

PERBANDINGAN ALGORITMA SVM DAN NBC DALAM ANALISA SENTIMEN PILKADA PADA TWITTER

COMPARISON OF SVM AND NBC ALGORITHM IN SENTIMENT ANALYSIS OF LOCAL LEADERS
ELECTION ON TWITTER

Elisabet Sinta Romaito^{*1}, M. Khairul Anam², Rahmaddeni³, Aniq Noviciate Ulfah⁴

^{*1,3,4} Jurusan Teknik Informatika, STMIK Amik Riau, Pekanbaru

²Jurusan Teknologi Informasi, STMIK Amik Riau, Pekanbaru

e-mail: ^{*1}1710031802042@stmik-amik-riau.ac.id, ²khairulanam@sar.ac.id,

³rahmaddeni@sar.ac.id, ⁴aniqnoviciatieulfah@sar.ac.id

Abstrak

Pemilihan Kepala Daerah atau PILKADA merupakan hal yang wajib dilakukan setiap 5 tahun sekali, Dimana masyarakat sangat berperan penting dalam PILKADA. Dalam setiap PILKADA tidak semua masyarakat dapat menerima setiap calon kepala daerah dengan baik, sering sekali terjadi pro dan kontra masyarakat terhadap PILKADA dimedia sosial terutama media sosial twitter. Penelitian ini bertujuan untuk melihat perbandingan algoritma support vector machine dan naive bayes classifier dalam analisa sentimen PILKADA berdasarkan data Twitter. Hasil yang diharapkan dari penelitian ini adalah mengetahui perbandingan akurasi dari algoritma SVM dan NBC. Perbandingan yang didapatkan adalah Accuracy 81,7 recall 81,7 dan precision 80% adalah hasil NBC, Sedangkan Accuracy 80,7 recall 80,7 dan precision 84% adalah hasil SVM. Maka dapat disimpulkan bahwa algoritma NBC lebih unggul dalam accuracy & recall sedangkan dalam precision yang lebih unggul adalah algoritma SVM.

Kata kunci— Twitter, PILKADA, Support Vector Machine, Naïve Bayes Classifier

Abstract

Local leaders election or PILKADA is something that has to do every five years, where the community plays an important role in the Regional Head Election. In every PILKADA not all community can accept the candidates for regional head well, there are often pro and contra of the community towards the PILKADA in social media, especially social media twitter. This study aims to compare support vector machine algorithm and Naive Bayes classifier in PILKADA sentiment analysis based on Twitter data. The expected result of this research is to know the accuracy comparison of SVM algorithm and NBC. The comparisons obtained were Accuracy 81.7 recall 81.7 and precision 80% were the results of NBC, while Accuracy 80.7 recall 80.7 and precision 84% were the results of SVM. So it can be denied that the NBC algorithm is superior in accuracy & recall, while the superior precision is the SVM algorithm.

Keywords— Twitter, PILKADA, Support Vector Machine, Naïve Bayes Classifier

1. PENDAHULUAN

Pemilihan umum atau pemilu adalah salah satu mekanisme yang ada pada salah satu tahap penyelenggaraan negara dan pembentukan pemerintahan Indonesia [1]. Salah satunya yaitu termasuk PILKADA atau Pemilihan Kepala Daerah. PILKADA adalah wujud nyata dari demokrasi pada rakyat Indonesia terkhusus nya pada pemerintahan daerah. Pemilihan kepala daerah secara

langsung telah diadakan mulai tahun 2005 sampai pada tahun 2020, berdasarkan ketentuan UU No. 32 Tahun 2004 yang berlandaskan pada Pasal 18 ayat (4) UUD 1945 yang mengesahkan bahwa Gubernur, Bupati, dan Walikota sebagai kepala pemerintahan daerah provinsi, kabupaten, dan kota dipilih secara demokratis [2]. PILKADA sering menjadi pro dan kontra di tengah masyarakat Indonesia tidak jarang Pilkada sering menjadi trending topic di media sosial twitter.

Media sosial *twitter* adalah media sosial dengan jutaan pengguna yang mampu memberikan wawasan unik tentang penggunaannya dan salah satu media yang populer dalam berinteraksi dengan sesama [3]. Data berupa opini tersebut disediakan oleh twitter yang kita bisa ambil datanya, data opini tersebut yang akan kita olah dan analisa agar menghasilkan data atau informasi sehingga informasi yang didapatkan membantu banyak pihak untuk mendukung pilihan. Dari data yang telah didapatkan maka akan dilakukan pengelompokan komentar berdasarkan dua kelompok yaitu komentar positif dan komentar negatif. *Twitter* menjadi tempat untuk mendapatkan data dari opini yang diungkapkan masyarakat melalui *tweets* [4].

Analisis sentimen adalah bagian dari teks mining yang dapat mengelompokkan polaritas dari teks yang dilakukan untuk melihat bagaimana polaritas dari suatu *tweet* [5]. Digunakan untuk mengungkap opini publik terhadap suatu isu pemberi opini yang banyak mendapat respon dari pengguna *twitter*. Pada penelitian ini menggunakan metode klasifikasi teks Support Vector Machine (SVM) dan Naïve Bayes Classifier (NBC).

Support Vector Machine (SVM) adalah suatu teknik untuk melakukan klasifikasi yang sangat populer belakangan ini [6]. metode ini mempelajari area yang memisahkan antar kategori dalam sebuah observasi. Sedangkan *Naïve bayes classifier (NBC)* merupakan metode sederhana tetapi memiliki akurasi dan performansi yang tinggi dalam pengklasifikasian teks [7]. Pada penelitian ini penulis akan melakukan perbandingan antara metode Support Vector Machine (SVM) dan *Naïve bayes classifier (NBC)* terhadap PILKADA. Hasil dari penelitian ini nantinya akan menunjukkan akurasi metode yang tertinggi dalam pengklasifikasian data. Di buktikan pada penelitian terdahulu yang di telah dilakukan oleh beberapa peneliti.

Arifin & Sasongko [8] melakukan Analisa Perbandingan pada Tingkat Performansi menggunakan Metode *Support Vector Machine* Dan *Naïve Bayes Classifier* Untuk mengklasifikasi Jalur Minat pada siswa/i SMA. Pada penelitian ini digunakan data training untuk membentuk model klasifikasi yaitu nilai peserta didik pada saat mendaftar di SMA pada tahun pelajaran 2013-2014 dengan jumlah 288 siswa. Siswa dengan jurusan IPA berjumlah 150 siswa dan siswa dengan jurusan IPS sebanyak 138 siswa. Dengan menggunakan alat bantu RapidMiner studio 8.0 sehingga penelitian ini menghasilkan algoritma SVM (Support Vector Machine) dengan kernel anova dan parameter C sebesar 5.0 relatif lebih unggul dibandingkan algoritma Naïve Bayes Classifier.

Penelitian lainnya yang menggunakan metode *NBC* adalah S.Wulan, Dkk [9] melakukan penelitian untuk melihat evaluasi kinerja dosen menggunakan kusioner agar mampu mencapai hasil yang maksimal. Penelitian senada juga dilakukan oleh Y.Cahyono [10] yang menggunakan metode *NBC feature selection PSO* dengan memisahkan tweet berbahasa Indonesia pada media sosial twitter.

Peneliti L.D.Mahbubah [11] yang menggunakan media sosial *twitter* pada penelitiannya dan mengolah data *tweets* pada #pilpres2019 dan #prabowo yang menghasilkan akurasi sebesar 73%. Penelitian tentang PEMILU juga dilakukan oleh S.N.J.Fitriyyah, Dkk [12] yang mengklasifikasi opini masyarakat pada calon presiden 2019 menggunakan media sosial twitter untuk melihat opini masyarakat pada PEMILU 2019.

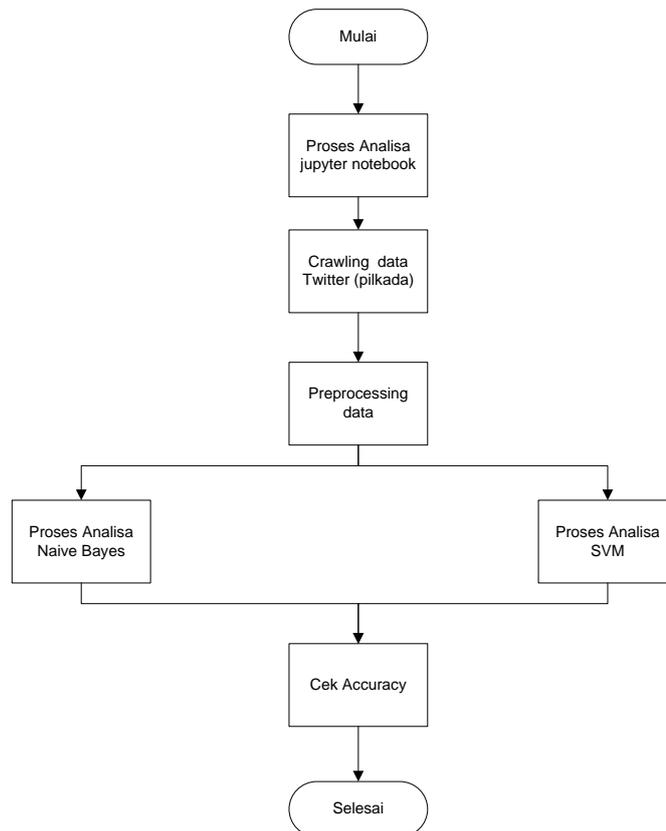
Peneliti M.F.Fibrianda and Bhawiyuga [13] juga melakukan analisa perbandingan dua metode dengan melakukan perbandingan akurasi deteksi serangan pada jaringan komputer. Penelitian ini menggunakan metode behavior based yang dalam proses kerjanya membutuhkan sebuah dataset dan metode. Dataset yang digunakan dalam penelitian ini adalah dataset dari ISCX2012 testbed tanggal 14 Juni 2012, dan metode yang digunakan dalam penelitian ini adalah metode NBC dan SVM. Hasil penelitian ini adalah perbandingan metode proses klasifikasi berdasarkan nilai precision, recall, akurasi confusion matrix dan f1 score. Naive Bayes, SVM Linear, SVM Polynomial dan SVM Sigmoid menghasilkan akurasi dengan persentase yang berurutan sebesar

85,055%, 99,995%, 99,999%, dan 99,995%.

Berdasarkan penelitian-penelitian yang pernah dilakukan sebelumnya maka penulis mengangkat sebuah penelitian tentang perbandingan metode naïve bayes classifier dengan metode support vector machine dalam analisa sentimen PILKADA pada media sosial twitter.

2. METODE PENELITIAN

Makalah Langkah-langkah pada penelitian ini dapat dilihat pada gambar 1 dibawah ini :



Gambar 1. Alur Proses Analisa Sentimen

2.1 Proses Analisa dengan Jupyter Notebook

Proses analisa data dimulai dengan menginstall aplikasi anaconda 3 dan menjalankan jupyter notebook untuk melakukan analisa sentimen dengan menggunakan bahasa pemrograman python.

2.2 Crawling Data

Proses crawling data dari twitter menggunakan library tweepy dengan twitter api untuk mengambil data dari twitter menggunakan query pilkada 2020. Untuk langkah detail dari proses crawling akan dijelaskan per poin sebagai berikut :

1. Buka jupyter notebook dan install !pip install tweepy. kemudian proses instalasi berjalan dan tunggu hingga selesai.
2. Lakukan proses crawling dengan memasukan keyword untuk masing-masing pilkada.
3. Maka secara default hasilnya dapat ditemukan pada C:/users/user/folder dalam format csv file.

4. Hasil dari proses crawling dalam bentuk .csv akan dilakukan pembersihan data sebelum data di proses dengan metode naive bayes classifier dan support vector machine, pembersihan data akan menghapus kolom-kolom yang lain dan menyisakan kolom teks dari tweet.

2.3 Preprocessing Data

Tahapan preprocessing data akan menggunakan proses casefolding, cleansing, normalisasi, tokenizing, stemming dan indexing dengan menggunakan tf-idf.

Tabel 1. Contoh data komentar

No	Komentar	Kelas
1	@enjipanges Hi Anggie, saat ini Pilkada sudah bisa dilakukan lho Kak, kuy lgsg cek Infonya. Thanks. -FN-	Positif
2	@danasd gak bisa apa pilkada diundur ini lagi pandemi.!!!	Negatif
3	@MallsCS #pilkada misi untuk produk apa aja yg bisa bayar pake Ovo ?	Netral

Pada Tabel 1 adalah contoh data komentar yang akan dipakai dalam preprocessing data pada penelitian ini. Dari data inilah peneliti akan melakukan tahapan preprocessing.

Tabel 2. Case folding

No	Komentar	Hasil Case Folding
1	@enjipanges Hi Anggie, saat ini Pilkada sudah bisa dilakukan lho Kak, kuy lgsg cek Infonya. Thanks. -FN-	@enjipanges hi anggie, saat ini pilkada sudah bisa dilakukan lho kak, kuy lgsg cek infonya. thanks. -fn-
2	@danasd gak bisa apa pilkada diundur ini lagi pandemi.!!!	@danasd gak bisa apa pilkada diundur ini lagi pandemi.!!!
3	@MallsCS #pilkada misi untuk produk apa aja yg bisa bayar pake Ovo ?	@mallscs #pilkada misi untuk produk apa aja yg bisa bayar pake ovo ?

Pada tabel 2 merupakan hasil data komentar yang telah di-Case Folding. Tahap case folding merupakan untuk mengubah seluruh huruf yang ada pada dokumen menjadi huruf kecil [14]

Tabel 3. Cleaning

No	Komentar	Hasil Case Folding
1	@enjipanges hi anggie, saat ini pilkada sudah bisa dilakukan lho kak, kuy lgsg cek infonya. thanks. -fn-	hi anggie saat ini pilkada sudah bisa dilakukan lho kak kuy lgsg cek infonya thanks fn
2	@danasd gak bisa apa pilkada diundur ini lagi pandemi.!!!	gak bisa apa pilkada diundur ini lagi pandemi
3	@mallscs #pilkada misi untuk produk apa aja yg bisa bayar pake ovo ?	misi untuk produk apa aja yg bisa bayar pake ovo

Pada tabel 3 merupakan hasil data komentar yang telah di cleaning, Cleaning adalah tahapan membersihkan dokumen dari komponen-komponen yang tidak memiliki hubungan dengan informasi yang ada pada dokumen, seperti karakter atau simbol, angka, emoticon dan link URL [15].

Tabel 4. Tokenizing

D1	D2	D3
hi	gak	misi
anggie	bisa	untuk
saat	apa	produk
ini	pilkada	apa
pilkada	diundur	aja
sudah	ini	yg
bisa	lagi	bisa
dilakukan	pandemi	bayar
lho		pake
kak		ovo
kuy		
lgsg		
cek		
infonya		
thanks		

Pada Tabel 4 merupakan hasil data komentar yang telah di Tokenizing. Tokenizing yaitu proses untuk mengelompokkan isi teks yang semula berupa kalimat hingga menjadi satuan kata-kata [15].

Tabel 5. Normalisasi

D1	D2	D3
hi	gak	misi
anggie	bisa	untuk
saat	apa	produk
ini	pilkada	apa
pilkada	diundur	aja
sudah	ini	yg
bisa	lagi	bisa
dilakukan	pandemi	bayar
lho		pake
kak		ovo
kuy		
lgsg		
cek		
infonya		
thanks		

Pada tabel 5 merupakan hasil data komentar yang telah di Normalisasi. Proses normalisasi yaitu mengubah kata yang tidak baku atau sesuai ejaan, pada tahap ini dilakukan dengan menggunakan database kamus kata Bahasa baku dan tidak baku yang dibuat sendiri berdasarkan dari data komentar yang digunakan [13].

Tabel 6. Filtering

D1	D2	D3
anggie	gak	misi
saat	bisa	untuk
pilkada	pilkada	produk
sudah	diundur	bisa
bisa	lagi	bayar
dilakukan	pandemi	pake
cek		ovo
infonya		

Pada tabel 6 merupakan hasil data komentar yang telah di Filtering. Tahapan filtering yaitu proses membuang kata yang tidak penting dari proses tokenizing sebelumnya. Kata akan di stopwords menggunakan database kamus KBBI V online berdasarkan jenis kata yang akan di stopwords seperti kata keterangan, kata seru, kata ganti. Kata depan dan kata hubung [16].

Tabel 7. Stemming

D1	D2	D3
anggie	gak	misi
saat	bisa	untuk
pilkada	pilkada	produk
sudah	undur	bisa
bisa	lagi	bayar
laku	pandemi	pake
cek		ovo
info		

Pada tabel 7 merupakan hasil data komentar yang telah di Stemming. Pada tahap stemming ini, akan dilakukan proses mencari kata dasar dari setiap kata hasil proses filtering sebelumnya. Stemming akan melakukan proses pengembalian berbagai bentukan kata menjadi bentukan kata dasar dengan menghilangkan imbuhan.

Tabel 8. Hasil Pembobotan

Term (t)	TF		
	POSITIF	NEGATIF	NETRAL
anggie	1	0	0
saat	1	0	0
pilkada	1	1	0
sudah	1	0	0
bisa	1	1	1
laku	1	0	0
cek	1	0	0
info	1	0	0
gak	0	1	0
bisa	0	1	0
undur	0	1	0
lagi	0	1	0

pandemi	0	1	0
misi	0	0	1
untuk	0	0	1
produk	0	0	1
bisa	0	0	1
bayar	0	0	1
pake	0	0	1
ovo	0	0	1
TOTAL	8	7	8

Pada tabel 8 adalah proses dari pembobotan kata. Pada proses pembobotan dilakukan perubahan kata menjadi bentuk angka atau vector. Sedangkan proses TF (term frekuensi) dilakukan untuk menentukan nilai frekuensi sebuah kata di dalam sebuah dokumen dan pemberian bobot pada setiap kata pada setiap dokumen untuk mencari dan menghitung berapa kali kata itu muncul.

A. Analisa Sentimen Naive Bayes Classifier

Analisa sentimen menggunakan metode naive bayes classifier akan dilakukan dengan python yang memberikan hasil seperti classification report, recall, precision, dan accuracy. Dimana proses analisa pada naïve bayes classifier ini menggunakan persamaan dengan 3 kategori yaitu positif, negatif, dan netral.

$$P(\text{Pos/Neg/Net}) = \frac{d(\text{Pos/Neg/Net})}{|c|} \quad (1)$$

$$P(\text{Pos}) = \frac{d(\text{Pos})}{|c|} = \frac{1}{3} \quad (2)$$

$$P(\text{Neg}) = \frac{d(\text{Neg})}{|c|} = \frac{1}{3} \quad (3)$$

$$P(\text{Net}) = \frac{d(\text{Net})}{|c|} = \frac{1}{3} \quad (4)$$

B. Analisa Sentimen Support Vector Machine

Analisa sentimen menggunakan metode Support Vector Machine akan dilakukan dengan python yang memberikan hasil seperti classification report, recall, precision, dan accuracy.

Akurasi, Presisi dan Recall

Akurasi didefinisikan sebagai tingkat kedekatan antara nilai prediksi dan aktual. Presisi memperlihatkan tingkat ketepatan dalam pengklasifikasian. Sedangkan fungsi recall adalah untuk mengukur proporsi positif aktual yang teridentifikasi. Confusion matrix digunakan untuk mengukur recall, presisi dan akurasi. Confusion matrix merupakan alat ukur berbentuk matrix untuk mendapatkan jumlah ketepatan klasifikasi pada kelas dengan algoritma. Bentuk confusion matrix disajikan pada tabel 9 dibawah ini:

Tabel 9. Bentuk Confusion Matrix Dari Dua Kelas

Confusion Matrix		Nilai Sebenarnya	
		TRUE	FALSE
Nilai Prediksi	TRUE	TP(True Positive) Correct result	FP(False Positive) Unexpected result
	False	FN(False Negative) Missing result	TN(True Negative)Cor rect absence of result

Pada Tabel 9 nilai TP (true positive) dan TN (true negative) memperlihatkan tingkat ketepatan pada klasifikasi. Umumnya semakin tinggi nilai TP dan TN maka akan semakin baik tingkat klasifikasi dari recall, akurasi dan presisi. Jika label prediksi keluaran adalah benar dan nilai seharusnya bernilai salah maka disebut false positive (FP). Sebaliknya jika prediksi label keluaran adalah salah dan nilai seharusnya bernilai benar maka hal ini disebut false negative (FN).

2.4 Cek Accuracy

Proses cek accuracy dilakukan dengan perhitungan precision, recall, f1-score dan support.

```

precision    recall  f1-score   support

 0         0.87    0.85    0.86     93
 1         0.77    0.37    0.50     27
 2         0.75    0.91    0.82     77

 accuracy   0.80
 macro avg  0.80    0.71    0.73
 weighted avg 0.81    0.81    0.80

SVM Accuracy --> 80.71065989847716
SVM Recall --> 80.71065989847716
SVM Precision --> 84.59966732440942
    
```

Gambar 2. Hasil Accuracy SVM

Berdasarkan hasil perhitungan, didapatkan accuracy SVM sebesar 80,7%, recall sebesar 80,7% dan Precision sebesar 84,6%.

```

precision    recall  f1-score   support

 0         0.84    0.87    0.86     93
 1         0.67    0.44    0.53     27
 2         0.82    0.88    0.85     77

 accuracy   0.78
 macro avg  0.78    0.73    0.75
 weighted avg 0.81    0.82    0.81

NBC Accuracy --> 81.7258883248731
NBC Recall --> 81.7258883248731
NBC Precision --> 80.99141489817136
    
```

Gambar 3. Hasil Accuracy NBC

Berdasarkan hasil perhitungan, didapatkan accuracy NBC sebesar 81,7%, recall sebesar 81,7% dan Precision sebesar 81%

3. HASIL DAN PEMBAHASAN

Berdasarkan tahapan yang telah dilakukan maka didapatkanlah sebuah diagram yang mendeskripsikan data hasil dari analisa.

- Inform. Polinema*, vol. 6, no. 2, pp. 33–39, 2020, doi: 10.33795/jip.v6i2.316.
- [5] R. KURNIAWAN and A. APRILIANI, “Analisis Sentimen Masyarakat Terhadap Virus Corona Berdasarkan Opini Dari Twitter Berbasis Web Scraper,” *Jurnal INSTEK (Informatika Sains dan Teknologi)*, vol. 5, no. 1, p. 67, 2020, doi: 10.24252/instek.v5i1.13686.
- [6] A. S. Ritonga and E. S. Purwaningsih, “Penerapan Metode Support Vector Machine (SVM) Dalam Klasifikasi Kualitas Pengelasan Smaw (Shield Metal Arc Welding),” *Ilm. Edutic*, vol. 5, no. 1, pp. 17–25, 2018.
- [7] D. M. E. D. M. Hussein, “A survey on sentiment analysis challenges,” *J. King Saud Univ. - Eng. Sci.*, vol. 30, no. 4, pp. 330–338, 2018, doi: 10.1016/j.jksues.2016.04.002.
- [8] O. Arifin and T. B. Sasongko, “Analisa perbandingan tingkat performansi metode support vector machine dan naïve bayes classifier,” *Semin. Nas. Teknol. Inf. dan Multimed. 2018*, vol. 6, no. 1, pp. 67–72, 2018.
- [9] S. Wulan, U. Vitandy, A. A. Supianto, and F. A. Bachtiar, “Analisis Sentimen Evaluasi Kinerja Dosen menggunakan Term Frequency- Inverse Document Frequency dan Naïve Bayes Classifier,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 6, 2019.
- [10] Y. Cahyono, “ANALISIS SENTIMENT PADA SOSIAL MEDIA TWITTER MENGGUNAKAN NAÏVE BAYES CLASSIFIER DENGAN FEATURE SELECTION PARTICLE SWARM OPTIMIZATION DAN TERM FREQUENCY,” 2017.
- [11] L. D. Mahbubah *et al.*, “Analisa Sentimen Twitter Pada Pilpres 2019 Menggunakan,” *Sintak*, pp. 194–195, 2019, [Online]. Available: <https://www.unisbank.ac.id/ojs/index.php/sintak/article/view/7585>.
- [12] S. N. J. Fitriyyah, N. Safriadi, and E. E. Pratama, “Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes,” *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 3, p. 279, 2019, doi: 10.26418/jp.v5i3.34368.
- [13] M. F. Fibrianda and A. Bhawiyuga, “Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode Naïve Bayes Dan Support Vector Machine (SVM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 9, pp. 3112–3123, 2018.
- [14] F. S. Jumeilah, “Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 1, p. 19, 2017, doi: 10.29207/resti.v1i1.11.
- [15] N. P. Bunga, “ANALISIS SENTIMEN TERHADAP PENERAPAN TEKNOLOGI OLEH PEMERINTAH KOTA PEKANBARU BERDASARKAN INTERAKSI MASYARAKAT PADA MEDIA SOSIAL MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 21, no. 1, pp. 1–9, 2020, [Online]. Available: <http://mpoc.org.my/malaysian-palm-oil-industry/>.
- [16] D. S. Pamungkas, N. A. Setiyanto, and E. Dolphina, “Analisis Sentiment Pada Sosial Media Twitter Menggunakan Naive Bayes Classifier Terhadap Kata Kunci ‘Kurikulum 2013’,” vol. 14, no. 4, pp. 299–314, 2015.