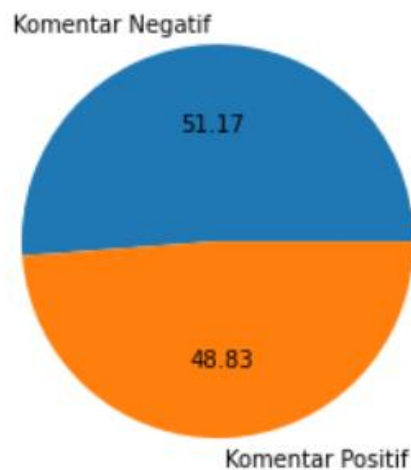
Tahapan terakhir dari pemodelan logistic regression adalah evaluasi. Evaluasi logistic regression dalam penelitian ini yaitu menggunakan confusion matrix

	precision	recall	f1-score	support
0	0.90	0.90	0.90	2434
1	0.90	0.91	0.90	2530
accuracy			0.90	4964
macro avg	0.90	0.90	0.90	4964
weighted avg	0.90	0.90	0.90	4964

Gambar 10. Hasil Confusion Matrix

Dari gambar diatas dapat dilihat bahwa nilai precision, recall, dan f-1 score untuk label kelas 1 dan label kelas 0 berada di atas 90%, yang menandakan bahwa model logistic regression ini baik dalam melakukan prediksi kelas label. Hasil akurasi yang diperoleh yaitu 90%.

3.5 Hasil



Gambar 11. Hasil Prediksi Data Tes

Berdasarkan gambar, model menghasilkan prediksi label pada data tes dengan jumlah label negatif sebesar 51.17% dan label positif sebesar 48.83%

4. Kesimpulan

Berdasarkan hasil pembahasan di atas, maka disimpulkan bahwa metode logistic regression menghasilkan prediksi yang baik dalam melakukan proses klasifikasi pada ulasan aplikasi Zenius pada google play store. Untuk melakukan pengujian, data hasil scrapping dilakukan preprocessing terlebih dahulu untuk menghasilkan data yang bersih kemudian data hasil preprocessing dibagi menjadi data latih dan data testing dengan perbandingan 80:20. Tahapan selanjutnya mengubah data tersebut ke dalam bentuk vector supaya bisa dilatih pada model logistic regression. Hasil dari pelatihan model pada data uji menunjukkan bahwa pada kelas 0, nilai precision, recall, dan f-1 score yang dihasilkan sebesar 90%, sedangkan untuk kelas 1 nilai precision dan f-1 score yang dihasilkan sebesar 90% dan recall sebesar 91%. Hasil akurasi dari

model logistic regression yaitu sebesar 90% yang menandakan bahwa model baik digunakan untuk melakukan kalsifikasi pada teks. Saran untuk penelitian selanjutnya yaitu diharapkan menggunakan jumlah dataset yang lebih banyak dalam pelatihan model serta mencoba metode baru untuk melakukan klasifikasi data teks seperti penggunaan metode deep learning.

Daftar Pustaka

- [1] "Zenius - #GantiCaraBelajar - Aplikasi di Google Play." <https://play.google.com/store/apps/details?id=net.zenius.mobile&hl=id&gl=US> (accessed May 07, 2023).
- [2] M. R. Firdaus, F. M. Rizki, F. M. Gaus, and I. K. Susanto, "Analisis Sentimen Dan Topic Modelling Dalam Aplikasi Ruangguru," *J-SAKTI (Jurnal Sains Komput. dan Inform.,* vol. 4, no. 1, p. 66, 2020, doi: 10.30645/j-sakti.v4i1.188.
- [3] E. Fitri, Y. Yuliani, S. Rosyida, and W. Gata, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *J. Transform.,* vol. 18, no. 1, pp. 71–80, Jul. 2020, Accessed: Apr. 25, 2023. [Online]. Available: <https://journals.usm.ac.id/index.php/transformatika/article/view/2317>.
- [4] A. S. Simbolon, N. I. Pangaribuan, and N. M. Aruan, "Analisis Sentimen Aplikasi E-Learning Selama Pandemi Covid-19 Dengan Menggunakan Metode Support Vector Machine Dan Convolutional Neural Network," *Seminastika,* vol. 3, no. 1, pp. 16–25, 2021, doi: 10.47002/seminastika.v3i1.236.
- [5] K. Kelvin, J. Banjarnahor, E. I. -, and M. NK Nababan, "Analisis perbandingan sentimen Corona Virus Disease-2019 (Covid19) pada Twitter Menggunakan Metode Logistic Regression Dan Support Vector Machine (SVM)," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA),* vol. 5, no. 2, pp. 47–52, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2365.