

ANALISIS SENTIMEN TERHADAP PEMBERIAN VAKSINASI MENGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR*

1st Fajar Aryanto
Universitas Buana Perjuangan Karawang
Karawang, Indonesia
if18.fajararyanto@mhs.ubpkarawang.ac.id
085695418616

2nd Ahmad Fauzi
Universitas Buana Perjuangan Karawang
Karawang, Indonesia
Afauzi@ubp.karawang.ac.id

3rd Anis Fitri Nur Masruriyah
Universitas Buana Perjuangan Karawang
Karawang, Indonesia
anis.masruriyah@ubpkarawang.ac.id

Abstract— Wabah virus *COVID-19* yang menyerang negara Indonesia membuat pemerintah menerapkan kebijakan yaitu pemberian vaksinasi. Sejak diumumkan kebijakan pemerintah terhadap pemberian vaksin *COVID-19* bulan Januari 2021 mulai banyak perbincangan terutama pada media sosial. Salah satu media sosial yang banyak dimanfaatkan masyarakat untuk beropini yaitu twitter. hal ini mengakibatkan twitter menjadi media menuangkan pemikiran masyarakat terhadap pemberian vaksin *COVID-19*. Opini yang ditimbulkan dapat bersifat positif, negatif, maupun netral terhadap pemberian vaksin. Berdasarkan uraian masalah tersebut, penelitian bertujuan untuk menganalisis sentimen dari warganet terkait kebijakan pemerintah dalam pemberian vaksinasi. Metode yang digunakan yaitu Algoritma KNN untuk melakukan analisis sentimen melalui opini masyarakat pada media sosial twitter. Hasil dari sentimen yang mempunyai nilai paling tinggi yaitu *positive* yang memiliki 150 kalimat lebih. Lalu menghasilkan akurasi 86.6% *Precision* 85% dan *Recall* 81%.

Kata kunci — *covid-19, vaksinasi, twitter.*

I. PENDAHULUAN

Wabah virus Covid-19 yaitu wabah besar yang melanda negara Indonesia dan tersebar diseluruh dunia. Kemunculan dari virus Covid-19 ini terjadi bermula di Kota Wuhan, Cina, sejak akhir Desember 2019. Wabah virus corona ini mengakibatkan dampak yang krisis terhadap perekonomian dunia. Indonesia termasuk negara yang paling banyak terinfeksi oleh virus covid-19, pada masa pandemi ini pemerintah mulai menerapkan berbagai kebijakan yang dilakukan, salah satunya yaitu kebijakan vaksinasi. Vaksinasi adalah proses suatu zat melalui suntikkan maupun ditetaskan ke dalam mulut untuk meningkatkan antibodi dan menangkal penyakit tertentu [1]. Vaksin ini berfungsi sebagai suatu antibodi atau kekebalan tubuh untuk mencegah penyakit dan menjaga kondisi tubuh dari virus dan bakteri [2]. Awal kebijakan pemerintah tentang pemberian vaksinasi banyak perbincangan masyarakat, bahwa vaksin ini memberikan dampak yang serius bagi tubuh. Maka dari itu menimbulkan berbagai opini yang disampaikan masyarakat dengan adanya pemberian vaksinasi di media sosial twitter [3].

Saat ini penggunaan media sosial sudah banyak digunakan dari kalangan masyarakat karena perkembangan teknologi yang berkembang sangat cepat [4]. Media sosial tidak hanya sebagai media berkomunikasi tetapi juga digunakan untuk berbagi informasi, berita acara, media bisnis, serta berfungsi sebagai sarana bagi masyarakat untuk beropini. Maka dari itu vaksinasi membuat sebagian besar masyarakat khawatir karena banyak berita yang beredar di media sosial khususnya twitter bahwa setelah melakukan vaksinasi dapat mengalami sesak nafas, pembekuan darah, badan terasa sakit, demam yang cukup tinggi dan lain-lain [5]. Oleh karena itu, pada penelitian tersebut ingin menganalisa jumlah respon positif, negatif atau netral dari masyarakat terhadap pemberian vaksinasi yang terdapat pada di media sosial twitter dengan metode Algoritma KNN [6]. Solusi yang diberikan yaitu penelitian tersebut memberikan gambar atau model visualisasi data yang didapatkan melalui Bahasa Pemrograman R atau Rstudio.

II. DATA DAN METODE

A. Bahan dan Peralatan

Untuk melakukan analisis sentimen bahan penelitian yang digunakan yaitu data twitter pada bulan Januari 2022 yang memuat informasi kebijakan pemerintah terhadap pemberian vaksinasi. Selain itu, peralatan yang dijadikan yaitu Perangkat Keras diantaranya, Laptop dengan spesifikasi AMD E2-6110 APU with AMD Radeon R2 Graphics. Lalu Perangkat Lunak yaitu *Microsoft Excel, Microsoft Word* dan Rstudio.

B. Prosedur Penelitian

Prosedur Penelitian diawali dengan melakukan Pengumpulan data untuk mengambil sebuah data dari media sosial twitter melalui Bahasa Pemrograman R atau Rstudio. Text preprocessing merupakan informasi data yang telah diperoleh akan dilakukan proses penyeleksian data dengan tujuan menjadikan data tidak terstruktur menjadi data terstruktur. Lalu tahapan model untuk menghasilkan suatu visualisasi yaitu berbentuk wordcloud dan histogram kata. Setelah itu proses analisis sentimen yang bertujuan untuk mengetahui

sebuah opini masyarakat apakah positif, netral dan negative. Pembobotan kata yaitu proses untuk mengetahui term frekuensi atau mencari nilai disetiap kata menggunakan metode TF-IDF. Klasifikasi menggunakan KNN dilakukan untuk mengetahui atau mencari nilai k di sebuah data. Evaluasi dilakukan untuk mengetahui tingkat akurasi algoritma menggunakan metode *confusion matrix* untuk menguji performa algoritma KNN.

C. Akuisisi Data

Akuisisi Data yaitu program yang digunakan oleh mesin pencari yang mengambil link halaman yang terdapat dalam web secara keseluruhan. Tujuan akuisisi data adalah untuk melihat interaksi antara pemerintah dan masyarakat melalui jumlah postingan dan jumlah komentar.

Tabel 1 Data awal twitter

No	Kalimat
1	"Pos vaksinasi tambahan bakal disiagakan pada saat mudik lebaran 2022"
2	"Booster jadi syarat mudik, antrean vaksinasi di Surabaya membeludak."
3	"@milodinosauruz_ Selamat malam, bagi pengguna Lansia (usia 60 Tahun ke atas) hanya diizinkan naik KRL pada pukul 10"
4	"Serda Rivo Prasetyo melaksanakan giat pendampingan Vaksinasi ccovac dosis 1 dengan menggunakan akun pcare Kesehatan"
5	"Pembagian Paket Sembako Bagi Warga Masyarakat Yang telah melaksanakan Vaksinasi Covid 19 di Lobi Pasar Blok A Tanah"
.	.
.	.
"10"	.
1000	"Pemerintah Terus Dorong Vaksinasi Anak Untuk Pulihkan Kegiatan Belajar"

D. Text Pre processing

Tahapan *text preprocessing* merupakan bertugas untuk membersihkan teks sebelum diolah lebih lanjut. Data teks mentah yang tidak terstruktur mempunyai cukup banyak *noise* seperti tanda baca, angka, imbuhan, karakter khusus, *slang word* dan sebagainya. Tahapan yang dilakukan diantaranya:

- 1) *Case folding*
Case folding dilakukan untuk merubah huruf kapital menjadi huruf kecil (lowercase).
- 2) *Tokenizing*
Tokenizing dilakukan penghilangan angka, tanda baca dan karakter lain yang dianggap tidak memiliki pengaruh terhadap pemrosesan teks.
- 3) *Filtering*
Filtering yang digunakan untuk membuang kata tidak penting dan mengambil kata penting, kata umum yang biasanya muncul dan tidak memiliki makna disebut dengan *stopword*.

E. Model

Tahapan model yaitu representasi visual berdasarkan data yang sudah diolah pada tahapan *text preprocessing* yang menghasilkan model berbentuk wordcloud dan histogram kata. Visualisasi pada *wordcloud* ini untuk menampilkan kata sering muncul pada twitter yang membahas vaksinasi.

F. Analisis Sentimen

Penelitian melakukan analisis sentimen terhadap respon masyarakat tentang pemberian vaksinasi dimedia sosial twitter. Analisis sentimen yaitu proses menganalisis dari sebuah dokumen teks yang sudah diolah proses sebelumnya dengan berbahasa Indonesia untuk mendapatkan informasi sentimen. Pada proses ini digunakan untuk menentukan sentimen lalu mengelompokkan teks ke dalam beberapa kelas, yaitu positif, negative, dan netral.

G. Pembobotan Kata

Tahapan ini yaitu menghitung bobot dari setiap *term* atau kata berdasarkan frekuensi kemunculan *term* tersebut dalam dokumen menggunakan metode TF-IDF. TF-IDF sebuah statistik numerik yang dapat menunjukkan kata kunci dengan kata tertentu. Selain dari itu, TF-IDF dapat mengetahui kata apa yang sering muncul pada suatu dokumen.

H. Algoritma KNN

Algoritma *K-Nearest Neighbor* yaitu metode untuk mengklasifikasi terhadap objek K berdasarkan pembelajaran yang paling dekat dengan objek tersebut. Algoritma KNN termasuk kelompok *instance based learning* dan juga salah satu teknik *lazy learning*. *K-Nearest Neighbor* merupakan suatu pendekatan untuk mencari kasus baru maupun kasus lama dengan menghitung berdasarkan pencocokan bobot. Algoritma *K-Nearest Neighbor* berguna untuk mengklasifikasikan objek berdasarkan atribut dan data training. Algoritma K-NN merupakan sebuah model *supervised learning* yang mana database populasi yang dipakai diidentifikasi kedalam kelas tertentu. Algoritma ini juga mempunyai prinsip yang sederhana, bekerja berdasarkan jarak terpendek dari sampel uji ke sampel latih. Ada beberapa tahapan algoritma KNN, sebagai berikut:

1. Menentukan nilai K.
2. Menentukan nilai K yang paling besar.
3. Mengurutkan dari nilai K tinggi ke rendah.
4. Memprediksikan kategori objek dengan menggunakan kategori *nearest neighbors* yang paling mayoritas.

Pada penelitian ini akan mengimplementasikan klasifikasi algoritma *K-Nearest Neighbors* dengan penyelesaian dalam program R.

I. Evaluasi

Pada proses terakhir yaitu evaluasi yang menggunakan *confusion matrix* untuk melakukan tahapan pengujian. Proses evaluasi menggunakan *confusion matrix* banyak digunakan pada penelitian yang menggunakan algoritma K-NN. Proses evaluasi berguna untuk menghitung tingkat akurasi algoritma. Selain tingkat akurasi, dengan *confusion matrix* juga dapat menghitung nilai *precision*, dan *recall*.

III. HASIL DAN PEMBAHASAN:

A. Data hasil *Text Pre processing*

Data preprocessing data didapatkan dari proses implementasi twitter ke dalam sistem sebanyak 1000 data. Tahapan text preprocessing diantaranya case folding, tokenizing dan filtering. Hasil dari text preprocessing menghasilkan data pada tabel 2.

Tabel 2 Hasil Text Preprocessing

No	Proses	Sebelum	Sesudah
1	<i>Case Folding</i>	<p>“Teman-teman sudah tahu belum sekarang vaksinasi Covid bisa diberikan untuk anak usia 6-11 tahun?”</p> <p>“Cari tahu cara daftar vaksinasi anak melalui JAKI, dan apa saja dokumen yang harus disiapkan?”</p>	<p>teman-teman sudah tahu belum sekarang vaksinasi covid bisa diberikan untuk anak usia 6-11 tahun?</p> <p>cari tahu cara daftar vaksinasi anak melalui jaki, dan apa saja dokumen yang harus disiapkan?</p>
2	<i>Tokenizing</i>	<p>“Teman-teman sudah tahu belum sekarang vaksinasi Covid bisa diberikan untuk anak usia 6-11 tahun?”</p> <p>“Cari tahu cara daftar vaksinasi anak melalui JAKI, dan apa saja dokumen yang harus disiapkan?”</p>	<p>teman-teman sudah tahu belum sekarang vaksinasi covid bisa diberikan untuk anak usia - tahun?</p> <p>cari tahu cara daftar vaksinasi anak melalui jaki, dan apa saja dokumen yang harus disiapkan?</p>
3	<i>Filtering</i>	<p>“Teman-teman sudah tahu belum sekarang vaksinasi Covid bisa diberikan untuk anak usia 6-11 tahun?”</p> <p>“Cari tahu cara daftar vaksinasi anak melalui JAKI, dan apa saja dokumen yang harus disiapkan?”</p>	<p>temanteman sudah tahu belum sekarang vaksinasi covid bisa diberikan untuk anak usia tahun</p> <p>cari tahu cara daftar vaksinasi anak melalui jaki dan apa saja dokumen yang harus disiapkan</p>

B. Implementasi Algoritma

Implementasi algoritma data yang digunakan yaitu data twitter pada bulan Maret 2022 dalam melakukan analisis sentimen. Data yang dipakai dalam klasifikasi KNN yaitu 1000 data dengan menentukan jarak nilai K tetangga terdekat. Tahapan diawali dengan tidy text pada gambar 1,2 dan 3.

no	text
1	Pos vaksinasi tambahan bakal disiagakan pada saat mudik lebaran 2022 #KapoliListyoSigitPrab...
2	Booster jadi syarat mudik, antrean vaksinasi di Surabaya membeludak. https://t.co/538hgFDnn5
3	@milodinosauruz_ Selamat malam, bagi pengguna Lansia (usia 60 Tahun ke atas) hanya diizinkan...
4	Serda Rivo Prasetyo melaksanakan giat pendampingan Vaksinasi ccovac dosis 1 dengan mengg...
5	Pembagian Paket Sembako Bagi Warga Masyarakat Yang telah melaksanakan Vaksinasi Covid 1...
6	RT @Deelee_007: UMAT ISLAM DIAM saat -NATARU 2021 PPKM dihapus (Des 21-Jan 22) -IML...
7	RT @PolsekPagu: Percepatan Vaksinasi Polres Kediri @pidpolreskediri #polreskediri #kediri #pol...
8	Dan sampai sekarang mereka adalah temak dan mereka siap menjadi tentara untuk setiap age...
9	Percepatan Vaksinasi Polres Kediri @pidpolreskediri #polreskediri #kediri #polripresisi #logola...
10	RT @KKMPutrajaya: Definisi status vaksinasi bermula 1 April 2022. https://t.co/jKiqMQmHbj
11	RT @Abe_Mukti: MUI menetapkan vaksin merah putih halal. Ini kabar bagus. Semoga dapat me...
12	Kunjungan Bupati Mamuju tersebut dalam rangka bakti sosial, vaksinasi COVID-19, serta pemb...
13	RT @GratisTerbaik: \Vaksinasi booster tampaknya kurang diminati warga
14	Lucu. Seorang Napi Teriak Histeris saat Hendak Vaksinasi di Rutan Apa ... https://t.co/jt1W2Y53...

Gambar 1 Data Twitter

no	baris	word
1	1	pos
2	2	1 vaksinasi
3	3	1 tambahan
4	4	1 disiagakan
5	5	1 mudik
6	6	1 lebaran
7	7	2 booster
8	8	2 syarat
9	9	2 mudik
10	10	2 antrean
11	11	2 vaksinasi
12	12	2 surabaya
13	13	2 membeludak
14	14	3 selamat

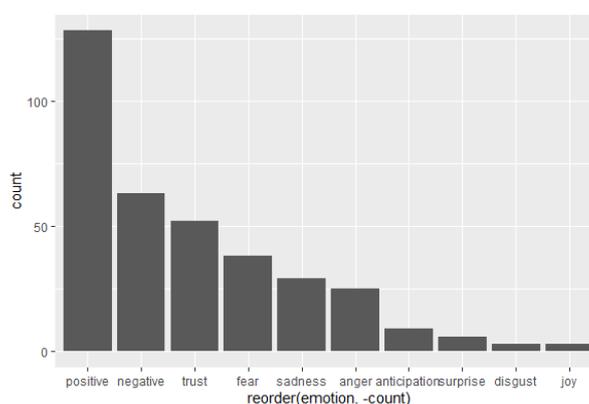
Gambar 2 Data Pengelompokkan *tidy text*

Data awal twitter kemudian dilakukan, pencarian kata sering muncul menggunakan TF IDF diawali dengan proses tidy text untuk melakukan pengelompokan sentimen ditunjukkan pada gambar 3.

word	n	
1	vaksinasi	732
2	booster	181
3	vaksin	160
4	covid	158
5	polr	116
6	indonesia	82
7	malaysia	82
8	status	82
9	warga	81
10	mudik	80
11	sinovac	79
12	dosi	67
13	masyarakat	66
14	melaksanakan	65

Gambar 3 Frekuensi *tidy text*

Setelah melakukan pencarian kata sering muncul selanjutnya proses sentimen pada suatu data twitter yang sudah diolah sebelumnya. Hasil pengelompokkan sentimen terlihat pada gambar 6.



Gambar 4 Visualisasi analisis sentimen

Setelah sudah melakukan pengelompokan sentimen selanjutnya mengelompokkan pembobotan kata yang menggunakan metode TF-IDF dengan frekuensi kata yang telah didapatkan.

baris	word	wordTotal	n	tf	idf	tf_idf
1	lebaran	23	6	0.0303030303	4.043051	0.122516705
2	mudik	80	6	0.0087082729	4.043051	0.035207994
3	vaksinasi	732	6	0.0009062075	4.043051	0.003663843
4	booster	181	7	0.0040603248	3.755369	0.015248019
5	mudik	80	7	0.0101596517	3.755369	0.038153243
6	syarat	37	7	0.0222929936	3.755369	0.083718422
7	vaksinasi	732	7	0.0010572421	3.755369	0.003970334
8	ata	21	8	0.0365296804	5.141684	0.187823326
9	dosi	67	12	0.0179640719	3.532226	0.063453155
10	giat	21	12	0.0510638298	3.532226	0.180368969
11	kesehatan	43	12	0.0247422680	3.532226	0.087395274
12	melaksanakan	65	12	0.0156453716	3.532226	0.055262983
13	vaksinasi	732	12	0.0018124150	3.532226	0.006401859
14	covid	158	12	0.0080213904	3.532226	0.028333361
15	masyarakat	66	12	0.0205479452	3.532226	0.072579979
16	melaksanakan	65	12	0.0156453716	3.532226	0.055262983
17	vaksinasi	732	12	0.0018124150	3.532226	0.006401859
18	warga	81	12	0.0165745856	3.532226	0.058545176

Gambar 5 Hasil TF-IDF

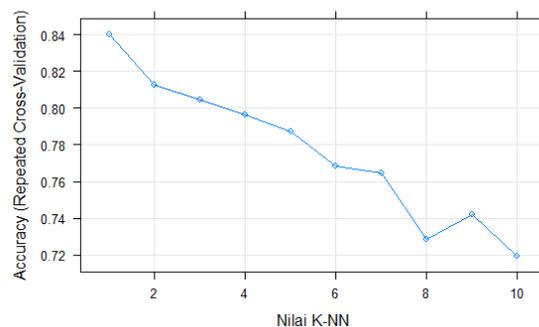
Tahapan klasifikasi pada algoritma KNN untuk mengetahui jarak nilai k dari tetangga terdekatnya. Nilai k yang digunakan yaitu nilai k 1 sampai nilai k 10 yang ditunjukkan pada tabel 3.

Tabel 3 Klasifikasi KNN

K	Accuracy	Kappa
1	0.8402630	0.7494160
2	0.8124906	0.7072022
3	0.8043825	0.6994380
4	0.7965072	0.6883668
5	0.7871700	0.6758114
6	0.7686257	0.6469452
7	0.7645697	0.6414396
8	0.7284363	0.5875599
9	0.7418816	0.6057878
10	0.7191926	0.5718328

Dari hasil Tabel 3 diketahui bahwa terdapat nilai k 1 sampai 10 serta dengan hasil akurasi. Hasil nilai akurasi menunjukkan rentang 0.84 sampai 0.71 dari nilai k 1 sampai 10. Nilai k = 1 memperoleh nilai tertinggi yaitu 0.84 atau 84%, dan nilai k = 10 memperoleh nilai terendah yaitu 0.71 atau 71%.

Grafik Akurasi Nilai k-



Gambar 6 Visualisasi hasil Algoritma KNN

C. Evaluasi

Pada tahapan yang telah dilakukan dengan klasifikasi KNN, oleh karena itu hasil klasifikasi KNN akan dievaluasi menggunakan confusion matrix. Pada algoritma KNN menggunakan 1000 data latih. Lalu menggunakan 752 data sebagai data training dan 248 data sebagai data testing. Berikut Hasil evaluasi menggunakan confusion matrix yang ditunjukkan pada tabel 4.

Tabel 4 confusion matrix

Prediction	negative	neutral	positive
negative	63	2	7
neutral	7	47	3
positive	8	6	105

Berdasarkan tabel 4 informasi dari hasil pengujian klasifikasi *K-Nearest Neighbor* terhadap data testing terdapat tabel matriks. Hasil tersebut ditampilkan perhitungan hasil kinerja dengan mencari nilai *Accuracy*, *Precision*, *Recall* sebagai informasi tambahan dan memperjelas. Untuk memperoleh nilai *Accuracy*, *Precision*, dan *Recall*. Maka membutuhkan rumus perhitungan dibawah ini.

$$\begin{aligned}
 a. \text{ Accuracy} &= \frac{TP}{\text{Jumlah Data}} \\
 &= \frac{63+47+105}{248} \\
 &= \frac{215}{248} = 0.866 \\
 &= 0.866 \times 100 = 86.6\%
 \end{aligned}$$

$$\begin{aligned}
 b. \text{ Precision} &= \frac{TP}{TP+FP} \\
 &= \frac{2.57}{3} = 0.85 \\
 &= 0.85 \times 100 = 85\%
 \end{aligned}$$

$$\begin{aligned}
 c. \text{ Recall} &= \frac{TP}{TP+FN} \\
 &= \frac{2.43}{3} = 0.81 \\
 &= 0.81 \times 100 = 81\%
 \end{aligned}$$

IV. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian maka diperoleh kesimpulan yaitu menghasilkan nilai sentimen yang dihasilkan tweet kebijakan pemerintah terhadap pemberian vaksinasi yaitu bersentimen positif. Lalu pada evaluasi pengujian model klasifikasi algoritma KNN dengan *confusion matrix*. Dari pengujian tersebut diperoleh akurasi sebesar 86.6%, *precision* sebesar 85%, *recall* sebesar 81%.

Saran untuk penelitian selanjutnya yaitu model analisis sentimen yang dibuat dapat dikembangkan lebih jauh untuk otomatisasi analisis sentimen terkait kebijakan pemerintah yang lainnya.

PENGAKUAN

Naskah ilmiah ini adalah Sebagian dari penelitian Tugas Akhir milik Fajar Aryanto dengan judul Analisis Sentimen Kebijakan Pemerintah Terhadap Pemberian Vaksinasi Pengguna Twitter Menggunakan Algoritma *K-Nearest Neighbor* yang dibimbing oleh Bapak Dr. Ahmad Fauzi, M.Kom dan Ibu Anis Fitri Nur Masruriyah, M.Kom.

DAFTAR PUSTAKA

- [1] Agrani, A., & Dr. Brady Rikumahu, S.E., M. (2020). Perbandingan Analisis Sentimen Terhadap Digital Payment “Go_Pay” Dan “Ovo” Di Media Sosial Twitter Menggunakan Algoritma Naive Bayes Dan Word Cloud. *E-Proceeding of Management*, 7, No.2
- [2] Alpito, D., Etris, R., Fikri, & Sadyanti, K. (2021). Respon Masyarakat Terhadap Vaksin Covid-19. *Seminar Nasional Karya Ilmiah Multidisiplin*, Vol. 1, No.
- [3] Claudy, Y. I., Perdana, R. S., & Fauzi, M. A. (2018). Klasifikasi Dokumen Twitter Untuk Mengetahui Karakter Calon Karyawan Menggunakan Algoritma K-Nearest Neighbor (KNN). *Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2, No. 8.
- [4] Darwis, D., & Dkk. (2020). Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia. *Ilmiah Edutic*, 7, No.1.

- [5] Fitriana, F., Utami, E., & Fatta, H. Al. (2021). Analisis Sentimen Opini Terhadap Vaksin Covid-19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes. *Komputasi Dan Informatika*, . 5 No. 1
- [6] Gunawan, B., Pratiwi, H. S. P., & Esyudha, E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Edukasi Dan Penelitian Informatika*, 4 No. 2..