JEPIN
(Jurnal Edukasi dan Penelitian Informatika)
ISSN(e): 2548-9364 / ISSN(p) : 2460-0741

Vol. 8
No. 3
Desember
2022

Implementasi *Support Vector Machine* untuk Analisis Sentimen Terhadap Pengaruh Program Promosi *Event* Belanja pada *Marketplace*

Gientry Rachma Ditami[#1], Eva Faja Ripanti[#2], Herry Sujaini[#3]

[#]*Program Studi Informatika, Fakultas Teknik, Universitas Tanjungpura*
Jl. Prof. Dr. H. Hadari Nawawi, Kota Pontianak, 78115

[1]gientryr@student.untan.ac.id

[2]evaripanti@untan.ac.id

[3]hs@untan.ac.id

Abstrak— **Tren belanja online membuat berbagai brand *marketplace* di Indonesia menerapkan strategi pemasaran terbaiknya untuk menarik minat pelanggan, salah satunya program promosi *event* belanja. Shopee dan Tokopedia merupakan dua *brand marketplace* teratas di Indonesia dengan pengunjung terbanyak berdasarkan data Similarweb tahun 2021. Pengalaman pengguna seputar promosi *event* belanja *marketplace* berlangsung di media sosial, salah satunya *Twitter*. Tujuan dari penelitian ini adalah membangun model analisis sentimen yang mampu mengklasifikasikan *tweets* masyarakat terkait dengan program promosi *event* belanja yang dilakukan oleh Shopee dan Tokopedia. Penelitian ini menggunakan data *tweets* pada periode yang telah ditentukan. Rangkaian *text preprocessing* yang dilakukan adalah *case folding*, *tokenizing*, *filtering*, normalisasi kata, dan *stemming*. Pembobotan kata menggunakan TF-IDF, *Support Vector Machine* sebagai algoritma pengklasifikasian, *Grid Search* untuk mencari parameter optimal, dan *K-Fold Cross Validation* serta *Confusion Matrix* untuk validasi dan pengujian model. Berdasarkan hasil analisis dan observasi, penelitian ini mengidentifikasi *event* belanja pada Shopee tanggal 25, *flash sale*, gratis ongkir, COD, tanggal kembar, dan Shopee 12.12. Sedangkan untuk Tokopedia tanggal 25, kejar diskon, bebas ongkir, COD, WIB, dan Tokopedia 12.12. Dari hasil pelabelan data, distribusi sentimen masyarakat untuk program promosi *event* belanja Tokopedia cenderung positif, Shopee cenderung negatif, dan sentimen masyarakat terhadap program promosi *event* belanja kedua *marketplace* didominasi oleh sentimen positif. Dari hasil pengujian, model yang menggunakan data set Shopee yaitu Skenario 3 dan Skenario 4 mendapat nilai akurasi tertinggi sebesar 72.12% dan 71.52%. Adapun dari hasil pencarian parameter terbaik menggunakan *Grid Search* meningkatkan nilai akurasi data set Tokopedia sebesar 1.44% dan data set Shopee sebesar 0.54%.**

***Kata kunci*— *Support Vector Machine*, *Grid Search*, *K-Fold Cross Validation*, Analisis Sentimen, *Twitter*, *Marketplace*, Program Promosi *Event* Belanja**

I. PENDAHULUAN

Di masa sekarang dengan berbagai fasilitas teknologi yang canggih mayoritas masyarakat melakukan banyak hal melalui digital, termasuk berbelanja online untuk memenuhi kebutuhannya masing-masing. Berdasarkan data [1] pengguna aktif bulanan *marketplace* peringkat lima besar di Indonesia yaitu, Tokopedia 126,4 juta, Shopee 117 juta, Bukalapak 31,27 juta, Lazada 28,20 juta, dan Blibli 18,52 juta. Perkembangan *marketplace* di Indonesia diikuti dengan meningkatnya jumlah pengguna dan orang yang berbelanja membuat tim pemasaran dari setiap *marketplace* bersaing dalam menerapkan strategi pemasarannya. Shopee dan Tokopedia merupakan *marketplace* di Indonesia yang menggunakan strategi pemasaran program promosi *event* belanja. Berbagai promosi yang ditawarkan oleh kedua *marketplace* ini membuat orang-orang yang ingin berbelanja berburu promo tersebut dan bertukar info hingga opini dengan orang lain menggunakan sosial media salah satunya, Twitter. Survei [2] menyebutkan pada tahun 2021 pengguna Twitter di Indonesia mencapai 16.32 juta. Jumlah pengguna Twitter yang besar maka terdapat banyak pula tweets atau opini yang diutarakan oleh setiap penggunanya. Berbagai informasi akan didapat dari data Twitter salah satunya pandangan masyarakat terkait dengan promosi *event* belanja pada Marketplace. Dalam mengolah data *tweets* dapat dilakukan menggunakan pendekatan metode analisis sentimen yang dapat mengklasifikasikan sentimen masyarakat.

Analisis sentimen adalah cabang penelitian dari *text mining* yang berfokus melakukan analisis dari suatu teks dokumen. *Opinion mining* dilakukan agar dapat melihat opini atau *trend* sebuah permasalahan atau objek [3]. Adapun terdapat tahapan dalam melakukan analisis sentimen meliputi tahapan pengumpulan data, klasifikasi, evaluasi, dan visualisasi data.

Submitted 25-07-2022; Revised 29-11-2022; Accepted 10-12-2022

Support Vector Machine adalah algoritma klasifikasi yang diimplementasikan untuk analisis sentimen. Konsep *Support Vector Machine* yaitu menemukan *hyperplane* paling baik dengan cara memaksimumkan jarak antar kelas [4]. Penelitian [5] yang berjudul "*Cyberbullying Classification using Text Mining* dengan menggunakan algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes*" membuktikan bahwa algoritma *Support Vector Machine* memiliki nilai akurasi lebih unggul dibandingkan algoritma *Decision Tree* dan *K-Nearest Neighbour* (KNN). Penelitian serupa yang dilakukan [6] terkait klasifikasi teks "Komparasi Algoritma *Naive Bayes* dan *Support Vector Machine* untuk Analisa Sentimen Review Film" menunjukkan nilai akurasi untuk algoritma SVM lebih unggul yaitu sebesar 90.00% dibandingkan Naive Bayes yang memiliki nilai akurasi sebesar 84.50%. Adapun algoritma *Support Vector Machine* (SVM) dinilai cocok untuk menganalisis sentimen data *tweets* [7].

Dari permasalahan yang ada didapatkan rumusan pada penelitian ini yaitu bagaimana membangun model klasifikasi teks dengan mengimplementasikan *Support Vector Machine* (SVM) kemudian dapat mengklasifikasikan *tweets* terkait dengan program promosi *event* belanja pada *marketplace*.

Penelitian ini bertujuan untuk mengimplementasikan algoritma *Support Vector Machine* dalam membangun model klasifikasi sentimen pengguna sosial media Twitter dengan topik *event* belanja *marketplace* Shopee dan Tokopedia pada kelas opini positif, negatif, dan netral. Kemudian model dapat menunjukkan kecenderungan sentimen masyarakat terhadap masing-masing *marketplace*.

II. TINJAUAN PUSTAKA

A. Text Mining

Text mining digunakan dalam menambang data berbentuk tekstual. Proses *text mining* sendiri yaitu menganalisa data teks dalam jumlah besar dalam mendapatkan informasi maupun *trend* baru yang sebelumnya belum diketahui [9]. *Text mining* dilakukan dengan mengambil sekumpulan bahasa alami yang tidak terstruktur [8]. Tujuan *text mining* yaitu untuk menemukan informasi yang sebelumnya belum diketahui, sesuatu yang belum pernah diketahui orang lain dimana tidak bisa didefinisikan [10]. Secara garis besar *text mining* adalah metode menambang data berbentuk teks yang belum terstruktur untuk menemukan informasi yang sebelumnya belum diketahui.

B. Analisis Sentimen (Opinion Mining)

Analisis sentimen ialah turunan *text* mining. Analisis sentimen adalah bidang keilmuan berfokus dalam menganalisis pendapat publik, sentimen, evaluasi, dan emosi terhadap objek baik produk, layanan, individu, isu, peristiwa, dan topik [11]. Analisis dilakukan dengan mengekstrak opini, mendalami dan melakukan olah data berupa teks secara otomatis kemudian menampilkan sentimen yang terkandung pada sebuah pendapat [12].

C. Scraping

Scraping merupakan metode atau teknik dalam pengambilan data. *Scraping* mengambil data yang tidak terstruktur di web yang kemudian disimpan dalam sebuah *database* atau *spreadsheet* [13]. *Scraping* bermanfaat agar informasi yang digali lebih ke inti sehingga lebih mudah dalam melakukan pencarian [14].

D. Text Preprocessing

Proses *text preprocessing* dilakukan dengan mengubah data yang belum terstruktur menjadi data terstruktur agar dapat melakukan proses selanjutnya yaitu analisis sentimen, peringkasan, dan *clustering* dokumen [15]. *Text preprocessing* diperlukan sebagai solusi untuk masalah *noisy* data, redudansi, dan nilai data yang hilang [16].

Terdapat enam tahapan *text preprocessing* yang dilakukan yakni *case folding*, *cleaning*, *tokenizing, filtering*, normalisasi kata, dan *stemming*.

1. *Case Folding*, mengganti keseluruhan teks dokumen menjadi *lowercase* [17].
2. *Cleaning*, membersihkan dokumen dari hal-hal yang tidak berhubungan dengan informasi di dalam dokumen, seperti tag, tautan, html, dan *script* [15].
3. Tokenizing, memotong string masukan dengan memecah kalimat menjadi kata [18].
4. *Filtering*, menghilangkan *term* (kata) yang tidak bermakna atau *stopword* [19].
5. *Normalisasi Kata*, mengganti kata yang tidak baku menjadi baku dan mengganti akronim menjadi kata sebenarnya [20].
6. *Stemming*, mencari akar kata dari setiap kata dengan menghapus imbuhan, dari prefix maupun sufiks[17].

E. Term Frequency – Inverse Document Frequency

Term Frequency - Inverse Document Frequency (TF-IDF) adalah cara untuk mendapatkan nilai bobot kata dari jumlah suatu kata muncul (*term*) yang terkandung pada satu dokumen dan keseluruhan dokumen [16]. Hasil kali dari nilai TF dan IDF yang disebut nilai TF-IDF.

F. Klasifikasi Support Vector Machine

Vladimir Vapnik menciptakan algoritma klasifikasi *Support Vector Machine* yang dapat memprediksi kelas berdasarkan pola berdasarkan hasil pembelajaran berbasis *machine learning* (*supervised learning*) [7]. Konsep SVM secara sederhana dapat artikan sebagai upaya menemukan *hyperplane* paling baik untuk memisahkan dua *class* dalam ruang *input*. Dengan mengukur nilai *margin* dan menemukan titik maksimumnya, kita dapat menemukan *hyperplane* terbaik[21].

Pada algoritma SVM terdapat fungsi *kernel* yang berfungsi menyelesaikan masalah *non-linear* menjadi *linear separable*. Beberapa fungsi *kernel* yang digunakan dalam klasifikasi yaitu *kernel Polynomial, Linear, Sigmoid*, dan *Radial Basis Function* (RBF). Setiap *kernel trick* memiliki nilai parameter sendiri. Penggunaan metode *Grid Search* dilakukan untuk menemukan nilai parameter terbaik dari fungsi *kernel* yang digunakan.

G. K-Fold Cross Validation

K-Fold Cross Validation merupakan metode untuk memvalidasi model dengan membagi data menjadi k-subset dan setelahnya dilakukan pengulangan sebanyak k kali [22]. Satu subset digunakan sebagai data uji dan subset lain digunakan sebagai data pembelajaran disetiap pengulangan. Selain memvalidasi data, membagi data *training* dan data *testing* juga dapat menggunakan *K-Fold Cross Validation*.

H. Confusion Matrix

Confusion Matrix merupakan alat untuk melakukan analisis dalam *Supervised Learning* agar dapat menampilkan hasil tes dari model yang telah diprediksi [23]. *Confusion matrix* ini digunakan untuk menghitung nilai *accuracy*, *precision*, dan *recall*. Adapun tabel *confusion matrix* ditampilkan pada Tabel 1.

TABEL I
CONFUSION MATRIX

		PREDICTION VALUES		
		TRUE	FALSE	NEUTRAL
ACTUAL VALUES	TRUE	TP	FPosNeg	FPosNet
	FALSE	FNegPos	TNeg	FNegNet
	NEUTRAL	FNetPos	FNetneg	TNet

Pada penelitian ini, pengujian *confusion matrix* menggunakan tiga parameter, yaitu nilai *accuracy*, *precision*, dan *recall*. *Accuracy* yaitu nilai persentase pendekatan dari keseluruhan data yang diidentifikasi dan dinilai. *Precision* merupakan tingkat ketepatan antara data yang diminta dengan hasil prediksi yang diberikan oleh model. *Recall* adalah nilai keberhasilan model dalam mendapatkan kembali sebuah informasi [24]. Berikut ini rumus yang digunakan dalam menghitung nilai *accuracy* (1), *precision* (2), dan *recall* (3).

$$\left(\frac{TP + TNeg + TNet}{Jumlah\ data}\right) \quad (1)$$

$$\left(\frac{\frac{TP}{TP+FPosNet+FPosNeg}+\frac{TNeg}{TNeg+FNegNet+FNegPos}+\frac{TNet}{TNet+FNetPos+TNetNeg}}{3}\right) \quad (2)$$

$$\left(\frac{\frac{TP}{TP+FNegPos+FNetPos}+\frac{TNeg}{FPosNeg+TNeg+FNetNeg}+\frac{TNet}{TNet+FNegNet+FPosNet}}{3}\right) \quad (3)$$

Pengukuran klasifikasi juga dilakukan dengan menggunakan nilai AUC untuk mengetahui golongan dari model klasifikasi yang dibangun. Nilai parameter yang digunakan untuk menghitung nilai AUC yaitu nilai *specificity* (4) dan *false positive rate* (FPR) (5) dengan rumus nilai AUC (6) sebagai berikut.

$$Specificity = \frac{TNeg}{TNeg+FPosNeg+FNetNeg} x100\% \quad (4)$$

$$\text{FPR} = FPR = 1 - Specificity \quad (5)$$

$$\text{AUC} = \frac{1+recall-FPR}{2} \quad (6)$$

I. Bahasa Pemograman Python

Menurut [25], *Python* adalah bahasa pemrograman yang ditafsirkan, interaktif, dan *object-oriented*. *Python* cocok digunakan sebagai *scripting language*, implementasi bahasa pemrograman web, dan lainnya. *Python* bisa di *extended* ke bahasa C dan C++ dan memberikan kecepatan yang memadai dalam mengkomputasi suatu pekerjaan komputasi yang intensif [26].

J. Twitter

Twitter adalah situs *web* yang dimiliki dan dioperasikan oleh Twitter Inc. Twitter menyediakan jaringan sosial *microblog* yang memungkinkan pengguna mengirim dan membaca pesan dalam bentuk kicauan (*tweets*). Selain digunakan untuk mencari informasi, Twitter digunakan pula untuk memberikan informasi umum hingga informasi pribadi [27].

K. Marketplace

Marketplace adalah *platform* digital untuk melakukan proses jual-beli yang mempertemukan penjual dan pembeli [28]. Shopee merupakan perusahaan *marketplace* yang menyediakan layanan penjualan melalui website dan aplikasi. Forrest Li merupakan *founder* yang mengenalkan Shopee pada tahun 2015 di tujuh negara Asia termasuk Indonesia. Tokopedia atau PT. Tokopedia adalah perusahaan *marketplace* karya anak bangsa Indonesia melalui *website* dan aplikasi yang dibangun oleh William Tanuwijaya dan Leontinus Alpha Edison di tahun 2009. Tokopedia menyediakan layanan bagi penjual dan pembeli dalam melakukan transaksi digital. Tokopedia sendiri mendorong pelaku UMKM di Indonesia untuk berkontribusi dalam mendukung terpenuhinya kebutuhan masyarakat di Indonesia melalui *platform* digital.

Dengan kemunculan berbagai *marketplace* di Indonesia maka setiap *marketplace* telah menyiapkan strategi pemasaran agar dapat bersaing dalam menarik dan mempertahankan pengguna salah satunya program promosi *event* belanja. Program promosi *event* belanja ini dilakukan oleh pihak *marketplace* terutama sebagai ciri khas masing-masing, selain itu *event* belanja juga diadakan setiap tanggal cantik maupun hari perayaan tertentu.

III. METODOLOGI PENELITIAN

A. Program Penelitian

Penelitian ini berfokus pada *text mining* dan analisis sentimen yang melibatkan algoritma *Support Vector Machine* pada objek *marketplace*. Terdapat 5 tahapan dalam melakukan penelitian ini. Tahapan meliputi pendalaman materi, analisis kebutuhan sistem,

perancangan model, implementasi, dan pengujian. Tahapan penelitian ini diilustrasikan pada Gambar 1.

Gambar. 1 Metodologi penelitian

Tahapan pertama adalah pendalaman materi yang dilakukan untuk memperoleh literatur dan landasan teori agar mendapat gambaran penelitian. Peneliti melakukan kajian pustaka atau studi literatur melalui jurnal, skripsi, dan artikel terkait dengan analisis sentimen atau *opinion mining* dan penggunaan algoritma *support vector machine* (SVM) dalam mengklasifikasikan data teks. Peneliti juga melakukan observasi melalui laman *marketplace* baik Shopee dan Tokopedia serta melakukan observasi melalui media sosial Twitter.

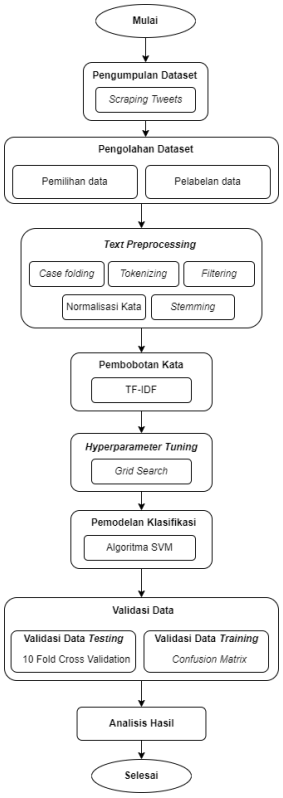
Tahapan kedua yaitu analisis kebutuhan dilakukan guna mengidentifikasi elemen-elemen yang dibutuhkan dan mendefinisikan penelitian. Dari analisis kebutuhan peneliti menghasilkan poin utama dalam penelitian ini yaitu *marketplace*, diikuti dengan parameternya yaitu program promosi *event* belanja. Analisis kebutuhan dalam proses penelitian ini meliputi pengambilan data, pengolahan, pembobotan kata, pengklasifikasian, validasi dan pengujian.

Tahapan berikutnya adalah perancangan model konseptual untuk menggambarkan sebuah gambaran penelitian yang dilakukan secara umum. Gambaran ini dapat memberikan pemahaman serta informasi kepada pembaca terkait dengan analisis sentimen yang dilakukan pada penelitian ini.

Tahapan keempat mengimplementasikan rancangan mulai dari melakukan pengumpulan data dengan Twitter sebagai sumber data menggunakan teknik *scraping*. Data *tweets* yang dikumpulkan dalam rentang waktu 6 bulan dari bulan Juli hingga Desember 2021. Data yang dihasilkan kemudian diseleksi agar menghasilkan data relevan untuk membangun model. Data *tweets* dikelompokkan menjadi 3 kelas yaitu positif, negatif, dan netral. Kemudian data *tweets* melalui tahapan *text preprocessing*, pembobotan kata TF-IDF, pengklasifikasian menggunakan algoritma *Support Vector Machine*, validasi dan pengujian menggunakan *K-Fold Cross Validation* dan *Confusion Matrix*.

Tahapan berikutnya yaitu tahapan validasi dan pengujian. *K-fold cross validation* dimanfaatkan untuk menentukan dan membagi data testing dan data *training* yang digunakan pada learning model. Penggunaan *k-fold cross validation* dapat juga menentukan akurasi dari hasil klasifikasi dan validitas data yang digunakan. Pengujian dilakukan menggunakan c*onfusion matrix* yaitu dengan parameter pengujiannya adalah nilaii dari *accuracy*, *precision*, dan *recall*. Dari hasil validasi dan pengujian maka akan dilakukan analisis hasil untuk mendapatkan kesimpulan dari penelitian ini.

B. Flowchart Penelitian



Gambar. 2 *Flowchart* penelitian

Flowchart penelitian dapat dilihat pada Gambar 2, dari proses analisis sentimen dari pengumpulan data hingga menghasilkan *output* dari pengujian. Tahapan dimulai dengan mengumpulkan data *tweets* dari Twitter menggunakan teknik *scraping*. Data *tweets* mentah hasil dari *scraping* kemudian akan diolah melalui proses pemilihan data dan dilakukan pelabelan dengan membagi data ke dalam tiga kelas, positif, negatif, dan netral. Dari proses pengolahan data kemudian dataset masuk ke dalam proses tahapan analisis sentimen, mulai dari *text preprocessing*, kemudian pembobotan kata. Pembobotan diperlukan agar dataset dapat diolah oleh algoritma klasifikasi. *Gridsearch* sebagai *hyperparameter tuning* dan

kemudian dilakukan pemodelan menggunakan Algoritma klasifikasi *Support Vector Machine*. Setelah itu, dihasilkan *learning model* yang kemudian model tersebut divalidasi menggunakan *K-Fold Cross Validation*. Kemudian proses pengujian menggunakan *data testing* yang dilakukan menggunakan *confusion matrix* dengan parameter-parameter yang telah ditentukan. Dari pengujian tersebut dilakukan analisis hasil.

IV. HASIL DAN PEMBAHASAN

A. Pemilihan Parameter

Pemilihan parameter pada penelitian ini dilakukan dengan mengobservasi *marketplace* baik Shopee dan Tokopedia serta intensitas *tweet* yang ada pada Twitter. Pemilihan parameter program promosi *event* belanja ini mengambil program-program *event* unggulan yang dipromosikan oleh kedua *marketplace* dengan kedudukan *event* yang setara. Parameter program *event* belanja dari Shopee dan Tokopedia dipaparkan pada Tabel 2.

TABEL II
EVENT BELANJA SHOPEE DAN TOKOPEDIA

Event Shopee	*Event* Tokopedia
Tanggal 25	Tanggal 25
Flash sale	Kejar diskon
Gratis Ongkir	Bebas Ongkir
COD	COD
Tanggal kembar	WIB
Shopee 12.12	Tokopedia 12.12

B. Scraping Tweet

Scraping *tweet* dilakukan menggunakan bantuan *tools online* yaitu *Advanced Search* Twitter dan *Scrape Hero*. *Advanced Search* Twitter ini digunakan untuk menambang *raw data tweet* yang telah diatur *keywords* hingga tahun pencarian *raw data tweet* yang kemudian di *scrape* menggunakan *Scrape Hero*. Data *tweets* yang diambil merupakan data *tweets* dari bulan Juli hingga Desember 2021. Berikut ini hasil *scraping tweets* terkait promosi *event* Shopee dan Tokopedia pada Tabel 3 dan Tabel 4.

TABEL III
HASIL *SCRAPING TWEETS* SHOPEE

Parameter	Juli	Agt	Sept	Okt	Nov	Des	Total
Tanggal kembar	78	68	74	117	114	121	572
Free Ongkir	122	183	32	28	45	17	427
Tanggal 25	8	73	51	0	20	12	164
Flash sale	159	698	436	52	1031	503	2879
COD	927	11	371	971	790	3	3073
Shopee 12.12						322	322
Total keseluruhan							7437

TABEL IV
HASIL *SCRAPING TWEETS* TOKOPEDIA

Parameter	Juli	Agt	Sept	Okt	Nov	Des	Total
Kejar diskon	4	13	20	6	22	4	69
COD	105	862	161	93	315	41	1577
Bebas ongkir	85	1315	225	193	210	13	2041
Tanggal 25	49	117	40	173	222	7	608
WIB	81	1584	334	182	433	23	2637
Tokopedia 12.12						178	178
Total keseluruhan							7110

C. Dataset

Dataset penelitian merupakan data *tweet* mentah yang telah lolos seleksi yang dilakukan secara manual oleh penulis. Dari hasil seleksi data didapatkan data *tweets* yang relevan untuk Shopee sebanyak 1647 *tweets* dan Tokopedia sebanyak 1623 *tweets*. Tingginya pengurangan jumlah data *tweets* dikarenakan banyaknya pengguna Twitter yang melakukan *spam tweet* atau *tweet* yang berulang.

Penelitian ini membagi dataset menjadi tiga, dataset Shopee, dataset Tokopedia, serta gabungan dataset Shopee dan Tokopedia. Hal ini bertujuan untuk membandingkan nilai prediksi sentimen terhadap Shopee dan Tokopedia serta mendapatkan nilai prediksi sentimen masyarakat terhadap program promosi *event* belanja keseluruhan. Berikut ini jumlah dataset penelitian yang dipaparkan pada Tabel 5.

TABEL V
DATASET PENELITIAN

Nama Data set	Jumlah Data
Shopee	1647
Tokopedia	1623
Shopee dan Tokopedia	3206

D. Pelabelan Data

Pelabelan data adalah proses pengolahan data *tweet* yang sebelumnya telah diperoleh dari *scraping tweet* kemudian diseleksi manual oleh penulis. Pelabelan data dibagi ke dalam tiga label kelas data, positif, negatif, dan netral. Pelabelan dilakukan dengan mengadopsi gaya berpikir konvergen yang dapat memberi simpulan logis dari informasi yang ada berdasarkan perseptual dan analitik. Hasil pelabelan data manual didapatkan dari pendapat sebanyak 7 orang yang memahami program *event* belanja dan menggunakan *marketplace* Shopee dan Tokopedia. Pengumpulan data label dari 7 orang ini dilakukan agar pelabelan manual bersifat objektif. Proses pelabelan ditampilkan pada Gambar 3.

Pelabel 1	Pelabel 2	Pelabel 3	Pelabel 4	Pelabel 5	Pelabel 6	Pelabel 7	label	tweet
Netral	positif	positif	netral	positif	netral	positif	positif	@__Yuki48 gw sih tokopedia aja di penerbit haru ma
Netral	positif	positif	netral	positif	positif	positif	positif	@_alpacaaaaa '@tokopedia Mjb, cari aja di search ba
Netral	positif	positif	netral	positif	positif	positif	positif	@_belanjarnettobelanjaklat '@tokopedia Mjb, Boleh
Netral	negatif	negatif	negatif	negatif	negatif	negatif	negatif	@_everysun '@wonunuj_ karna kalo tokopedia COD
Netral	netral	negatif	netral	netral	netral	negatif	netral	@_taetaekoo '@andinidinar03 '@bangtansbby_ '@to
Netral	negatif	negatif	negatif	negatif	negatif	negatif	negatif	@00belanjaa '@tokopedia Aku gitu jg. Bebas ongkir
Netral	netral	netral	negatif	negatif	netral	netral	negatif	@4neasy '@dittameliaa '@tokopedia Nggak semua t

Gambar. 3 Pelabelan dataset

E. Text Preprocessing

Dataset yang telah melewati proses pelabelan selanjutnya diproses ke tahapan *text preprocessing*. Terdapat 6 tahapan yang dilakukan yakni *case folding*, *cleaning*, *tokenizing*, *filtering*, normalisasi kata, dan *stemming*. Berikut ini hasil *text preprocessing* pada suatu dokumen yang dipaparkan pada Tabel 6.

TABEL VI
HASIL IMPLEMENTASI *TEXT PREPROCESSING*

Tweet	Keterangan
@Nerokumaaa sy sih mengusulkan peniadaan sistem COD maupun paylater di marketplace. selain potensi penyalahgunaan spt ini tinggi, 'korban' paylater jg banyak. yg kena kbykn org awam yg gagap mengikuti teknologi. tlg '@ShopeeID '@tokopedia '@bukalapak '@LazadaID	Tanpa *Text Preprocessing*
@nerokumaaa sy sih mengusulkan peniadaan sistem cod maupun paylater di marketplace. selain potensi penyalahgunaan spt ini tinggi, 'korban' paylater jg banyak. yg kena kbykn org awam yg gagap mengikuti teknologi. tlg '@shopeeid '@tokopedia '@bukalapak '@lazadaid	*Case Folding*
sy sih mengusulkan peniadaan sistem cod maupun paylater di marketplace selain potensi penyalahgunaan spt ini tinggi korban paylater jg banyak yg kena kbykn org awam yg gagap mengikuti teknologi tlg	*Cleaning*
[sy', 'sih', 'mengusulkan', 'peniadaan', 'sistem', 'cod', 'maupun', 'paylater', 'di', 'marketplace', 'selain', 'potensi', 'penyalahgunaan', 'spt', 'ini', 'tinggi', 'korban', 'paylater', 'jg', 'banyak', 'yg', 'kena', 'kbykn', 'org', 'awam', 'yg', 'gagap', 'mengikuti','teknologi', 'tlg']	*Tokenizing*
['mengusulkan', 'peniadaan', 'sistem', 'cod', 'paylater', 'marketplace', 'potensi', 'penyalahgunaan', 'spt', 'korban', 'paylater', 'kbykn', 'org', 'awam', 'mengikuti', 'tlg']	*Filtering*
['mengusulkan', 'peniadaan', 'sistem', 'cod', 'paylater', 'marketplace', 'potensi', 'penyalahgunaan', 'seperti', 'korban', 'paylater', 'kebanyakan' , 'orang', 'asing', 'mengikuti', 'tolong']	Normalisasi Kata
['usul', 'tiada', 'sistem', 'cod', 'paylater', 'marketplace', 'potensi', 'penyalahgunaan', 'seperti', 'korban', 'paylater', 'banyak', 'orang', 'asing', 'ikut', 'tolong']	*Stemming*

F. Pembobotan Kata Term Frequency Inverse-Document Frequency (TF-IDF)

Setelah dataset melalui *text preprocessing*, dataset kemudian masuk ke tahapan pembobotan kata agar dataset dapat diklasifikasikan. Hasil implementasi dari pembobotan kata menggunakan metode TF-IDF dapat dilihat pada Gambar 4.

```
(0, 1078)      0.19957126335355804
(0, 1011)      0.8458519616215389
(0, 202)       0.4946774402218677
(1, 1078)      0.1259924601374083
```

Gambar. 4 Hasil pembobotan TF-IDF

G. Pemodelan Klasifikasi Support Vector Machine (SVM)

Pada pemodelan klasifikasi penulis menggunakan fungsi *kernel* yang disediakan oleh algoritma SVM yaitu, *kernel rbf* (*radial basis function*). Pemilihan parameter dalam melakukan pemodelan sangat berpengaruh pada hasil model yang dibangun. Dalam penelitian ini penggunaan metode *Grid Search* dilakukan untuk menentukan parameter terbaik. Adapun pada fungsi *kernel* RBF parameter yang digunakan adalah parameter C dan parameter *gamma* (ɣ) dengan memberikan nilai 1, 10, 100 untuk parameter C dan 0.005, 0.001, 0.01, 0.1, 1 untuk parameter *gamma* (ɣ).

H. Validasi dan Hasil Pengujian Model

Terdapat enam skenario pengujian yang dilakukan. Dari tiga dataset yang digunakan, yaitu Dataset Tokopedia, Dataset Shopee, dan Dataset Gabungan ketiga dataset tersebut diujikan pada *learning model* pada saat menggunakan *parameter default* dari kernel RBF dan *parameter* hasil dari *grid search* yang dilakukan menggunakan masing-masing dataset.

TABEL VII
SKENARIO PENGUJIAN MODEL

Skenario Pengujian	Dataset	Parameter C dan *gamma* yang digunakan
Skenario 1	Dataset Tokopedia	Parameter *default*
Skenario 2	Dataset Tokopedia	*Best combination* dari *Grid Search*
Skenario 3	Dataset Shopee	Parameter *default*
Skenario 4	Dataset Shopee	*Best combination* dari *Grid Search*
Skenario 5	Dataset Gabungan	Parameter *default*
Skenario 6	Dataset Gabungan	*Best combination* dari *Grid Search*

Setelah dilakukan pengujian maka didapatkan *confusion matrix* dengan nilai parameter, yaitu nilai *accuracy*, *precision*, dan *recall*. Dari ketiga nilai pula didapatkan nilai *specificity*, *false positive rate* (FPR), dan nilai AUC. Adapun hasil pengujian menggunakan tiga dataset dan enam skenario dipaparkan pada Tabel 7.

TABEL VII
HASIL PENGUJIAN

Parameter Pengujian	Skenario	Hasil Pengujian		Δ (%)
		Data Training (%)	*Data Testing* (%)	
Accuracy	Skenario 1	66,23	70,55	+ 4,32
	Skenario 2	67,67	69,94	+ 2,27
	Skenario 3	66,94	72,12	+ 5,18
	Skenario 4	67,48	71,52	+ 4,04
	Skenario 5	68,15	71,03	+ 2,88
	Skenario 6	66,9	67,6	+ 0,7
Precision	Skenario 1	64,18	70	+ 5,82
	Skenario 2	65,31	67,75	+ 2,44
	Skenario 3	67,82	69,72	+ 1,9
	Skenario 4	67,66	69,15	+ 1,49
	Skenario 5	66,98	70,09	+ 3,11
	Skenario 6	65,92	66,46	+ 0,54
Recall	Skenario 1	61,33	65,97	+ 4,64
	Skenario 2	64,96	68,13	+ 3,17
	Skenario 3	63	66,11	+ 3,11
	Skenario 4	63,77	66,04	+ 2,27
	Skenario 5	67,01	70,27	+ 3,26
	Skenario 6	65,95	66,46	+ 0,51

Hasil pengujian yang dipaparkan pada Tabel 7 didapatkan dari nilai akurasi *data training* yang dihasilkan menggunakan *K-Fold Cross Validation* kemudian nilai akurasi *data testing* dihasilkan dari *confusion matrix* yang dipaparkan dalam *classification report*. Dari tabel tersebut dapat dilihat nilai *delta* atau selisih dari *data training* dan *data testing*. Adapun nilai selisih akurasi tertinggi dihasilkan oleh Skenario 3 sebesar +5.18%. Pengujian yang baik adalah yang memiliki nilai akurasi *data training* dan *data testing* tidak memiliki nilai *gap* yang jauh atau nilai selisih yang rendah. Dalam hal ini, Skenario 6 memiliki gap atau selisih terkecil sebesar +0.7% yang kemudian diikuti oleh Skenario 2 yang memiliki nilai selisih sebesar +2.27%. Dari hasil pengujian yang dilakukan nilai akurasi *data training* tertinggi dihasilkan oleh Skenario 5 dengan nilai akurasi sebesar 68.15%, selanjutnya Skenario 2 sebesar 67.67%, dan Skenario 4 sebesar 67.48%. Dari Tabel 7, nilai akurasi *data training* mengalami peningkatan setelah melakukan *parameter tuning* menggunakan *gird search*. Hal ini terlihat dari perbedaan nilai yang dihasilkan dari *training* Skenario 1 ke Skenario 2, dan Skenario 3 ke Skenario 4. Namun, terjadi penurunan nilai akurasi *training* setelah dilakukan parameter *tuning* menggunakan *grid search* pada Skenario 6 yang berarti parameter *tuning* menggunakan *grid search* kurang efektif untuk data set gabungan Tokopedia dan Shopee.

Berdasarkan hasil pengujian didapatkan juga nilai perbandingan pengujian model yaitu nilai akurasi *training model* yang dihasilkan dari parameter *default* dan parameter hasil dari *grid search*. Nilai parameter yang diujikan pada model yaitu dari rentang 10^-4 hingga 10^4 untuk parameter C dan parameter *gamma*.

TABEL VIII
TABEL NILAI AUC

Dataset	Nilai Akurasi *Training Model*		Δ (%)
	Default (%)	*Grid Search* (%)	
Dataset Tokopedia	66,23	67,67	1,44
Dataset Shopee	66,93	67,47	0,54
Dataset Gabungan	68,14	66,89	-1,25

Dari hasil perbandingan nilai akurasi pengujian pada Tabel 8 menunjukkan bahwa setelah melakukan parameter *tuning* menggunakan *grid search* terhadap data set Tokopedia dan data set Shopee, nilai akurasi model mengalami peningkatan sebesar 1.44% untuk data set Tokopedia dan 0.54% untuk data set Shopee. Namun pada data set Gabungan Tokopedia dan Shopee nilai akurasi model mengalami penurunan setelah dilakukan parameter *tuning* menggunakan *grid search* yaitu sebesar 1.25%.

Berdasarkan [29] efektivitas *information retrieval system* diukur menggunakan nilai *presision* dan *recall* yang dikategorikan menjadi dua, yaitu efektif dan baik jika nilai persentase diatas 50% sebaliknya tidak efektif jika dibawah 50%. Dari hasil pengujian menunjukkan nilai *precision* dan *recall* baik saat *training* maupun *testing* menunjukkan range keseluruhan nilai diatas 50%. Hal ini membuktikan bahwa model yang dibangun sudah efektif dan tingkat keberhasilan model dalam memberikan informasi sudah tergolong baik.

Dari hasil perhitungan nilai parameter pengujian yaitu nilai AUC, penelitian ini mengadopsi tabel rentang nilai AUC dari penelitian [8] pada Tabel 9.

TABEL IX
NILAI AUC DAN GOLONGAN KLASIFIKASI

Nilai AUC	Keterangan
0.90 - 1.00	Klasifikasi sempurna
0.80 - 0.90	Klasifikasi sangat baik
0.70 - 0.80	Klasifikasi baik
0.60 - 0.70	Klasifikasi cukup
≤ 0.60	Klasifikasi buruk

Berdasarkan hasil pengujian model maka didapatkan hasil perhitungan dari nilai AUC setiap model yang dibangun sebagaimana aspek-aspek yang digunakan untuk mengukur nilai AUC adalah nilai *Specificity* dan *False Positive Rate* (FPR). Dari hasil pengujian model nilai AUC

yang dihasilkan pada setiap skenario sudah termasuk ke dalam golongan klasifikasi baik. Maka dari itu, dapat diartikan bahwa model yang dibangun sudah mampu mengklasifikasikan data set yang digunakan dengan baik. Adapun nilai AUC setiap model dipaparkan pada Tabel 10.

TABEL X
TABEL NILAI AUC

Aspek Penilaian	Skenario	Hasil Pengujian Model *Data Testing*
Specificity	Skenario 1	0,64
	Skenario 2	0,76
	Skenario 3	0,94
	Skenario 4	0,91
	Skenario 5	0,79
	Skenario 6	0,78
False Positive Rate (FPR)	Skenario 1	0,36
	Skenario 2	0,24
	Skenario 3	0,06
	Skenario 4	0,09
	Skenario 5	0,21
	Skenario 6	0,22
Nilai AUC	Skenario 1	0,64985
	Skenario 2	0,72065
	Skenario 3	0,80055
	Skenario 4	0,7852
	Skenario 5	0,74635
	Skenario 6	0,7223

V. KESIMPULAN

Berdasarkan hasil penelitian, model klasifikasi yang dibangun sudah mampu untuk memprediksi nilai sentimen masyarakat terhadap pengaruh promosi *event* belanja pada *marketplace* dengan nilai sentimen cenderung Positif untuk Tokopedia dan cenderung Negatif untuk Shopee. Adapun sentimen masyarakat Twitter terhadap Program Promosi *Event* Belanja pada *Marketplace* cenderung Positif.

Dari hasil analisis dan observasi *event* belanja dari setiap marketplace yang digunakan sebagai parameter penelitian yakni, Tanggal 25, COD, WIB, Kejar Diskon, Bebas Ongkir, Tokopedia 12.12, Tanggal Kembar, *Flashsale*, Gratis Ongkir, Shopee 12.12. Data penelitian yang digunakan yaitu dataset Tokopedia, dataset Shopee, dataset gabungan Tokopedia dan Shopee.

Berdasarkan *scraping* data *tweet* yang diambil dari rentang waktu bulan Juli hingga Desember 2021 didapatkan data sebanyak 7437 *tweets* untuk Shopee dan 7110 *tweets* untuk Tokopedia. Data tersebut kemudian melalui tahapan seleksi hingga dihasilkan data relevan yang digunakan untuk penelitian sebanyak 1647 data tweets Shopee dan 1623 data tweets Tokopedia. Proses pelabelan data *tweet* dilakukan secara manual menggunakan metode kuisioner dengan mengadopsi gaya berpikir konvergen agar pelabelan data secara manual bersifat objektif. Data penelitian yang digunakan dibagi menjadi tiga yaitu data set Tokopedia, data set Shopee, dan data set gabungan Tokopedia dan Shopee. Dari hasil pelabelan *tweet* yang dilakukan distribusi sentimen data set Tokopedia cenderung positif, data set Shopee cenderung negatif, data set gabungan Tokopedia dan Shopee cenderung positif.

Dari hasil analisis pengujian model Skenario 3 memang memiliki selisih nilai akurasi model *training* dan *testing* tertinggi sebesar +5.18%. Namun apabila dilihat dari segi model yang *good fit* maka Skenario 2 (data set Tokopedia-parameter *grid search*) dan Skenario 6 (data set gabungan Tokopedia dan Shopee-parameter *grid search*) merupakan model yang baik karena memiliki nilai akurasi yang hampir sama dengan nilai selisih antara model *training* dan *testing* terkecil yaitu sebesar 0.7% dan 2.27%.

Berdasarkan hasil perbandingan model yang dilakukan didapatkan bahwa penggunaan *grid search* untuk mencari parameter terbaik dapat meningkatkan nilai akurasi pada data set Tokopedia dengan kenaikan nilai akurasi sebesar 1.44% dan data set Shopee dengan kenaikan nilai akurasi sebesar 0.54%. Namun terjadi penurunan nilai akurasi saat menggunakan data set gabungan Tokopedia dan Shopee sebesar -1.25%. Terjadinya penurunan nilai akurasi setelah dilakukan *tuning* dapat terjadi karena faktor data set yang digunakan maupun karena parameter *tuning* merupakan proses *trial* dan *error* dimana pencarian parameter melalui rentang serta nilai yang kita tentukan sehingga tidak menutup kemungkinan bila parameter *default* dari kernel SVM yang digunakan sudah merupakan parameter yang menghasilkan nilai akurasi tertinggi.

REFERENSI

[1] Similarweb, "Top Websites Ranking for Marketplace in Indonesia", [Online]. Available: https://www.similarweb.com/top-websites/indonesia/category/e-commerce-and-shopping/marketplace/ [Accessed 28 Agustus 2021]

[2] Statista, "Forecast of the number of Twitter users in Indonesia from 2017 to 2025", [Online]. Available: https://www.statista.com/forecasts/1145550/twitter-users-in-indonesia [Accessed 28 Agustus 2021]

[3] I. F. Rozi, S. H. Pramono, and E. H. Dahlan, "Implementasi *Opinion Mining* (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi," *Jurnal EECCIS*, vol. 6, no. 1, pp 37-43, 2012.

[4] Samsudiney, "Penjelasan Sederhana tentang Apa Itu SVM?," [Online]. Available: https://medium.com/@samsudiney /penjelasan-sederhana-tentang-apa-itu-svm-149fec72bd02

[5] Noviantho, S. M. Isa, dan L. Ashianti, *"Cyberbullying Classification using Text Mining," 1st International Conference on Informatics and Computational Sciences* (ICICoS). pp 241-246, 2017.

[6] E. Indrayuni, "Komparasi Algoritma *Naive Bayes* Dan *Support Vector Machine* Untuk Analisa Sentimen Review Film," *Journal of Computing and Information System* (PILAR), vol. 14, No.2, pp 175-180, 2018.

[7] F. F. Haranto and B. W. Sari,"Implementasi Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter Terhadap

Pelayanan Telkom Dan Biznet," Jurnal PILAR Nusa Mandiri, vol. 15, no. 2, pp 171-176, 2019.
[8] D. A. Agustina, S. Subanti, and E. Zukhronah, "Implementasi *Text Mining* Pada Analisis Sentimen Pengguna Twitter Terhadap Marketplace di Indonesia Menggunakan Metode *Support Vector Machine*," *Indonesian Journal of Applied* Statistics, vol. 3, no. 2, pp 109-122, 2020.
[9] I. Adiwijaya, "*Text Mining* dan *Knowledge Discovery*. Kolokium Bersama Komunitas Datamining Indonesia & Soft-Computing Indonesia", Sept. 2006.
[10] M. Hearst, "What Is Text Mining?," SIMS, *University of California, Berkeley*. Oktober, 2003.
[11] B. Liu, "*Sentiment Analysis and Opinion Mining*," Morgan & Claypool Publishers, May 2012.
[12] F. V. Sari and Wibowo, A, "Analisis Sentimen Pelanggan Toko Online JD.ID Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi," *Jurnal SIMETRIS*, vol.10, pp 681-686, 2019.
[13] S. C. M. de S Sirisuriya, "A Comparative Study on Web Scraping," Proceedings of 8th International Research Conference, KDU, P. Nov. 2015
[14] D. D. A. Yani, H. S. Pratiwi, and H. Muhardi, "Implementasi *Web Scraping* untuk Pengambilan Data pada Situs *Marketplace*," Jurnal Sistem dan Teknologi Informasi, vol. 7, no.4, pp 257-262.
[15] M. A. Fauzi, "*Text Pre-Processing*", [Online]. *Available*: http://malifauzi.lecture.ub.ac.id/files/2016/02/Text-Pre-Processing.pdf. [*Accessed* 12 November 2021].
[16] S. Jurniasih, E. F. Ripanti, and E. E. Pratama, "Implementasi *Naive Bayes Classifier* pada Opinion Mining Berdasarkan Tweets Masyarakat Terkait Kinerja Presiden dalam Aspek Ekonomi," *Jurnal Sistem dan Teknologi Informasi*, vol. 8, no. 3, pp. 239-249, 2020.
[17] P. M. Prihartini, "Implementasi Ekstraksi Fitur pada Pengolahan Dokumen Berbahasa Indonesia," *Jurnal Matrix*, vol.6, no.3, pp 174-178, 2016.
[18] F. S. Jumeilah, "Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian," *Jurnal RESTI* (*Rekayasa Sistem dan Teknologi Informasi*), vol. 1, no. 1, pp. 19 – 25, 2017.
[19] A. Nurzahputra and M. A. Muslim, "Analisis Sentimen pada Opini Mahasiswa Menggunakan *Natural Language Processing,"* in *Seminar Nasional Ilmu Komputer (SNIK 2016)*, Semarang, 2016.
[20] 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)," *Jurnal Media Informatika Budidarma*, vol. 5, no.2, pp 406-414, 2021.
[21] A. S. Nugroho, A. B. Witarto, and D. Handoko, "*Support Vector Machine* Teori dan Aplikasinya dalam Bioinformatika", [Online]. Available: https://asnugroho.net/papers/ikcsvm.pdf [Accessed 14 November 2021]
[22] A. Widjaya, L. Hiryanto, and T. Handhayani, "Prediksi Masa Studi Mahasiswa Dengan *Voting Feature Interval* 5 Pada Aplikasi Konsultasi Akademik *Online*," *Journal of Computer Science and Information Systems*, vol.1, pp 25-33, 2017.
[23] L. Sa'adah, "Analisis Sentimen Review E-Commerce pada Twitter Menggunakan dan Metode Klasifikasi Support Vector Machine. Tel-U Collection", [Online]. Available: https:/ /repository.telkomuniversity. ac.id/pustaka/158487/analisis-sentimen-review-e-commerce-pada-twitte r-menggunakan-metode-klasifikasi-support-vector-machine.html. [Accessed 14 November 2021]
[24] D. Iskandar and Y. K. Suprapto, "Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan Antara Algoritma C 4.5 Dan Naïve Bayes," Jurnal Ilmiah NERO, vol. 2, no.1, pp 37-43, 2015.
[25] M. F. Sanner, "*Python: A Programming Language for Software Integration and* Development", [Online]. Available: https://t.ly/9dpq [Accessed 20 November 2021]
[26] D. Khulman, "*A Python Book: Beginning Python, Advanced Python, and Python Exercises*", [Online]. Available: https://web.arch ive.org /web/20120623165941/ http://cutter.rexx.com/~dkuhlman/python_book_01.html#part-1-beginning-python. [Accessed 14 November 2021]
[27] H. Basri, "Peran Media Sosial Twitter dalam Interaksi Sosial Pelajar Sekolah Menengah Pertama di Kota Pekanbaru (Studi Kasus Pelajar SMPN 1 Kota Pekanbaru)," *Jom FISIP*, vol.4, no. 2, pp 1-15, 2017.
[28] D. Apriadi and A. Y. Saputra, "*E-Commerce* Berbasis *Marketplace* dalam Upaya Mempersingkat Distribusi Penjualan Hasil Pertanian," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol.1, no.2, pp 131-136, 2017.
[29] L. Adriani, H. Sujaini, and Tursina, "Implementasi *Sentiment Analysis* Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak," *Jurnal Sistem dan Teknologi Informasi*, vol. 8, no. 2, pp. 183-190, 2019