

## SENTIMENT ANALYSIS ONLINE SHOP ON THE PLAY STORE USING METHOD SUPPORT VECTOR MACHINE (SVM)

Muhammad Iqbal Ahmadi <sup>(1)\*</sup>, Fuji Apriani <sup>(2)</sup>, Mia Kurniasari <sup>(3)</sup>, Siti Handayani <sup>(4)</sup>,  
Dudih Gustian<sup>(5)</sup>

Sistem Informasi, Universitas Nusaputra  
Jl Raya Cibatu Cisaat No.21, Cibolang Kaler, Kec.Cisaat, Sukabumi, Jawa Barat 43155  
e-mail : muhammad.iqbal\_si17@nusaputra.ac.id <sup>(1)\*</sup>

### Abstract

*Online shopping activities have become a necessity for today's society, where now more people choose to shop online rather than offline. The growth of shopping through E-commerce continues to show an increasing trend. It is important for companies to know the public's response to the products or services they offer. Responses from the public can affect the image of a company, there are positive and negative responses from the public. If a lot of products or services give negative responses, a company or service will be badly rated by users and can affect the company. However, to monitor or classify responses from the public is not easy, because the numbers are too large to be processed manually. In this study the authors will use the Support Vector Machine method to classify the positive and negative responses of users on online shopping applications on the Google play store. In this study, the authors used 10 fold cross validation as an evaluation model with 3,000 reviews and 1500 training data. 300 reviews were taken from each application as testing data. This research uses the support vector machine method which produces the application with the highest level of accuracy is Tokopedia with accuracy values, namely Tokopedia 90.67%, JD.ID 75.33%, Blibli 74.00%, Shopee 70.00%, Lazada 69.00% .*

**Keywords : Google Play Store, Support Vector Machine, Sentimen Analysis**

Abstrak- Aktifitas belanja online telah menjadi kebutuhan masyarakat saat ini, dimana sekarang lebih banyak masyarakat memilih belanja online daripada offline. Pertumbuhan belanja melalui E-commerce terus menunjukkan tren meningkat. Penting bagi perusahaan untuk mengetahui tanggapan publik mengenai produk atau layanan yang mereka tawarkan. Tanggapan dari publik dapat mempengaruhi citra dari sebuah perusahaan, tanggapan dari publik tersebut ada positif dan negatif. Jika dari sebuah Produk atau layanan banyak yang memberi tanggapan negatif maka sebuah perusahaan atau layanan akan di nilai buruk oleh para pengguna dan dapat mempengaruhi pada perusahaan tersebut. Akan tetapi untuk memantau atau mengklasifikasi tanggapan dari publik bukanlah hal yang mudah, karena jumlah yang terlalu banyak jika diproses secara manual. Dalam penelitian ini penulis akan menggunakan metode Support Vector Machine untuk mengklasifikasi tanggapan positif dan negatif para pengguna pada aplikasi belanja online di Google play store. Dalam penelitian ini penulis menggunakan 10 *fold cross validation* sebagai model evaluasi dengan data berjumlah 3.000 ulasan, sebagai data training sebanyak 1500. yang diambil dari masing-masing aplikasi sebanyak 300 ulasan sebagai data testing. penelitian ini menggunakan metode support vector machine yang menghasilkan aplikasi dengan tingkat akurasi paling tinggi adalah Tokopedia dengan nilai akurasi yaitu Tokopedia 90,67% , JD.ID 75,33%, Blibli 74.00%, Shopee 70,00%, Lazada 69,00%.

**Kata Kunci : Google Play Store, Support Vector Machine, Analisis Sentimen**

### 1. PENDAHULUAN

Di era 4.0 ini perkembangan teknologi telah mengubah cara manusia untuk melakukan aktifitasnya, perkembangan teknologi tersebut dapat mempermudah kebutuhan manusia melalui penggunaan gadget dan internet. Dimana kita dapat mengakses berbagai informasi dan komunikasi melalui teknolgi tersebut. Berbagai aktifitas manusia seperti komunikasi menggunakan media sosial, belajar secara online, dan berbelanja secara online pun dapat di akses melalui internet tanpa harus pergi ke toko secara langsung. Kemajuan teknologi ini

membuat proses transaksi menjadi lebih mudah dengan adanya *E-commerce*. *E-commerce* adalah perdagangan, penyebaran, penjualan, pembelian, serta pemasaran barang dan jasa yang mengandalkan sistem elektronik, seperti internet, TV, atau jaringan komputer yang lainnya. Sedangkan online marketplace adalah sebuah situs yang memfasilitasi proses jual beli dari berbagai toko, online market place paling banyak diminati oleh kebanyakan masyarakat karena mempunyai banyak kelebihan seperti tidak harus pergi ke tokko secara langsung, bisa membeli barang berbeda wilayah, dapat di akses 24 jam, dan banyaknya diskon, promo dan bebas ongkos kirim.

Di Indonesia sendiri begitu banyak online marketplace yang bisa di akses atau di download melalui google playstore, seperti Lazada, Tokopedia, Shopee, JD.ID dan masih banyak lagi. Dari masing-masing online shop tersebut mempunyai kelebihan dan kekurangannya masing-masing. Kekurangannya seperti banyaknya penipuan pada saat transaksi, penjual tidak mengirimkan barang yang sesuai dengan konsumen, serta kemasan yang sudah rusak. Masalah tersebut sangat berpengaruh pada kepuasan pelanggan yang seharusnya menjadi tolak ukur kualitas pelayanan dari online shop tersebut.

Dari latar belakang diatas terdapat beberapa rumusan masalahnya yaitu :

- Aspek yang di ulas oleh para pengguna dari berbagai aplikasi online Marketplace.
- Pengkategorian aplikasi online marketplace yang ada di play store.
- Bagaimana membangun sistem sentimen analisis untuk mengetahui aplikasi online marketplace mana yang terbaik di playstore.

Pada penelitian ini penulis akan melakukan pengambilan data dengan menggunakan google playstore. Dari hasil pengolahan data tersebut berupa analisis sentimen. Analisis sentimen adalah sebuah proses menemukan pendapat pengguna tentang beberapa topik atau teks yang disampaikan. Tugas dasar dalam analisis ini adalah mengelompokkan sentimen pelanggan menjadi positif dan negatif berdasarkan komentar yang ditulis oleh para pengguna aplikasi, kemudian membandingkan antara kelima online shop tersebut untuk mengetahui posisi atau tingkatan dari masing-masing online marketplace tersebut di goole playstore.

Penelitian ini menggunakan *metode Support Vectore Mesin (SVM)* metode tersebut merupakan satu metode yang menggunakan machine learning yang memiliki hasil lebih baik dalam hal klasifikasi. Pada penelitian ini penulis berharap agar mampu membantu masyarakat untuk mengetahui online marketplace mana yang baik, aman dan nyaman saat melakukan transaksi, sedangkan untuk perusahaan online marketplace itu sendiri yaitu Dapat memperoleh masukan inti dan mendorong peningkatan kulaitas pelayanan mereka.

Penelitian terdahulu dilakukan oleh Nuraeni Herlinawati STMIK Nusa Mandiri melakukan penelitian mengenai sentimen analisis terhadap aplikasi zoom meeting menggunakan metode *Naïve Bayes dan Support Vector Machine*. Dimana penelitian ini untuk melihat ulasan para pengguna zoom meeting, dan ulasan tersebut dimasukan ke dalam dua kategori yaitu positif dan negatif.

Jumlah dataset setelah preprocessing menjadi 1.007 record. Data hampir seimbang dengan label positif sebanyak 546 dan label negatif 461 ulasan. Evaluasi model menggunakan 10 fold cross validation diperoleh nilai akurasi dan nilai AUC dari masing-masing algoritma yaitu untuk NB nilai akurasi = 74,37% dan nilai AUC = 0,659. Sedangkan untuk algoritma SVM nilai akurasi = 81,22% dan nilai AUC = 0,886. Dalam penelitian ini dapat diketahui bahwa tingkat akurasi yang didapatkan algoritma Support Vector Machine (SVM) lebih unggul 6,85% dibandingkan algoritma Naïve Bayes (NB).

## **2. METODE PENELITIAN**

Pada penelitian ini penulis menggunakan *metode support vector machine* untuk menghasilkan tingkat akurasi dalam menganalisa label sentimen pada ulasan yang sudah di ambil dari google play store yang dimana ulasan tersebut diberikan oleh pengguna aplikasi terkait sebagai penggambaran tingkat kepuasan dalam penggunaan aplikasi pembalajaan online tersebut,

adapun proses pengolahan data agar menghasilkan akurasi sentimen aplikasi adalah sebagai berikut :

### **2.1 Pengambilan data**

Data yang diambil yaitu ulasan yang sudah diberikan oleh pengguna aplikasi terkait di google play store, dengan menggunakan metode web scraping (data miner yang disediakan oleh google chrome) yakni suatu teknik pengambilan data yang digunakan untuk memecah data dalam jumlah besar yang diambil dari web, data yang sudah dipilih akan di rubah menjadi deretan tabel (spreadsheet) yang di kumpulkan di dalam microsoft excel, Untuk tahap selanjutnya yaitu tahap processing atau Proses data.

### **2.2 Proses data**

Tahap berikutnya adalah tahap persiapan data yang telah di ambil, agar dapat meningkatkan tingkat efektivitas dalam proses selanjutnya. Dalam tahap ini meliputi pembangunan data dan membersihkan data agar siap untuk diproses di tahap selanjutnya. Berikut adalah tahapan Proses data diantaranya :

- i. Tokenization : Dalam tahap ini data akan dipecah menjadi beberapa bagian karakter atau kata-kata yang sesuai dengan kebutuhan, sering juga disebut token agar dapat menjadi suatu kata yang memiliki arti tertentu.
- ii. Transform case : Megubah semua huruf yang asalnya kapital atau huruf besar menjadi sebaliknya atau huruf kecil agar dalam tahap selanjutnya terdapat keseragaman huruf dan akan meminimalkan kesalahan proses tokenize.
- iii. Filter tokens (by length) : Yaitu tahap menghapus kata-kata dengan panjang huruf tertentu. Disini penulis memberikan batasan yaitu minimal 2 karakter dan maksimal 25 karakter. Jadi kata-kata yang memiliki 1 karakter dan lebih dari 25 karakter akan otomatis dihilangkan.
- iv. Filter stopwords : Merupakan tahap penghapusan terhadap kata-kata yang tidak memiliki arti yang biasanya berupa kata sambung, kata keterangan, dan sebagainya. Yang dimana disini penulis sudah menyiapkan sebuah data pendukung yaitu data stoplist yang berisi kata-kata yang memiliki kemungkinan besar tidak akan memberi pengaruh dalam tahapan perediksi. Kata kata yang tidak berguna nantinya akan di buang dan tidak akan di gunakan pada tahap selanjutnya.
- v. Stemming : Menghilangkan imbuhan pada masing masing kata sehingga menjadi kata dasar, dan dalam tahap ini juga bertujuan untuk membersihkan suatu kata dari pengejaan yang kurang tepat.
- vi. Labeling : Disini data yang sudah di proses akan masuk ke tahap pelebelan yaitu tahap memberikan identitas dan dibagi menjadi dua kategori yaitu ulasan positif atau ulasan negatif.

### **2.2 Pemodelan data**

Setelah data sudah melalui proses pembangunan data, selanjutnya data akan masuk ke tahap analisis data menggunakan algoritma yang telah di tentukan yaitu support vector machine. Pada tahap ini data akan dibagi menjadi dua bagian yaitu data testing dan data training. Dalam tahap ini akan di lakukan proses pengklasifikasian label positif dan negatif pada ulasan setiap aplikasi pembelian online di google play store menggunakan algoritma support vector machine untuk mendapatkan tingkatan akurasi teratas. Di tahap ini juga penulis menggunakan tools berupa aplikasi pendukung yakni aplikasi repidminer, untuk versi dari aplikasi rapidminer yang digunakan saat ini adalah versi 9.8.

### 2.3 Evaluasi

Selanjutnya masuk ke tahap terakhir dalam penelitian ini yaitu melakukan tahap evaluasi terhadap data yang di hasilkan dari tahap-tahap sebelumnya yaitu berupa akurasi dan AUC (Area Under Curve). Dalam tahap ini juga penulis dapat membandingkan tingkat akurasi dan tingkat AUC (Area Under Curve) aplikasi mana yang tertinggi dan terendah.

## 3. HASIL DAN PEMBAHASAN

Untuk hasil dan pembahasan ini penulis akan menjelaskan proses yang di jalani selama penelitian. Disini penulis menggunakan tools yakni repidminer dalam mengolah data menggunakan metode algoritma Support Vector Machine. Berikut beberapa proses yang penulis alami :

### 3.1 Pengambilan data

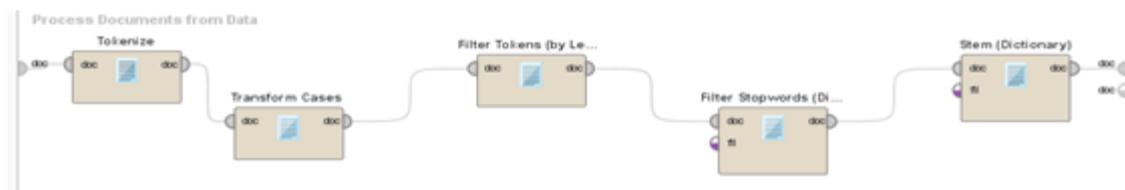
Dalam penelitian ini data yang diambil adalah ulasan para pengguna dari beberapa aplikasi *online marketplace* yang terdapat di google play store, adapun beberapa aplikasi yang akan dipilih untuk penelitian ini adalah Tokopedia, Shopee, Lazada, JD.ID, dan Blibli. Data yang di ambil menggunakan teknik Web Scrapping (add-on Google Chrome yaitu Data Miner). Jumlah data yang diambil yaitu 1500 ulasan dengan perbandingan ulasan positif dan negatif 50 : 50, data tersebut akan di jadikan data latih untuk data ulasan dari setiap aplikasi yang berjumlah 300 ulasan.

Row No.	Sentiment	TEXT
1	positive	Hai tokopedia..saya sudah ikut pelatihan di sini.pelatihannya bagus , mudah di fahami , dan ga ngebosenin..dulu saya mikirnya ...
2	positive	Hello tokopedia,, saya sudah mengikuti pelatihan prakerja,, dan saya sangat puas karena aplikasi ini sangat membantu dan m...
3	positive	Terimakasih Tokopedia, berkat pelatihan kerja yang di berikan, saya banyak mendapatkan ilmu-ilmu dan pengalaman baru . Su...
4	negative	sy beli paket smartfren yg 20rbu,dari tadi siang sampe saat ini blm masuk juga,keterangannya,menunggu respon dari operator,...
5	positive	Hai tokopedia..saya sudah ikut pelatihan di sini.pelatihannya bagus , mudah di fahami , dan ga ngebosenin..dulu saya mikirnya ...
6	positive	Hello tokopedia,, saya sudah mengikuti pelatihan prakerja,, dan saya sangat puas karena aplikasi ini sangat membantu dan m...
7	positive	Terimakasih Tokopedia, berkat pelatihan kerja yang di berikan, saya banyak mendapatkan ilmu-ilmu dan pengalaman baru . Su...
8	negative	sy beli paket smartfren yg 20rbu,dari tadi siang sampe saat ini blm masuk juga,keterangannya,menunggu respon dari operator,...
9	positive	taeek...!! lama lama tokped makin ancurr.. jualan di tokped semakin tinggi cost nya.. sementara untung gk seberapa.. klo kita g...
10	negative	Tolong untuk jual belinya bisa saling komunikasi dulu (telepon) biar jelas tidaknya.... hp bisa aja buat mainan anak-anak./ ke...

Gambar 1.Sampel data set

### 3.2 Proses data

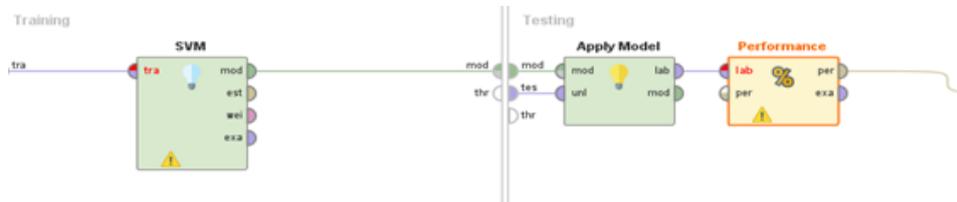
Setelah data set sudah di dapatkan selanjutnya masuk ke tahap Proses data, yang dimana disini data akan di persiapan terlebih dahulu sebelum masuk ke tahap pemodelan. adapun tahap ini mencakup kegiatan pembangunan data dan membersihkan data agar `siap di proses. Dalam Proses data melewati beberapa langkah diantaranya Tokennize, Transform Case, Filter Tokenize (By Length), Filter Stopwords (Dictionary), Stem (Dictionary).



Gambar 2. Proses Data

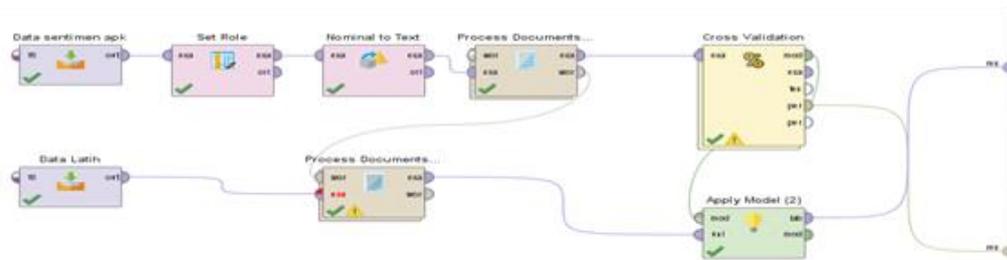
### 3.3 Pemodelan data

Dalam pemodelan data akan dilakukan pengklasifikasian label positif dan negatif pada setiap ulasan aplikasi Online shop di Google Play Store menggunakan metode Support Vector Machine yang akan menghasilkan tingkatan akurasi dari beberapa Online Shop.



Gambar 3. Cross Validation Support Vector Machine

Dalam tahap pemodelan ini juga dilakukan 10 Fold Cross Validation. Berikut bentuk cross validation dari Support vector Machine.



Gambar 4. Pemodelan Data

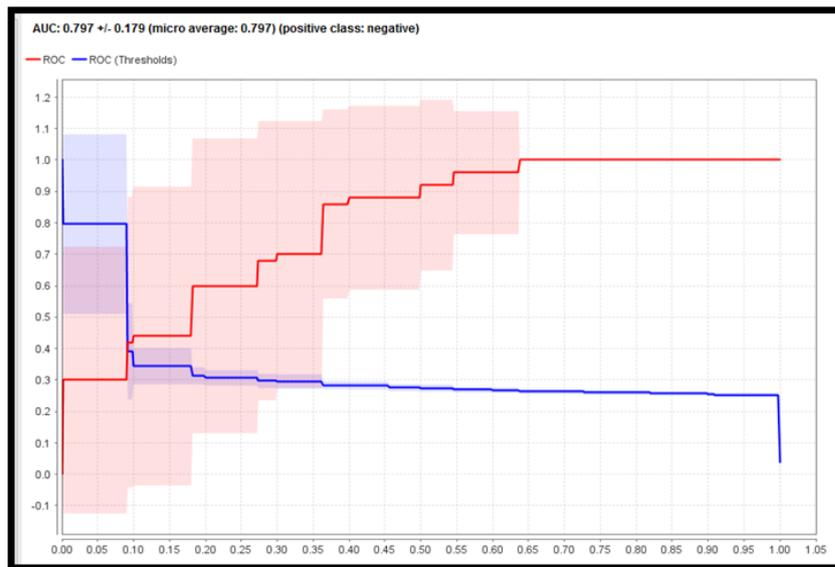
### 3.4 Evaluasi

Setelah data melalui tahap modeling selanjutnya masuk ke tahap evaluasi atas hasil dari tahap sebelumnya. Disini penulis akan membandingkan tingkat akurasi dan tingkat Area Under Curve (AUC) dari ulasan setiap aplikasi, dalam perbandingan akurasi penulis memberikan indikator True Positive (TP), True Negative (TN), False Pesitive (FP), False Negative (TN). Yang disimpulkan bahwa True Positive (TP) berada di kelas positif yang diklasifikasikan sebagai kelas positif, True Negative (TN) berada di kelas negatif yang diklasifikasikan sebagai kelas negatif, False Pesitive (FP) berada di kelas positif yang diklasifikasikan kedalam kelas negatif, False Negative (TN) berada di kelas negatif yang di klasifikasikan kedalam kelas positif.

Dalam hal ini proses pengolahan data menggunakan Support Vector Machine (SVM) menghasilkan tingkat Accuracy dan Area Under Curve (AUC).dari kelima aplikasi online shop penulis memilih Tokopedia sebagai contoh hasil dari Proses data yang di lakukan oleh tool rapidminer ditampilkan pada gambar 4 untuk tingkat akurasi dan pada gambar 5 untuk tingkat Area Under Curve (AUC).

accuracy: 90.67% +/- 3.66% (micro average: 90.67%)			
	true positive	true negative	class precision
pred. positive	270	27	90.91%
pred. negative	1	2	66.67%
class recall	99.63%	6.90%	

Gambar 2. tingkat accurasy pada aplikasi Tokopedia



Gambar 3. tingkat AUC pada aplikasi Tokopedia

Dapat disimpulkan dari gambar 4 adalah tingkat akurasi yang didapatkan dari Proses analisis sentiment menggunakan metode Support Vector Machine (SVM) adalah 90,67% dengan jumlah True Positive 270 record yang di klasifikasikan sebagai positif, False Positive 1 record diklasifikasikan sebagai negatif, True Negative 2 record diklasifikasikan sebagai positif, dan False Negative 27 record di klasifikasikan sebagai positif. Sedangkan pada gambar 5 bahwa dapat disimpulkan nilai Area Under Curve (AUC) pada marketplace Tokopedia sebesar 0,797 dengan performa yang cukup baik.

## KESIMPULAN

Berdasarkan hasil dari keseluruhan analisis sentiment online shop di google play store dengan metode Support Vector Machine menggunakan Tool Repidminer maka hasilnya dapat dirangkum dalam tabel dan grafik berikut ini :

Tabel 1. Confusion matrix setiap market place

Marketplace	TP	FP	TN	FN
Tokopedia	270	1	2	27
Shopee	4	89	206	1
Lazada	5	92	202	1
Jd.Id	222	0	4	74
Blibli	0	77	222	1

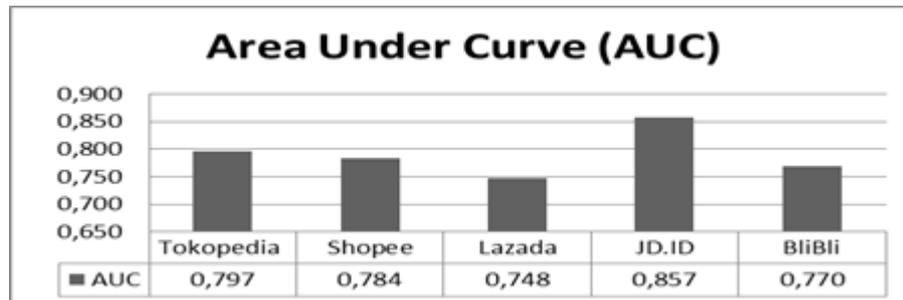
Keterangan :

- = Positif
- = Negatif

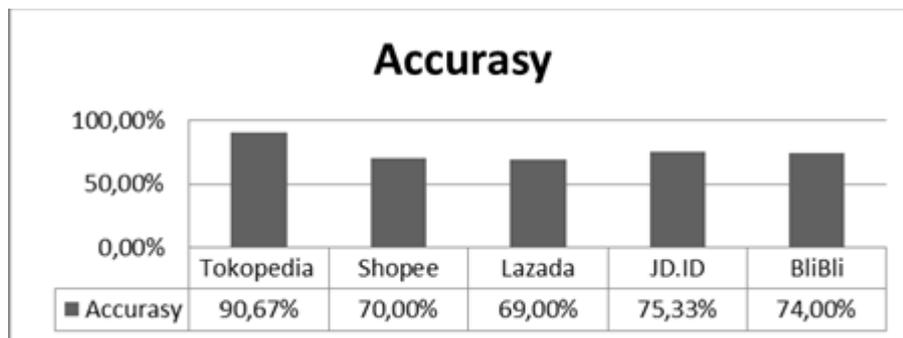
Tabel 2. Perbandingan Performance Algoritma

Marketplace	Accuracy	AUC
Tokopedia	90,67%	0,797

Marketplace	Accuracy	AUC
Shopee	70,00%	0,784
Lazada	69,00%	0,748
Jd.Id	75,33%	0,857
Blibli	74,00%	0,770



Gambar 4. Grafik tingkat akurasi



Gambar 5. Grafik tingkat Area Under Curve (AUC)

Setelah melakukan kegiatan analisis sentimen aplikasi online shop di google play store menggunakan metode support vector machine maka dapat disimpulkan bahwa aplikasi yang memiliki tingkat akurasi tertinggi di tempati oleh Tokopedia yang dimana memiliki tingkat akurasi sebesar 90,67%. Sedangkan marketplace yang memiliki tingkat akurasi terendah dalam penelitian ini jatuh kepada Lazada dengan tingkat akurasi 69,00%.

Sedangkan dalam kategori Area Under Curve (AUC) tertinggi di raih oleh JD.ID dengan perolehan 0.857, dan tingkat Area Under Curve (AUC) terendah di dapatkan oleh marketplace lazada dengan perolehan Area Under Curve (AUC) 0,748.

#### DAFTAR PUSTAKA

- Cahyono, Y. (n.d.). *ANALISIS SENTIMENT PADA SOSIAL MEDIA TWITTER MENGGUNAKAN NAÏVE BAYES CLASSIFIER DENGAN FEATURE SELECTION PARTICLE SWARM OPTIMIZATION DAN TERM FREQUENCY*. 14–19.
- Herlinawati, N., Yuliani, Y., Faizah, S., & Gata, W. (2020). *ANALISIS SENTIMEN ZOOM CLOUD MEETINGS DI PLAY STORE MENGGUNAKAN NAÏVE BAYES DAN SUPPORT VECTOR MACHINE*. 5(2), 293–298.
- Pangkey, M., Poekoel, V., Lantang, O., Informatika, T., & Ratulangi, U. S. (2016). *Sistem Pakar Pendeteksi Kerusakan Handphone Berbasis Android*. 8(1).
- Polinema, J. I., Machine, S. V., Vector, S., Machine, S. V., Analysis, T. S., Machine, S. V., Vector, S., & Analysis, S. (2016). *IMPLEMENTASI TWITTER SENTIMENT ANALYSIS UNTUK REVIEW FILM MENGGUNAKAN ALGORITMA SUPPORT*. 93–100.

- Sarjana, P. P., Elektro, J. T., & Industri, F. T. (2015). *SENTIMENT ANALYSIS MENGGUNAKAN QQQWWQQQQQQSUPPORT VECTOR MACHINE ( SVM ) SENTIMENT ANALYSIS USING SUPPORT VECTOR MACHINE ( SVM )*.
- Fauzi, M. A. (2018). Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion Automatic Essay Scoring View project Twitter Sentiment Analysis View project.
- Gani, N. S. dkk. (2020). *KOMUNIKASI, COVID 19 DALAM BINGKAI* (T. Bahfiarti (ed.)). IAIN Parepare Nusantara Press.
- Gheytsi, M., Azizifar, A., & Gowhary, H. (2015). The Effect of Smartphone on the Reading Comprehension Proficiency of Iranian EFL Learners. *Procedia - Social and Behavioral Sciences*, 199(June), 225–230.
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115.
- Mujilawati, S. (2016). Pre-Processing Text Mining Pada Data Twitter. *Seminar Nasional Teknologi Informasi Dan Komunikasi, 2016(Sentika)*, 2089–9815.
- Santoso, V. I., Virginia, G., & Lukito, Y. (2017). Penerapan Sentiment Analysis Pada Hasil Evaluasi Dosen Dengan Metode Support Vector Machine. *Jurnal Transformatika*, 14(2), 72.
- Simanihuruk, L. dkk. (2019). *E-Learning: Implementasi, Strategi dan Inovasinya* (T. Limbong (ed.)). Yayasan Kita Menulis.
- Widiastuti, N. I., Rainarli, E., & Dewi, K. E. (2017). Peringkasan dan Support Vector Machine pada Klasifikasi Dokumen. *Jurnal Infotel*, 9(4), 416.