



Penerapan Machine Learning Pada Analisis Sentimen Aplikasi MyTelkomsel Menggunakan Data Ulasan Google Playstore

Farin Junita Fauzan¹, M. Afdal², Rice Novita³, Mustakim⁴

12050323918@students.uin-suska.ac.id, m.afdal@uin-suska.ac.id, rice.novita@uin-suska.ac.id, mustakim@uin-suska.ac.id

Universitas Islam Negeri Sultan Syarif Kasim Riau

Informasi Artikel

Diterima : 13 Mei 2024
Direview : 17 Jun 2024
Disetujui : 30 Jun 2024

Kata Kunci

MyTelkomsel, support vector machine, random forest, analisis sentimen, ulasan

Abstrak

Aplikasi MyTelkomsel merupakan sebuah platform akses digital yang menyediakan layanan terhadap telekomunikasi, Oleh karena itu, analisis sentimen terhadap pengguna Aplikasi MyTelkomsel menjadi relevan untuk mendapatkan wawasan yang berharga bagi pengembang dan pengelolaan aplikasi. Penelitian ini bertujuan untuk melakukan analisis sentimen dan mengetahui perbandingan metode pada data ulasan terhadap Aplikasi Mytelkomsel. Dataset yang digunakan terbagi menjadi dua sesuai topik ulasan yaitu pelayanan dan pengguna. Metode pelabelan penelitian ini menggunakan Lexicon Based dan Pakar Bahasa Indonesia dengan tiga kelas yaitu positif, negatif dan netral. Dataset ulasan yang dilabeli diterapkan dengan metode SVM dan Random Forest. Dimana hasil yang didapatkan dari menerapkan dua dataset dengan dua pelabelan yaitu pendekatan oleh pakar cenderung lebih akurat dibandingkan pendekatan lexicon based karena akurasi tertinggi lexicon based sebesar 79% sedangkan pelabelan Pakar sebesar 83%. Selain itu, dalam penelitian ini, algoritma SVM menunjukkan akurasi tertinggi, yakni 83%, pada dataset pengguna yang dianalisis oleh Pakar Bahasa Indonesia.

Keywords

MyTelkomsel, support vector machine, random forest, sentiment analysis, reviews

Abstract

The MyTelkomsel application is a digital access platform that provides telecommunications services. Therefore, sentiment analysis of MyTelkomsel application users is relevant for obtaining valuable insights for application development and management. This research aims to conduct sentiment analysis and compare methods on review data of the MyTelkomsel application. The dataset used is divided into two topics: service and user reviews. The labeling method in this research uses Lexicon Based and Indonesian Language Experts with three classes: positive, negative, and neutral. The labeled review dataset is then applied with SVM and Random Forest methods. The results obtained from applying two datasets with two labeling approaches indicate that the approach by experts tends to be more accurate compared to the lexicon-based approach because the highest accuracy of the lexicon-based approach is 79%, while the expert labeling achieves an accuracy of 83%. Additionally, in this study, the SVM algorithm demonstrates the highest accuracy, namely 83%, on the user dataset analyzed by Indonesian Language Experts.

A. Pendahuluan

Perkembangan teknologi informasi kini semakin pesat, terlihat dari semakin meluasnya penggunaan smartphone di masyarakat. Salah satu contoh dari kemajuan teknologi dalam bidang komunikasi adalah Telekomunikasi. Telekomunikasi adalah proses pengiriman informasi dari satu lokasi ke lokasi lainnya dengan menggunakan teknologi informasi. Saat ini ada empat operator seluler yang beroperasi di Indonesia yaitu Telkomsel sebagai anak usaha PT Telkom Indonesia Tbk. (TLKM), Indosat Ooredoo Hutchison (ISAT) sebagai perusahaan hasil merger Indosat dan Tri, PT XL Axiata Tbk. (EXCL), serta PT Smartfren Telecom Tbk. (FREN) [1]. Sebuah laporan dari Ookla Speedtest Intelligence Telkomsel berada di posisi pertama mengenai penyedia internet seluler dengan kecepatan tertinggi di Indonesia selama kuartal keempat tahun 2023 [2].

Telkomsel menyediakan aplikasi MyTelkomsel gratis di Google Play Store. Hingga September 2023, sudah lebih dari 50 juta unduhan dan lebih dari 35 juta monthly active users. Pertumbuhan transaksi aplikasi ini mencapai sekitar 28% dibandingkan tahun sebelumnya [3]. MyTelkomsel adalah layanan self-service dengan berbagai fitur, termasuk pembelian pulsa, paket kuota internet, Indihome, paket menonton seperti Maxstream dan Prime Video Mobile, serta layanan customer service online, dll.

Meskipun MyTelkomsel menawarkan banyak fitur yang memudahkan pelanggan, aplikasi ini masih menerima keluhan dan respon negatif di Google Play Store. Keluhan tersebut meliputi lamanya proses login, kesulitan masuk ke menu utama, kegagalan transaksi, serta ketidakpuasan terhadap customer service online lainnya. Keluhan ini dapat berdampak buruk pada kepuasan dan loyalitas pelanggan.

Seiring waktu, pengguna sering menyampaikan pendapat dan opini mereka melalui ulasan di aplikasi MyTelkomsel di Google Play. Opini yang diungkapkan dalam komentar mencakup berbagai topik, namun peneliti membaginya menjadi dua topik utama: pelayanan aplikasi dan pengguna aplikasi. Pelayanan aplikasi mencakup layanan atau jasa yang ditawarkan kepada pengguna, dengan tujuan memenuhi kebutuhan dan keinginan mereka serta membantu dalam tindakan atau tugas yang dilakukan. Sementara itu, pengguna aplikasi dalam konteks penelitian ini merujuk pada respons dari individu atau organisasi terhadap aplikasi itu sendiri.

Komentar pengguna MyTelkomsel dapat digunakan sebagai evaluasi untuk meningkatkan aplikasi.[4]. Analisis sentimen membantu mengidentifikasi sentimen positif, negatif, dan netral dalam komentar, memungkinkan perusahaan untuk memahami pandangan pengguna dan membuat perbaikan yang sesuai. [5].

Penelitian sebelumnya [6] tentang perbandingan algoritma klasifikasi menunjukkan bahwa Random Forest memiliki akurasi tertinggi (94%), diikuti oleh Support Vector Machine (91%), dan Logistic Regression (86%). Penelitian lainnya [7][8][9][10] juga menunjukkan bahwa Random Forest dan Support Vector Machine selalu memberikan hasil akurasi tinggi dalam analisis sentimen.

Penelitian ini bertujuan membandingkan kinerja SVM dengan metode klasifikasi lain seperti Random Forest dalam analisis sentimen ulasan berbahasa

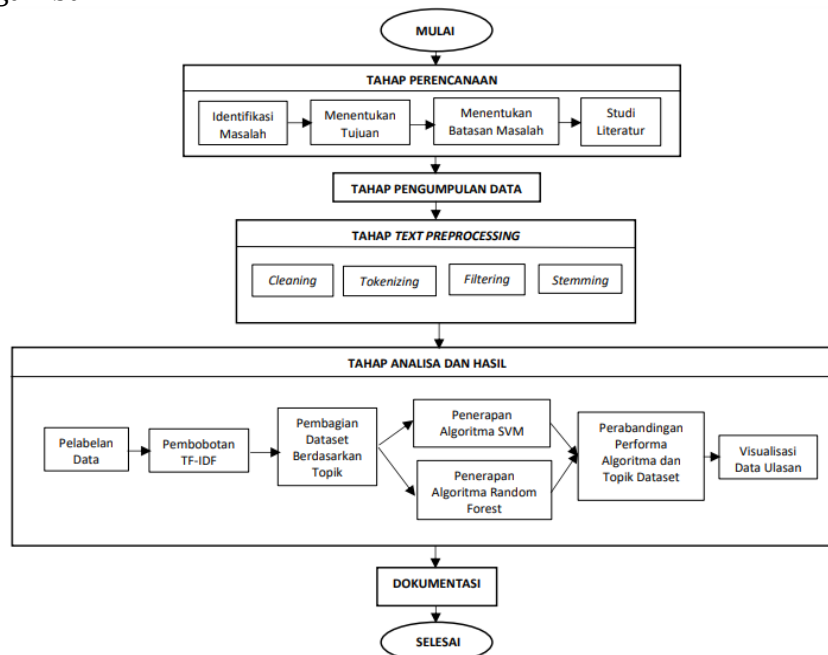
Indonesia di Google Play Store. SVM dan Random Forest dikenal efektif dalam menangani data besar dan data dari berbagai media.

Sedangkan pada penelitian sebelumnya [11][12][13][14][15] menunjukkan bahwa metode lexicon based memiliki akurasi tinggi. Metode ini menggunakan kamus dengan sentimen terkait untuk analisis opini. Pendekatan lexicon-based secara otomatis mengklasifikasikan teks opini dengan menguji data latih berupa teks opini yang telah diklasifikasikan secara manual. Berdasarkan hasil tersebut, penelitian ini akan menggunakan dua pendekatan: lexicon based dan pendekatan oleh pakar bahasa Indonesia.

Berdasarkan permasalahan latar belakang yang diuraikan diatas, maka diperoleh rumusan masalah yaitu bagaimana penerapan machine learning pada analisis sentimen aplikasi mytelkomsel menggunakan data ulasan google playstore menggunakan algoritma svm dan random forest.

B. Metode Penelitian

Adapun metodologi penelitian yang digunakan pada penelitian ini dapat dilihat pada gambar :



Gambar 1. Metodologi Penelitian

1. Tahap Perencanaan

a. Identifikasi Masalah

Mengamati dan menemukan permasalahan yang terjadi mengenai Aplikasi MyTelkomsel berdasarkan opini masyarakat dari ulasan google playstore.

b. Menentukan Tujuan

Memilih metode yang tepat serta algoritma yang akan digunakan untuk mencapai hasil yang diinginkan.

c. Menentukan Batasan Masalah

Agar penelitian tidak keluar dari pembahasan dari cakupan objek penelitian.

d. Studi Literatur

Sumber studi mencakup publikasi ilmiah seperti jurnal, skripsi atau tesis, buku, serta sumber-sumber lain yang tervalidasi.

2. Tahap Pengumpulan data

Data yang akan digunakan dalam penelitian ini adalah ulasan Aplikasi MyTelkomsel yang diambil dari Google Play Store. Teknik yang digunakan untuk mengumpulkan data ulasan Aplikasi MyTelkomsel di Google Play Store adalah scraping menggunakan salah satu tool dalam bahasa pemrograman Python, yaitu google-play-scraper. Setelah proses scraping selesai, data yang terkumpul akan dikonversikan ke format yang lebih terstruktur menggunakan Microsoft Excel.

3. Tahap Text Pre-processing

Agar machine learning dapat berfungsi dengan baik, preprocessing data harus dilakukan. Ini melibatkan penghapusan kata penghubung, tanda titik koma, dan kata-kata lain yang tidak relevan dari informasi penelitian. Berikut pemaparan tahapan text preprocessing yang digunakan.

a. Cleaning

Bertujuan untuk membuang tanda baca, angka, dan juga emoticon

b. Tokenizing

Bertujuan memisahkan kata dari kalimat menjadi bentuk satuan kata (token)

c. Filtering

Pengambilan kata-kata penting dari hasil tokenisasi serta penghilangan kata umum dan tidak memiliki makna seperti kata hubung dan kata depan.

d. Stemming

Mengestrak kata ke dalam bentuk kata dasar atau dapat diartikan sebagai penghapusan imbuhan pada kata.

4. Tahap Analisa dan Hasil

1. Pelabelan Data

a. Pelabelan Data menggunakan Lexicon Based

Pendekatan Lexicon Based diterapkan menggunakan Kamus InSet untuk memberi label/kelas pada data[16].

b. Pelabelan Data oleh Pakar

Dilakukan juga pelabelan data komentar oleh Pakar Bahasa Indonesia untuk melabelkan data

2. Pembobotan kata menggunakan TF-IDF

Tahap dilakukannya pembobotan setiap kata dengan menggunakan Term Frequency dan Inverse Document Frequency.

3. Pembagian Dataset berdasarkan Topik

Dataset terbagi juga menjadi 2 dataset yaitu dataset ulasan terkait pelayanan aplikasi MyTelkomsel dan Ulasan pengguna terkait aplikasi MyTelkomsel.

4. Penerapan Algoritma SVM

Dalam menerapkan Algoritma SVM baik pada data pelabelan lexicon based maupun data pelabelan pakar, penentuan jumlah dan nilai parameter adalah kunci penting.

5. Penerapan Algoritma Rndom Forest

Pada penerapan Algoritma Random Forest baik pada data pelabelan lexicon based maupun data pelabelan pakar, ditentukan jumlah dan nilai dari parameter.

6. Perbandingan Performa Algoritma dan Topik Dataset

Untuk menentukan algoritma dengan performa terbaik, pengukurannya dilakukan menggunakan akurasi. Kemudian diperoleh perbandingan akurasi antara Algoritma SVM dan Algoritma Random Forest yang diterapkan pada data pelabelan berbasis leksikon dan data pelabelan pakar, serta perbandingan antara dataset yang telah dipisahkan berdasarkan topik.

7. Visualisasi Data Komentar

Visualisasi kata ditampilkan dalam bentuk word cloud berdasarkan tingkat frekuensi kata yang sering muncul pada setiap kelas dalam dataset.

5. Dokumentasi

Tahapan terakhir dari seluruh kegiatan penelitian.

C. Hasil dan Pembahasan

1. Pengumpulan Data

Berikut merupakan data awal dapat dilihat pada Tabel 1 :

Tabel 1. Data Awal

No	Ulasan
1.	Mantap
2.	Ganggu banget sih sms operator. Lg main game sms mulu dari 5111, 3636, hadiah coin, 777, +2000 coin. Ganggu woy sinyal jdi lambat klo ada sms
...	...
2331	Apk nya lemotnya masyallah . Payah

2. Tahap *Text Pre-processing*

a. Cleaning

Bertujuan untuk membuang tanda baca, angka, dan juga emoticon. Hasil cleaning data dapat dilihat pada Tabel 2 :

Tabel 2. Cleaning Data

No	Teks Cleaning
1.	mantap
2.	ganggu banget sih sms operator lg main game sms mulu dari hadiah coin coin ganggu woy sinyal jdi lambat klo ada sms
...	...
2331	apk nya lemotnya masyallah payah

b. Tokenizing

Memecahkan kalimat menjadi penggalan kata (token). Pada tahap ini akan dilakukan pengambilan minimal 3 kata pada setiap komentar[17], untuk mengetahui makna dalam sebuah komentar.

Setelah dilakukan tahapan tersebut didapatkanlah 1758 data komentar. Hasil Tokenizing dapat dilihat pada Tabel 3 :

Tabel 3. Data Tokenizing

No	Teks	Token
1	ganggu banget sih sms operator lg main game sms mulu dari hadiah coin coin ganggu woy sinyal jdi lambat klo ada sms	['ganggu', 'banget', 'sih', 'sms', 'operator', 'lg', 'main', 'game', 'sms', 'mulu', 'dari', 'hadiah', 'coin', 'coin', 'ganggu', 'woy', 'sinyal', 'jdi', 'lambat', 'klo', 'ada', 'sms']
2	padahal paket kuota internet ada knp pulsa utama bisa berkurang yg dipakai wa fb dan ig masa iya pulsa ditarik juga saat main sosmed padahal kuota internet ada	['padahal', 'paket', 'kuota', 'internet', 'ada', 'knp', 'pulsa', 'utama', 'bisa', 'berkurang', 'yg', 'dipakai', 'wa', 'fb', 'dan', 'ig', 'masa', 'iya', 'pulsa', 'ditarik', 'juga', 'saat', 'main', 'sosmed', 'padahal', 'kuota', 'internet', 'ada']
...
1758	apk nya lemotnya masyallah payah	['apk', 'nya', 'lemotnya', 'masyallah', 'payah']

c. Filtering

Dilakukan penghapusan kata-kata yang tidak memiliki makna secara signifikan. Hasil proses filtering data dapat dilihat pada Tabel 4 :

Tabel 4. Data Filtering

No	Teks	Filtering
1	ganggu banget sih sms operator lg main game sms mulu dari hadiah coin coin ganggu woy sinyal jdi lambat klo ada sms	['ganggu', 'banget', 'sih', 'sms', 'operator', 'lg', 'main', 'game', 'sms', 'mulu', 'hadiah', 'coin', 'coin', 'ganggu', 'woy', 'sinyal', 'jdi', 'lambat', 'klo', 'sms']
2	padahal paket kuota internet ada knp pulsa utama bisa berkurang yg dipakai wa fb dan ig masa iya pulsa ditarik juga saat main sosmed padahal kuota internet ada	['paket', 'kuota', 'internet', 'knp', 'pulsa', 'utama', 'berkurang', 'yg', 'dipakai', 'wa', 'fb', 'ig', 'iya', 'pulsa', 'ditarik', 'main', 'sosmed', 'kuota', 'internet']
...
1758	apk nya lemotnya masyallah payah	['apk', 'nya', 'lemotnya', 'masyallah', 'payah']

d. Stemming

Mengubah kata yang memiliki imbuhan pada kalimat komentar menjadi kata dasar. Hasil proses stemming dapat dilihat pada Tabel 5 :

Tabel 5. Data Stemming

No	Teks	Stemming
1	ganggu banget sih sms operator lg main game sms mulu dari hadiah coin coin ganggu woy sinyal jdi lambat klo ada sms	['ganggu', 'banget', 'sih', 'sms', 'operator', 'operator', 'lg', 'main', 'game', 'lg', 'main', 'game', 'sms', 'mulu', 'hadiah', 'sms', 'mulu', 'hadiah', 'coin', 'coin', 'coin', 'coin', 'ganggu', 'woy', 'sinyal', 'jdi', 'lambat', 'klo', 'sms']

2	['paket', 'kuota', 'internet', 'knp', 'puls', 'utama', 'berkurang', 'yg', 'dipakai', 'wa', 'fb', 'ig', 'iya', 'puls', 'ditarik', 'main', 'sosmed', 'kuota', 'internet']	['paket', 'kuota', 'internet', 'knp', 'puls', 'utama', 'kurang', 'yg', 'pakai', 'wa', 'fb', 'ig', 'iya', 'puls', 'tarik', 'main', 'sosmed', 'kuota', 'internet']
...
1758	['apk', 'nya', 'lemotnya', 'masyallah', 'payah']	['apk', 'nya', 'lot', 'masyallah', 'payah']

3. Pelabelan Data

Dataset ulasan dikategorikan ke dalam tiga kelas yaitu positif, negatif, dan netral. Pada penelitian ini, pelabelan dilakukan dengan dua pendekatan yaitu *Lexicon Based* dan Pakar Bahasa Indonesia.

a. Pelabelan Data menggunakan Lexicon Based

Hasil pelabelan data menggunakan Lexicon Based pada 1758 data ulasan pada Aplikasi MyTelkomsel, terdapat sentimen positif berjumlah 549, sentimen negatif 953 dan sentimen netral 256. Hasil pelabelan data dengan pendekatan LBF dapat dilihat pada Tabel 6 :

Tabel 6. Hasil Pelabelan Data Lexicon Based

No	Skor Sentimen	Sentimen	content
1	-12	negatif	ganggu banget sih sms operator lg main game sms mulu hadiah coin coin ganggu woy sinyal jdi lambat klo sms
2	-12	negatif	paket kuota internet knp puls utama kurang yg pakai wa fb ig iya puls tarik main sosmed kuota internet
...
1758	0	netral	apk nya lot masyallah payah

b. Pelabelan Data oleh Pakar

Pelabelan oleh Pakar Bahasa Indonesia, pada 1758 data ulasan terdapat sentimen positif berjumlah 192, sentimen negatif 1276 dan sentimen netral 290. Hasil pelabelan data dengan pendekatan pakar bahasa dapat dilihat pada Tabel 7 :

Tabel 7. Hasil Pelabelan Data Pakar

No	Sentimen	content
1	negatif	ganggu banget sih sms operator lg main game sms mulu hadiah coin coin ganggu woy sinyal jdi lambat klo sms
2	negatif	paket kuota internet knp puls utama kurang yg pakai wa fb ig iya puls tarik main sosmed kuota internet
...
1758	negatif	apk nya lot masyallah payah

4. Pembobotan Data Menggunakan TF-IDF

TF-IDF mengasumsikan bahwa beberapa term lebih sering muncul daripada yang lain dalam dokumen secara umum. Frekuensi kata dalam dokumen meningkatkan bobot kata tersebut, sementara IDF menurunkan

bobot term jika muncul sering dalam koleksi dokumen. [18]. Hasil pembobotan data dengan TF-IDF dapat dilihat pada Tabel 8 :

Tabel 8. Hasil Pembobotan Kata dengan TF-IDF

No	Term						
	aa	abal	abih	abis	...	zonk	zoom
1	0	0	0	0	...	0	0
2	0	0	0	0	...	0	0
...
1758	0	0	0	0	...	0	0

5. Pembagian Dataset berdasarkan Topik

Dataset terbagi juga menjadi 2 dataset yaitu dataset ulasan terkait pelayanan aplikasi MyTelkomsel dan Ulasan pengguna terkait aplikasi MyTelkomsel.

a. Dataset Pelayanan

Pada penelitian ini dataset pelayanan memiliki 1436 ulasan.

- Dataset Pelayanan Lexicon Based

Terdapat 456 data positif, 827 data negatif dan 153 data netral. Dapat dilihat pada Tabel 9 :

Tabel 9. Dataset Pelayanan Lexicon Based

No	Skor Sentimen	Sentimen	content
1	2	positif	layan cs nya
2	6	positif	terima kasih layan nya
...
1436	0	netral	my telkomsil emang kualitas

- Dataset Pelayanan oleh Pakar

Terdapat 138 data positif, 1116 data negatif dan 182 data netral. Dapat dilihat pada Tabel 10 :

Tabel 10. Dataset Pelayanan Pakar

No	Sentimen	content
1	netral	layan cs nya
2	positif	terima kasih layan nya
...
1436	positif	my telkomsil emang kualitas

b. Dataset Pengguna

Pada penelitian ini dataset pengguna memiliki 1128 ulasan.

- Dataset Pengguna Lexicon Based

Terdapat 370 data positif, 652 data negatif dan 106 data netral. Dapat dilihat pada Tabel 11 :

Tabel 11. Dataset Pengguna Lexicon Based

No	Skor Sentimen	Sentimen	content
1	2	positif	kesini kesana ya t sel kasih rate star dah ckup ya hai teman mari alih operator
2	3	positif	coba kasih bntang
...
1128	1	positif	jaring ngeleg gw ngegame musuh jaring tolol telkom

- Dataset Pengguna oleh Pakar

Terdapat 118 data positif, 897 data negatif dan 113 data netral. Dapat dilihat pada Tabel 12 :

Tabel 12. Dataset Pengguna oleh Pakar

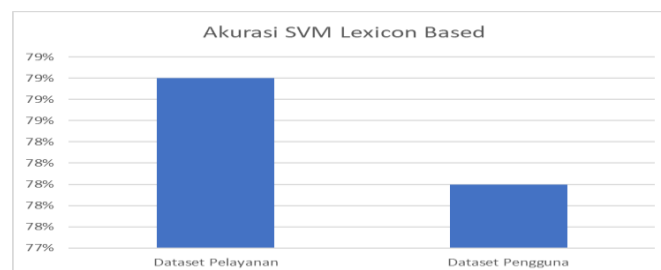
No	Sentimen	content
1	netral	kesini kesana ya t sel kasih rate star dah ckup ya hai teman mari alih operator
2	netral	coba kasih bntang
...
1128	negatif	jaring ngeleg gw ngegame musuh jaring tolol telkom

6. Pembagian Data

Pada penelitian ini pembagian data latih dan data uji dilakukan dengan teknik Holdout Validation sebesar 70% data latih dan 30% data uji.

7. Penerapan Algoritma SVM

Klasifikasi Algoritma SVM menggunakan Data Pelabelan Lexicon Based

**Gambar 2.** Akurasi SVM Lexicon Based

Berdasarkan Gambar 2, akurasi tertinggi terjadi pada dataset Pelayanan dengan nilai 79%, sementara dataset Pengguna memiliki akurasi 78%. Pengujian confusion matrix digunakan untuk mengevaluasi hasil prediksi. Detail hasil dari kedua dataset tersedia dalam Tabel 13 dan Tabel 14 :

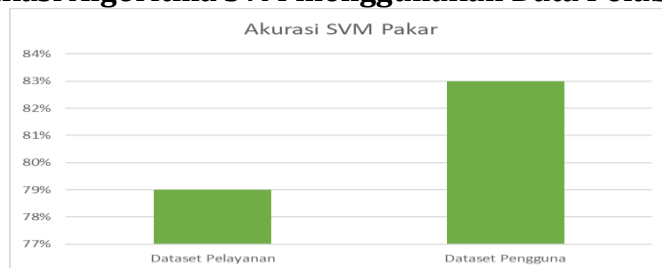
Tabel 13. Confusion Matrix SVM Lexicon Based

Prediksi	Dataset Pelayanan			Dataset Pengguna		
	Aktual			Aktual		
	Positif	Negatif	Netral	Positif	Negatif	Netral
Positif	235	3	15	174	0	10
Negatif	23	11	13	29	2	2
Netral	31	5	95	35	0	87

Tabel 14. Classification Report SVM Lexicon Based

Ukuran	Dataset Pelayanan			Dataset Pengguna		
	Positif	Negatif	Netral	Positif	Negatif	Netral
Presisi	77%	81%	58%	88%	73%	100%
Recall	73%	93%	23%	71%	95%	6%
F1-score	75%	87%	33%	79%	82%	11%

Klasifikasi Algoritma SVM menggunakan Data Pelabelan Pakar



Gambar 3. Akurasi SVM Pakar

Berdasarkan Gambar 3, dataset Pengguna memiliki akurasi tertinggi (83%), sedangkan dataset Pelayanan memiliki akurasi 79%. Pengujian confusion matrix dilakukan untuk mengevaluasi hasil prediksi. Detail hasil dari kedua dataset terdapat pada Tabel 15 dan Tabel 16 :

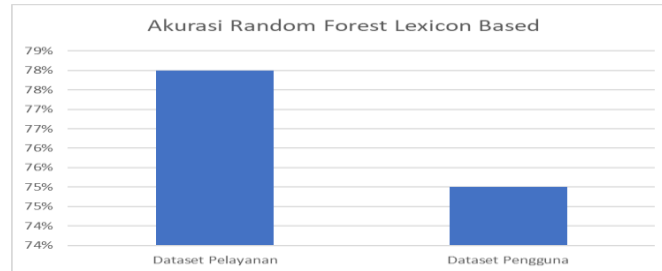
Tabel 15. Confusion Matrix SVM Pakar

Prediksi	Dataset Pelayanan			Dataset Pengguna		
	Aktual			Aktual		
	Positif	Negatif	Netral	Positif	Negatif	Netral
Positif	333	2	3	265	0	0
Negatif	54	0	0	30	0	0
Netral	31	1	7	29	0	15

Tabel 16. Classification Report SVM Pakar

Ukuran	Dataset Pelayanan			Dataset Pengguna		
	Positif	Negatif	Netral	Positif	Negatif	Netral
Presisi	70%	80%	0%	100%	82%	0%
Recall	18%	99%	0%	34%	100%	0%
F1-score	29%	88%	0%	51%	90%	0%

8. Penerapan Algoritma Random Forest
Klasifikasi Algoritma Random Forest menggunakan Data Pelabelan Lexicon Based



Gambar 4. Akurasi Random Forest Lexicon Based

Berdasarkan Gambar 4 menunjukkan bahwa akurasi tertinggi ditemukan pada dataset Pelayanan (78%), sedangkan dataset Pengguna memiliki akurasi sebesar 75%. Pengujian confusion matrix dilakukan untuk mengevaluasi hasil prediksi, dan detailnya dapat dilihat pada Tabel 17 dan Tabel 18 :

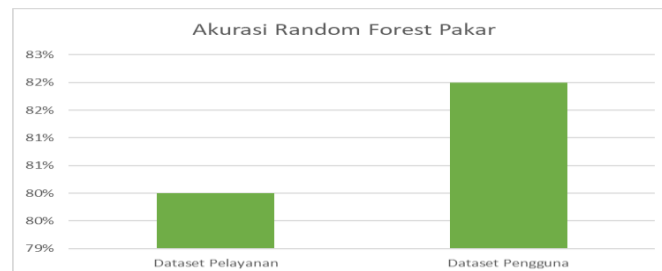
Tabel 17. Confusion Matrix Random Forest Lexicon Based

Prediksi	Dataset Pelayanan			Dataset Pengguna		
	Aktual			Aktual		
	Positif	Negatif	Netral	Positif	Negatif	Netral
Positif	231	6	16	175	0	9
Negatif	27	14	6	28	5	0
Netral	35	5	91	47	0	75

Tabel 18. Classification Report Random Forest Lexicon Based

Ukuran	Dataset Pelayanan			Dataset Pengguna		
	Positif	Negatif	Netral	Positif	Negatif	Netral
Presisi	81%	79%	56%	89%	70%	100%
Recall	69%	91%	30%	61%	95%	15%
F1- score	75%	85%	39%	73%	81%	26%

Klasifikasi Algoritma *Random Forest* menggunakan Data Pelabelan Pakar



Gambar 5. Akurasi Random Forest Pakar

Berdasarkan Gambar 5, akurasi tertinggi diperoleh pada dataset Pengguna dengan nilai 82%, sedangkan dataset Pelayanan memiliki akurasi 80%. Pengujian confusion matrix digunakan untuk mengevaluasi hasil prediksi. Detail hasil dari kedua dataset dapat dilihat pada Tabel 19 dan Tabel 20 :

Tabel 19. Confusion Matrix Random Forest Pakar

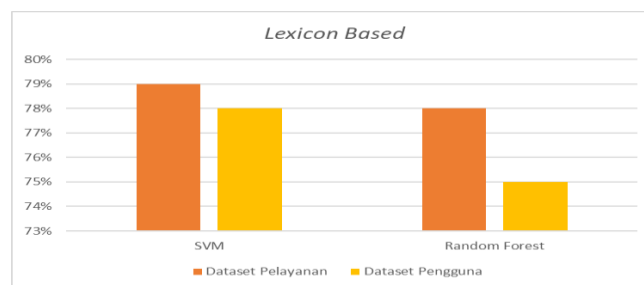
Prediksi	Dataset Pelayanan			Dataset Pengguna		
	Aktual			Aktual		
	Positif	Negatif	Netral	Positif	Negatif	Netral
Positif	334	1	3	265	0	0
Negatif	50	3	1	30	0	0
Netral	29	2	8	30	1	13

Tabel 20. Classification Report Random Forest Pakar

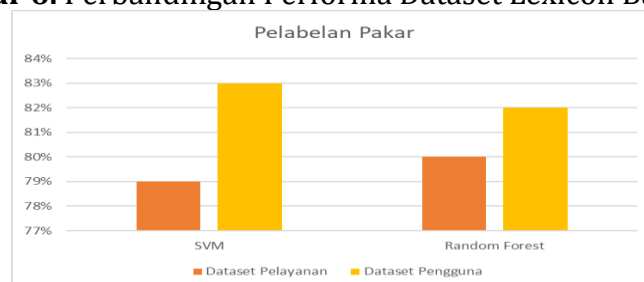
Ukuran	Dataset Pelayanan			Dataset Pengguna		
	Positif	Negatif	Netral	Positif	Negatif	Netral
Presisi	67%	81%	50%	100%	82%	0%
Recall	21%	99%	6%	30%	100%	0%
F1- score	31%	89%	10%	46%	90%	0%

9. Perbandingan Performa Algoritma dan Topik Dataset

Perbandingan performa algoritma SVM dan Random Forest pada data Aplikasi MyTelkomsel menggunakan 2 dataset topik dan menggunakan 2 pendekatan, didapatkan hasil akurasi algoritma SVM dan Random Forest *Lexicon Based* pada dataset pelayanan sebesar 79% dan 78%, dataset pengguna sebesar 78% dan 75%, sedangkan akurasi algoritma SVM dan Random Forest Pakar pada dataset pelayanan sebesar 79% dan 80%, sedangkan dataset pengguna sebesar 83% dan 82%. Hasil dapat dilihat pada Gambar 6 dan Gambar 7 :



Gambar 6. Perbandingan Performa Dataset Lexicon Based



Gambar 7. Perbandingan Performa Dataset Pakar

10. Visualisasi Data Komentar

Visualisasi data ulasan Google Play Store menggunakan WordCloud membantu mengidentifikasi kata-kata yang sering muncul dalam teks dan sentimen yang umum, seperti kata-kata positif, negatif, dan netral.



Gambar 8. Visualisasi Data Komentar Positif



Gambar 9. Visualisasi Data Komentar Negatif



Gambar 10. Visualisasi Data Komentar Netral

D. Simpulan

Dalam penelitian ini, perbandingan Pendekatan yang terbagi menjadi pelabelan *lexicon based* dan pelabelan pakar bahasa dan dibagi lagi menjadi 2 Dataset berdasarkan topik yaitu pelayanan dan pengguna. Hasilnya menunjukkan bahwa pada pendekatan *lexicon based*, dataset pelayanan menunjukkan kinerja lebih unggul dengan akurasi algoritma tertinggi mencapai 79% menggunakan algoritma SVM. Namun, pada pelabelan oleh pakar, dataset pengguna menunjukkan kinerja lebih unggul dengan akurasi algoritma tertinggi mencapai 83% menggunakan algoritma SVM. Oleh karena itu, pendekatan oleh pakar cenderung lebih akurat dibandingkan pendekatan *lexicon based*. Selain itu, dalam penelitian ini, algoritma SVM menunjukkan akurasi tertinggi, yakni 83%, pada dataset pengguna yang dianalisis oleh pakar.

E. Referensi

- [1] CNBC Indonesia, "No Title," [www.cnbcindonesia.com](https://www.cnbcindonesia.com/tech/20230929101832-37-476411/operator-hp-kebanyakan-ini-dampak-ke-warga-ri), 2023. <https://www.cnbcindonesia.com/tech/20230929101832-37-476411/operator-hp-kebanyakan-ini-dampak-ke-warga-ri>
- [2] GoodStats, "No Title," [goodstats.id](https://goodstats.id/article/inilah-provider-internet-tercepat-kuartal-iv-2024-adakah-providermu-qDi6Y), 2024. <https://goodstats.id/article/inilah-provider-internet-tercepat-kuartal-iv-2024-adakah-providermu-qDi6Y>
- [3] Telkomsel, "No Title," <https://www.telkomsel.com/>, 2023. <https://www.telkomsel.com/about-us/news/telkomsel-luncurkan->

- pengalaman-wisata-digital-jelajah-nusantara-20-di-aplikasi
- [4] S. E. Qoidah, "Analisis Atas Kinerja Karyawan Dalam Hubungannya Dengan Stres Kerja Dan Kepuasan Kerja Karyawan Customer Service Telkomsel Di Gerai Grapari Mall Pondok Indah 3 Jakarta Selatan," 2021.
 - [5] A. Zein and E. Suharyanto, "Sentimen Analisis Pada Komentar Pendek Evaluasi Dosen Oleh Mahasiswa (EDOM) Program Studi Sistem Informasi Universitas Pamulang," *J. Ilmu Komput.*, vol. 5, no. 1, p. 17, 2022.
 - [6] A. Saepudin, A. Faqih, and G. Dwilestari, "Perbandingan Algoritma Klasifikasi Support Vector Machine, Random Forest dan Logistic Regression Pada Ulasan Shopee," *J. Tekno Kompak*, vol. 18, no. 1, pp. 178–192, 2024.
 - [7] M. R. Adrian, M. P. Putra, M. H. Rafialdy, and N. A. Rakhmawati, "Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB," *J. Inform. Upgris*, vol. 7, no. 1, 2021.
 - [8] E. Fitri, "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine," *J. Transform.*, vol. 18, no. 1, pp. 71–80, 2020.
 - [9] A. Miftahusalam, A. F. Nuraini, A. A. Khoirunisa, and H. Pratiwi, "Perbandingan Algoritma Random Forest, Naive Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Penghapusan Tenaga Honorer," in *Seminar Nasional Official Statistics*, 2022, vol. 2022, no. 1, pp. 563–572.
 - [10] M. Fachriza and M. Munawar, "Analisis Sentimen Kalimat Depresi Pada Pengguna Twitter Dengan Naive Bayes, Support Vector Machine, Random Forest," *KOMPUTEK*, vol. 7, no. 2, pp. 49–58, 2023.
 - [11] F. ADLINA, "Analisis Sentimen Online Customer Review Pada Toko Smartphone Daerah Istimewa Yogyakarta Di E-Marketplace Shopee Menggunakan Lexicon Based Dan Word Cloud," 2022.
 - [12] G. N. Aulia and E. Patriya, "Implementasi Lexicon Based Dan Naive Bayes Pada Analisis Sentimen Pengguna Twitter Topik Pemilihan Presiden 2019," *J. Ilm. Inform. Komput.*, vol. 24, no. 2, pp. 140–153, 2020.
 - [13] S. S. Salim and J. Mayary, "Analisis Sentimen pengguna Twitter terhadap dompet elektronik dengan metode lexicon based dan k-nearest neighbor," *J. Ilm. Inform. Komput.*, vol. 25, no. 1, pp. 1–17, 2020.
 - [14] A. Fatihin, "Analisis sentimen terhadap ulasan aplikasi mobile menggunakan metode support vector machine (SVM) dan pendekatan lexicon based." Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah ..., 2022.
 - [15] R. ADHA, "PENDEKATAN LEXICON BASED FEATURES UNTUK IDENTIFIKASI CYBERBULLYING MENGGUNAKAN ALGORITMA KLASIFIKASI." Universitas Islam Negeri Sultan Syarif Kasim Riau, 2023.
 - [16] M. Huang, H. Xie, Y. Rao, Y. Liu, L. K. M. Poon, and F. L. Wang, "Lexicon-Based Sentiment Convolutional Neural Networks for Online Review Analysis," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1337–1348, 2022, doi: 10.1109/TAFFC.2020.2997769.
 - [17] U. Khaira, R. Johanda, P. E. P. Utomo, and T. Suratno, "Sentiment analysis of cyberbullying on twitter using SentiStrength," *Indones. J. Artif. Intell. Data Min*, vol. 3, no. 1, p. 21, 2020.

- [18] E. R. Indriyani, P. Paradise, and M. Wibowo, "Perbandingan Metode Naïve Bayes dan Support Vector Machine Untuk Analisis Sentimen Terhadap Vaksin Astrazeneca di Twitter," *J. Media Inform. Budidarma*, vol. 6, no. 3, pp. 1545–1553, 2022.