



## PENERAPAN METODE NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE PADA ANALISIS SENTIMEN TERHADAP DAMPAK VIRUS CORONA DI TWITTER

Cholid Fadilah Hasri<sup>1</sup>, Debby Alita<sup>2</sup>

<sup>1,2</sup> Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia  
[cholidfadilahhasri@gmail.com](mailto:cholidfadilahhasri@gmail.com)<sup>1</sup>, [debbyalita@teknokrat.ac.id](mailto:debbyalita@teknokrat.ac.id)<sup>2</sup>

Received: 10 January 2022 Accepted: 23 March 2022 Published: 30 June 2022

### Abstract

The use of social media, especially twitter grows rapidly each day, and anyone could get any information quicker with the help of hashtag search or by visiting any trending topic by recent viral newswire. Government of Indonesia had already imposes PPKM (Pemberlakuan Pembatasan Kegiatan Masyarakat) or social and activity restrictions in various regions amongst Indonesia throughout 2021. The restrictions of gather and crowd is already in motion, by solely purposes reducing the mobility of people, to ensure the COVID-19 won't go outbreak. The news spread quickly on various social media platform, one of which is twitter. At least numerous people expressing their opinions regarding the policies made by the central and regional government and also the restrictions of social activity in 2021. By utilizing the Text Mining Classification Method Technique, it will be revealed whether the value of the sentiments are positive, neutral or negative. One of the most widely used algorithm in sentiment analysis is Naïve Bayes Classification method. This research utilize the Naïve Bayes Classifier (NBC). The results of the research shows that Naïve Bayes without any additional features capable of classify the sentiment with accuracy value of 81.07%. To ensure these research results, it is necessary to do a testing with other method, such as Support Vector Machine, with the results 79.96% accuracy value.

**Keywords:** Sentiment Analysis, Naive Bayes, Twitter, COVID-19

### Abstrak

Penggunaan media sosial, khususnya twitter terus berkembang pesat, dan siapa saja dapat memperoleh informasi dengan cepat dengan mencari hashtag atau mengunjungi topik trending dari berita yang sedang hangat. Pemerintahan Indonesia telah memberlakukan PPKM (Pemberlakuan Pembatasan Kegiatan Masyarakat) di sejumlah daerah Indonesia sepanjang tahun 2021. Larangan untuk berkumpul dan berkerumun selama PPKM telah diberlakukan untuk mengurangi mobilitas masyarakat agar mencegah terjadinya penyebaran COVID-19. Berita ini pun cepat menyebar di berbagai platform media sosial salah satunya adalah media sosial twitter. Setidaknya sejumlah masyarakat berbondong-bondong mengungkapkan pendapat mereka mengenai kebijakan-kebijakan yang telah dibuat oleh pemerintah pusat dan daerah tentang PPKM di tahun 2021. Dengan memanfaatkan teknik Text Mining metode klasifikasi, akan diketahui suatu sentimen bernilai positif, netral atau negatif. Salah satu algoritma yang banyak digunakan dalam analisis sentimen adalah metode klasifikasi Naïve Bayes. Penelitian ini menggunakan metode Naïve Bayes Classifier (NBC). Hasil penelitian menunjukkan bahwa metode Naïve Bayes tanpa penambahan fitur mampu mengklasifikasi sentimen dengan nilai akurasi sebesar 81.07%. Untuk memastikan hasil penelitian ini, dilakukan juga pengujian dengan Support Vector Machine yang menghasilkan nilai akurasi sebesar 79.96%.

**Kata Kunci:** Sentiment Analysis, Naive Bayes, Twitter, COVID-19.

### To cite this article:

Hasri and Alita. (2022). Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter. *Jurnal Informatika dan Rekayasa Perangkat Lunak*, Vol. 3, No. 2, 145-160

## PENDAHULUAN

Analisis sentimen adalah proses memahami dan mengolah data tekstual secara otomatis untuk mendapatkan informasi. Analisis sentimen dilakukan untuk mendeteksi opini terhadap suatu subjek dan objek (misalnya individu, organisasi ataupun produk) dalam sebuah kumpulan data. Besarnya pengaruh dan manfaat dari analisis sentiment menyebabkan penelitian ataupun aplikasi mengenai analisis sentimen berkembang pesat (Alita & Isnain, 2020). Juga untuk mengetahui siapa pemberi opini yang banyak mendapatkan respon dari pengguna twitter. Analisis sentiment dapat digunakan untuk mengungkap opini publik terhadap suatu isu (Kurniawan & Apriliani, 2020). Tidak hanya itu, analisis sentimen merupakan cara mengumpulkan pendapat khalayak umum menggunakan jejaring sosial yang didalamnya terdapat mengandung pelayanan umum, serta isu terkini (Syarifuddin, 2020). Contohnya seperti yang sedang ramai saat ini yaitu virus corona.

Virus corona atau dikenal sebagai pandemi Covid-19 pertama kali di deteksi di kota wuhan Desember 2019 (Wikipedia, 2019). Penyebaran virus secara cepat melalui kontak fisik seperti droplets yang dikeluarkan selama berbicara, batuk, atau bersin (Dave dkk, 2020). Pada awal tahun 2020 virus ini menyebar ke berbagai negara, termasuk Indonesia. Tidak hanya itu, pengaruh yang ditimbulkan akibat virus corona berdampak hampir disemua aspek, seperti kehidupan masyarakat (Burhanuddin & Abdi, 2020).

Kehidupan masyarakat berubah semenjak adanya wabah pandemi corona terlebih sejak dilakukannya *social distancing* yang mana masyarakat diharuskan untuk menjaga jarak kontak fisik antara satu orang dengan yang lainnya, berdiam diri di rumah, menghindari kerumunan, selalu mencuci tangan hingga diharuskan untuk selalu menggunakan masker (Yanuarita dan Haryati, 2020). Banyak masyarakat yang mengutarakan pendapatnya melalui media sosial twitter dalam menanggapi dampak yang ditimbulkan akibat wabah virus corona (Kaur & Sharma, 2020).

Media sosial Twitter adalah salah satu sebagai media penghubung yang diminati oleh seluruh masyarakat di dunia. Hal ini dapat dilihat dari jumlah peningkatan pengguna Twitter yang tercatat di seluruh dunia salah satunya indonesia (Darwis dkk, 2021). Twitter digunakan tidak hanya untuk bersosialisasi dan berkomunikasi namun juga digunakan untuk menyampaikan aspirasi dan mempresentasikan hal-hal yang sedang terjadi pada masyarakat (Astari dkk, 2020). Hal ini dapat dijadikan sebagai sebuah acuan untuk dilakukan analisis sentimen masyarakat terhadap dampak virus Corona. Analisis sentimen merupakan cara untuk melakukan proses klasifikasi terhadap dokumen-dokumen teks yang dilandasi dari opini dan pandangan dari masyarakat tertentu bisa dilakukan dengan proses memahami, mengolah hingga mengekstrak data bersifat dokumen teks secara otomatis (Hendriani & Sianturi, 2021). Hal ini dilakukan untuk mengetahui apakah opini maupun tanggapan tersebut bersifat positif, negatif atau netral. Salah satu metode yang digunakan dalam analisis sentimen yaitu metode Naïve Bayes Classifier (Ratino dkk, 2019).

Algoritma Naive Bayes adalah merupakan salah satu algoritma klasifikasi yang banyak digunakan pada Data Mining ataupun Text Mining (Wijaya dkk, 2016). Salah satu metode klasifikasi untuk menentukan gambaran persepsi masyarakat di dalam Text Mining adalah Naive Bayes Classifier (Isnain dkk, 2020). Metode ini dipilih dikarenakan sangat cocok untuk short data teks. Kelebihan dari NBC adalah metode sederhana tetapi memiliki akurasi serta performansi yang tinggi dalam pengklasifikasian teks (Cahyono, 2017).

Berdasarkan uraian diatas maka diperlukannya analisis sentimen untuk memisahkan mana yang merupakan opini berisi informasi yang bersifat positif, netral dan negatif terutama dalam aspek sosial.

Penelitian ini bertujuan menganalisis opini publik untuk menemukan klasifikasi positif, negatif dan netral terhadap kehidupan masyarakat baik itu sebelum dan sesudah wabah pandemi Corona. Adapun metode klasifikasi data yang digunakan oleh penulis yaitu Naïve Bayes Classifier dikarenakan metode ini memiliki tingkat akurasi, penarikan dan presisi yang lebih baik daripada metode SVM.

### 1.1 Text Mining

Text mining merupakan bagian dari data mining, yaitu proses untuk memperoleh suatu pengetahuan menggunakan seperangkat alat analisis dimana pengguna berinteraksi dengan sekumpulan dokumen dari waktu ke waktu [5]. Konsep text mining biasanya digunakan dalam klasifikasi dokumen tekstual dimana dokumen-dokumen tersebut akan diklasifikasikan sesuai dengan topik dokumen tersebut[6].

### 1.2 Sentimen Analisis

Sentimen analisis atau opinion mining dapat diartikan sebagai bidang ilmu yang meneliti bagaimana menyampaikan sentimen, opini atau pendapat dan emosi yang diungkapkan di dalam teks atau kalimat. Ada beberapa topik pembahasan dalam sentimen analisis, salah satu yang paling sering diteliti yaitu

klasifikasi sentimen. Topik ini berpusat pada kegiatan pengelompokan sentimen berlandaskan teks opini terhadap pembahasan masalah yang menarik [7].

### 1.3 Naïve Bayes Classifier(NBC)

Naïve Bayes merupakan sebuah metode pengklasifikasian dengan menggunakan probabilitas sederhana yang berakar pada Teorema Bayes dan memiliki asumsi ketidaktergantungan (independent) yang tinggi dari masing – masing kondisi atau kejadian, Bentuk umum teorema bayes adalah sebagai berikut:

$$P(H|X) = (1)$$

Keterangan:

X = Data dengan kelas yang belum diketahui

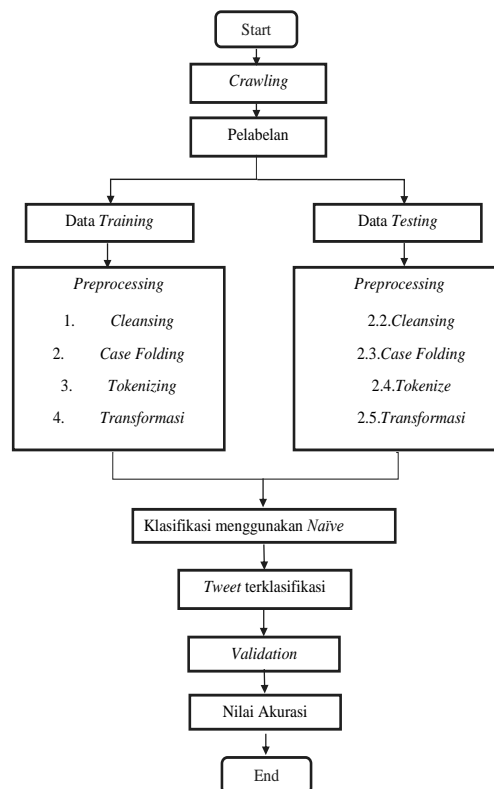
H = Hipotesa data X merupakan suatu kelas spesifik

$P(H|X)$  = Probabilitas hipotesis H berdasarkan kondisi X (posterior probability)

$P(H)$  = Probabilitas hipotesis H (prior probability)

## METODE PENELITIAN

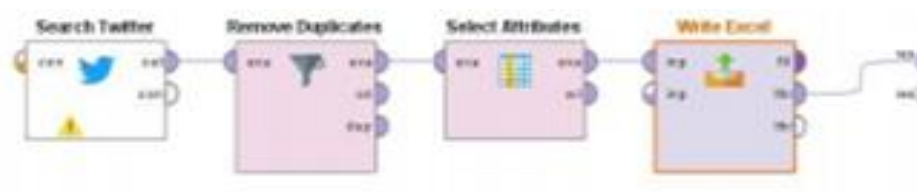
Data yang dikumpulkan bersumber dari komentar-komentar pada postingan berita di akun instagram. Kemudian dibagi menjadi tiga kelas yaitu kelas positif negatif dan netral. Berikut adalah tahapan penelitian :



Gambar 1. Tahapan Penelitian

### 2.1 Pengumpulan Data

Dalam pencarian data menggunakan sebuah kata kunci atau *query* ‘Kata Kunci’ atau ‘kata kunci lain’ dengan menggunakan aplikasi *Rapidminer*. Data yang didapat adalah seluruh data opini masyarakat tentang dampak corona terhadap kehidupan sosial masyarakat. Pengumpulan data dilakukan pada periode januari sampai dengan april 2021. Berikut merupakan gambar proses *crawling* data menggunakan *Rapidminer*.



Gambar 2. Proses Crawling Data Menggunakan Rapidminer

### 2.2 Proses Training dan Data Testing

Pada proses klasifikasi manual bertujuan untuk menentukan kelas dari kumpulan data set. Data set yang telah dilakukan klasifikasi manual akan dijadikan data latih (training). Dari klasifikasi manual akan menghasilkan tiga keluaran kategori, sentimen positif, negatif serta netral.

Tabel 1. Contoh Data Training

Tweet	Fitur	Kategori
Tweet1	@yoikidempal gara gara covid gajadi event ajg	Negatif
Tweet2	@lemmeflyhigh_ Asu. Malas sekali. Bakar2 ma tmn deket ga boleh. Emg covid anjg ya.	Negatif
Tweet3	@SatuReporter Tekan Penyebaran Covid Pemerintah Imbau Tidak Ada Perayaan Lomba Kemerdekaan	Positif
Tweet4	@inaymaa Untuk sementara mudiknya ditunda dulu yaa	Positif
Tweet5	@LtdAkbar Karena bagiku covid itu flu	Netral
Tweet6	@VIVAcoid Kampus-Kampus Gelar Wisuda Secara Online Saat Pandemi Covid	Netral

### 2.3 Pre-Proses

Pre-proses dilakukan untuk menghasilkan data bersih. Tahapan Preproses meliputi: Cleaning, Case Folding, Tokenize, Transformasi, Stemming dan Stopword Removal.

### 2.4 Pembentukan Model Probabilitas

Seluruh data latih (training) yang telah dikumpulkan akan diubah menjadi kumpulan komentar yang memiliki nilai probabilitas terhadap kategori kelas yang ditentukan.

Tabel 2. Hasil perhitungan *Prior Probability*

Atribut Kelas	P(Class)
Positif	3/9 atau 1/3
Negatif	3/9 atau 1/3
Netral	3/9 atau 1/3

Kemudian dilakukan perhitungan *conditional probability* masing-masing kelas seperti tabel 3 berikut:

**Tabel 3.** Perhitungan *Conditional Probability*

Trem	Kategori		
	Positif	Negatif	Netral
Acara	0/3	1/3	0/3
Ada	1/3	0/3	0/3
Anjing	0/3	2/3	0/3
Bakar	0/3	1/3	0/3
Boleh	0/3	1/3	0/3
Covid	1/3	2/3	1/3
Dekat	0/3	1/3	0/3
Dulu	1/3	0/3	0/3
Flu	0/3	0/3	1/3
gara-gara	0/3	1/3	0/3
Gelar	0/3	0/3	1/3
Imbau	1/3	0/3	0/3
Jadi	0/3	1/3	0/3
Kampus	0/3	0/3	1/3
Karena	0/3	0/3	1/3
Lomba	1/3	0/3	0/3
Malah	0/3	0/3	1/3
Mudik	1/3	0/3	0/3
Online	0/3	0/3	1/3
Pandemi	0/3	0/3	1/3
Pemerintah	1/3	0/3	0/3
Raya	1/3	0/3	0/3
Saat	0/3	0/3	1/3
Sebar	1/3	0/3	0/3
Secara	0/3	0/3	1/3
Sekali	0/3	1/3	0/3
Sementara	1/3	0/3	0/3
Tekan	1/3	0/3	0/3

**Tabel 4.** Perhitungan *Conditional Probability Lanjutan*

Trem	Kategori		
	Positif	Negatif	Netral
Teman	0/3	1/3	0/3
Tidak	1/3	2/3	0/3
Tunda	1/3	0/3	0/3
Wisuda	0/3	0/3	1/3

### 2.5 Klasifikasi

Pada tahapan ini akan dilakukan pengelompokan komentar kedalam beberapa kelas, yaitu: positif, negatif dan netral. Algoritma yang digunakan dipenelitian ini adalah *Naive Bayes Classifier*. Setelah diklasifikasi data tersebut akan diuji kualitas dari model yang telah dihasilkan dengan menghitung nilai akurasi.

### 2.6 Validasi

Pengujian dalam penelitian ini dengan menggunakan Confusion Matrix yang akan menghasilkan nilai akurasi.

### 2.7 Hasil

Hasil dari penerapan teknik data mining yang menggunakan metode *Naive Bayes Classifier* menghasilkan nilai akurasi dari proses analisis sentimen masyarakat terhadap kebijakan mudik 2021 dengan menggunakan *library python*. Yang menghasilkan dokumen terklasifikasi dengan tiga kategori kelas yaitu sentimen kelas positif, sentimen kelas negatif maupun sentimen kelas netral.

## HASIL DAN PEMBAHASAN

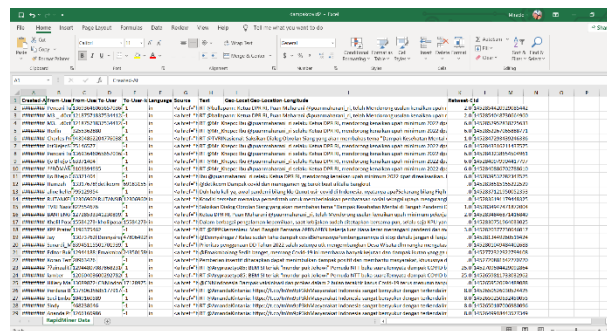
### 3.1 Crawling Data Instagram

Proses *crawling* dilakukan pada objek penelitian yaitu pada dampak virus corona terhadap masyarakat indonesia dengan pencarian sentimen diperoleh berdasarkan kata kunci “dampak virus corona”. Sistem *crawling* yang dibangun menggunakan *rapidminer*.



Gambar 3. Proses pengambilan data menggunakan *rapidminer*

Proses pengambilan data menggunakan *rapidminer* setelah semua *tweet* terkumpul, selanjutnya menyimpan hasil *crawling* data ke dalam file excel dengan format *xlsx*.



Gambar 4. Hasil Crawling dalam bentuk *Excel*

### 3.2 Klasifikasi Sentimen Manual (Tagging)

Setelah seluruh *tweet* berhasil disimpan selanjutnya dilakukan proses pelabelan. Dalam penelitian ini proses klasifikasi manual (tagging) dilakukan oleh seorang yang ahli dalam bahasa Indonesia yang latar belakang pendidikan bahasa dan sastra Indonesia.

Nama : Meri Eka Oktaviani., S.Pd  
Pendidikan Terahir : Sarjana Pendidikan

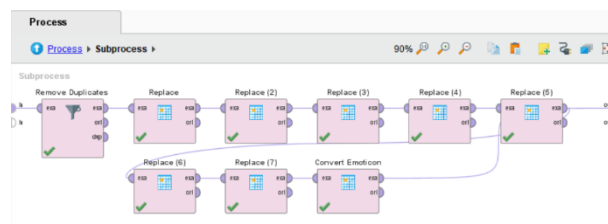
yang nantinya data tersebut akan digunakan sebagai data latih. Pengklasifikasian data yang digunakan sebanyak 1104 *tweet* ke dalam tiga kelas yaitu positif, negatif dan netral.

Nama : Meri Eka Oktayiani, S.Pd Pendidikan Terakhir : Sarjana Pendidikan	
Text	Label
Ibu @puanmaharani_ri selaku Ketua DPR RI, mendorong kenaikan upah minimum 2022 dapat direalisasikan. Ini tak terlepas dari dampak pandemi COVID-19. Serta selama thn 2021 blm ada kenaikan upah minimum seiring menurunnya perekonomian. #DPRruntukNegeri Terus mendengar aspirasi rakyat https://t.co/GATBxwvwaq	Positif
@detikcom Dampak covid dan manajemen ya buruk buat alitalia bangkrut Duh hahh kali ya, awal pandemi bilang klo Qurut usir covid di Indonesia nyatanya apa? Sekarang bilang Figih bantu atasi dampak covid. Masih percaya sama orang ini? @Jelantik5 @agama_musantara @AremaNita_ https://t.co/2641cVMutD	Negatif
Kondisi tersebut memaksa pemerintah untuk membatalkan pembatasan sosial sebagai upaya mengurangi dampak penyebaran Covid-19 sehingga menyebabkan menurunnya intensitas kegiatan ekonomi yang menyebabkan terjadinya economic searbacks. https://t.co/evSZKoaPGM	Netral
Ketua DPR RI, Puan Maharani @puanmaharani_ri, telah Mendorong usulan kenaikan upah minimum pekerja pada 2022 untuk meningkatkan kesejahteraan mereka yang terkena dampak pandemi COVID-19. #DPRruntukNegeri Terus mendengar aspirasi rakyat. https://t.co/YZVyaTMsb5	Positif
Dalam berbagai pengalaman kepemimpinan, saat kebijakan sudah ditetapkan bersama pun, selalu saja KPU yang menyalah dampak dari ketidakpastian. Bukan DPR & Pemerintah. Misalnya soal pilkada di tengah pandemi kemarin, yang terpacar covid itu kan penyelenggara. Bukan yang di atas.	Negatif
@Dennysiregar7 Kalau sudah tahu dampak covid Seharusnya Pembangunannya di stop dahulu jangan di lanjutkan, uangnya untuk penanganan covid	Positif
@Emakmbolang Sedih banget, memang Covid-19 ini membawa banyak kejutan dan dampak ilutan yang ga mudah.	Netral

Gambar 4. Hasil Crawling data yang sudah di tagging

### 3.3 Preprocessing

Tahapan ini terdiri dari beberapa proses karena data *tweet* tidak sepenuhnya menggunakan kata baku. Tahap preprocessing dilakukan dengan menggunakan bantuan aplikasi *rapidminer*.

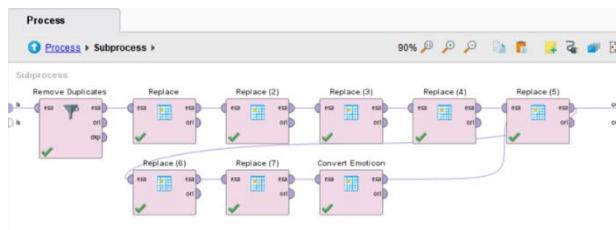


Gambar 5. Proses Penghapusan tanda baca atau angka

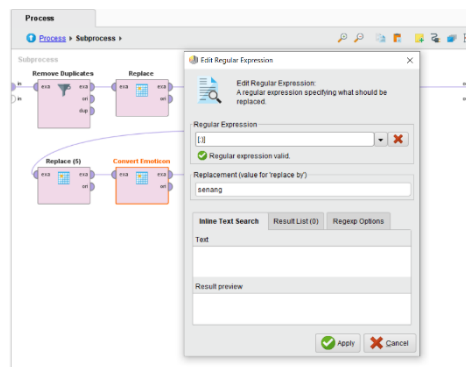
Dalam proses *cleaning* tersebut menggunakan *remove duplicate* sebagai operator yang digunakan untuk menghapus *duplicate* pada *attribute*. Dan operator *replace* untuk mengubah *symbol* dan angka yang tidak diperlukan, berikut hasil *cleaning* menggunakan *rapidminer*.

Row No.	Sesment	Text
1	Positif	Salipaini Ketua DPR RI Puan Maharani puamabarani telah mendorong usulan kenaikan upah minimum pekerja pada 2022 untuk mengatasi...
2	Positif	Salipaini puamabarani selaku Ketua DPR RI mendorong kenaikan upah minimum 2022 yang dilaksanakan ini tak terlepas dr damp...
3	Netral	TiGHasasari Sekretaris Ditlog Otoritas Ssang yang akan membahas loma Dampak Kesehatan Mental di Tengah Pandemi Covid 19 diamanah oleh...
4	Positif	Ibu puamabarani selaku Ketua DPR RI mendorong kenaikan upah minimum 2022 dapat dilaksanakan ini tak terlepas dr dampak pandemi COVID-19 selama ini...
5	Negatif	edificom Dampak covid dan manajemen yg benar buat atalala bangndi
6	Negatif	Duh hulu tali ya awal pandemi bilang lo Gaud user covid di Indonesia nyatanya asa
7	Netral	Kondisi tersebut memaksa pemerintah untuk membatalkan pembatasan sosial sebagai upaya mengurangi dampak penyebaran Covid19 sehingga masyarakat lainnya intensitas kegiatan...
8	Netral	Sekaligus Ditlog Otoritas Ssang yang akan membahas loma Dampak Kesehatan Mental di Tengah Pandemi Covid 19 diamanah oleh Dit Ssang
9	Positif	Ketua DPR RI Puan Maharani puamabarani telah mendorong usulan kenaikan upah minimum pekerja pada 2022 untuk meningkatkan kesejahteraan mereka yang terkena dampak pandemi...
10	Negatif	Dalam berbagai pengalaman kepastian saat kelijakan sudah dilakukan bersama pun selalu saja KPU yang menafikan dampak dari implementasinya Bukan DPR Pemerintah
11	Positif	SPPKormentaku Masi Tanggih Bersama APBN
12	Positif	Demosepati? Kalau sudah tahu dampak covid
13	Positif	Prioritas pengurusan CO Tahun 2022 salah satunya akan mengembangkan Desa Wisata dan rangka mengatasi dampak Pandemi Covid19
14	Netral	Emasibotang Ssang bangat memang Covid19 ini membawa banyak kejutan dan dampak kulan yang ga mudah
15	Positif	Pemberian insentif dibagikan dapat meredakan dampak positif dan membawa masyarakat khususnya dunia usaha Hibatobuh.DimHisa KaramTempo
16	Negatif	Appanyelid BGM Grlanah mudur pak Jonwil Pemuda NTT Ssang sanga
17	Netral	Chibabibenera Campa vakanisasi dan proses dilansi2 Sutan berakir kasus Covid19 terus menurut tanpa dibatarkan wajo DINDPDR terbahadua penunggang pesawat terbang kenaga...

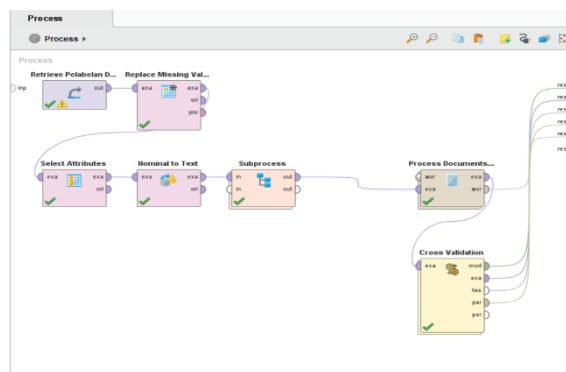
Gambar 6. Hasil *Cleaning* data



Gambar 7. Tahapan *Convert Emoticon*

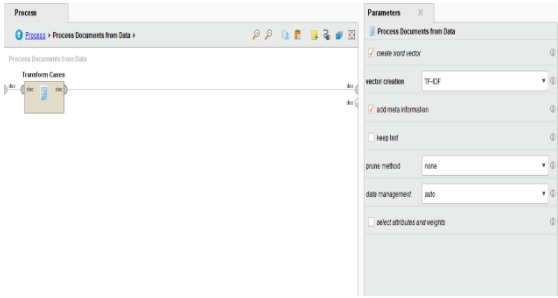


Gambar 8. Seleksi *Convert Emoticon*

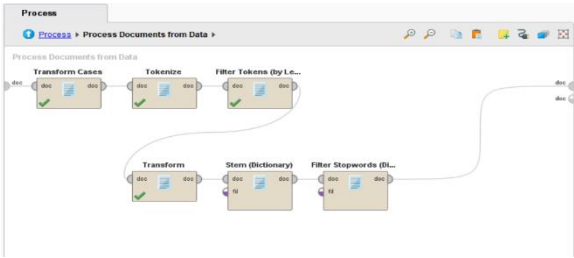


Gambar 9. Tahapan *Case Folding*

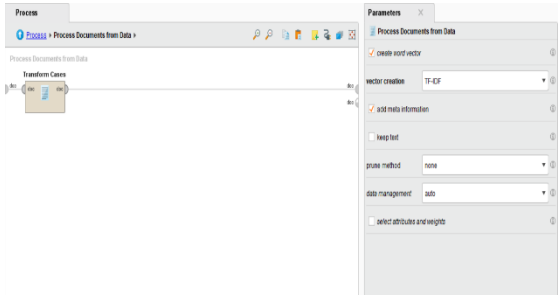




Gambar 10. Uji Case Folding



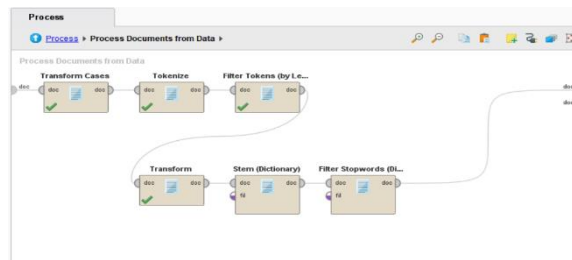
Gambar 11. Hasil Uji Case Folding



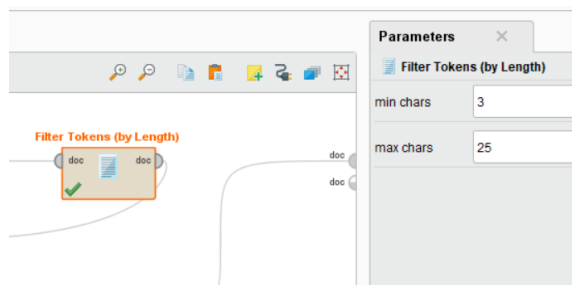
Gambar 10. Uji Case Folding

Word	Atribut Nama	Total Co.	Docum...	Positif	Netral	Neg
15	15	1	1	0	1	0
16	16	1	1	1	0	0
17	17	1	1	0	0	1
18	18	1	1	0	0	1
19	19	1	1	0	0	1
20	20	1	1	0	0	1
21	21	1	1	0	0	1
22	22	1	1	0	0	1
23	23	1	1	0	0	1
24	24	1	1	0	0	1
25	25	1	1	0	0	1
26	26	1	1	0	0	1
27	27	1	1	0	0	1
28	28	1	1	0	0	1
29	29	1	1	0	0	1
30	30	1	1	0	0	1
31	31	1	1	0	0	1
32	32	1	1	0	0	1
33	33	1	1	0	0	1
34	34	1	1	0	0	1
35	35	1	1	0	0	1
36	36	1	1	0	0	1
37	37	1	1	0	0	1
38	38	1	1	0	0	1
39	39	1	1	0	0	1
40	40	1	1	0	0	1
41	41	1	1	0	0	1
42	42	1	1	0	0	1
43	43	1	1	0	0	1
44	44	1	1	0	0	1
45	45	1	1	0	0	1
46	46	1	1	0	0	1
47	47	1	1	0	0	1
48	48	1	1	0	0	1
49	49	1	1	0	0	1
50	50	1	1	0	0	1

Gambar 11. Hasil Uji Case Folding



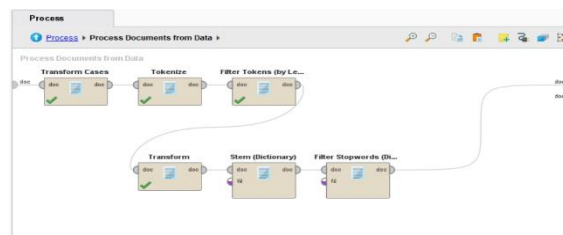
Gambar 12. Tahapan Uji *Tokenize*



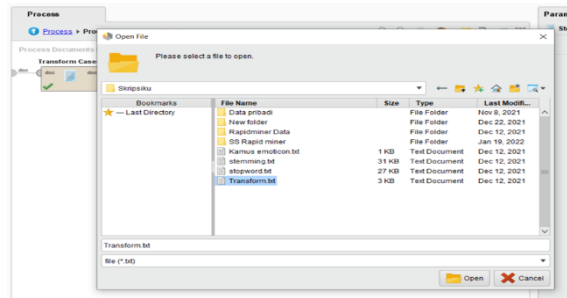
Gambar 13. Pergantian Parameter *Tokenize*

Word	Attribute Name	Total ...	Docum...	Positif	Netral	Negatif
about	about	4	4	0	4	0
abselyono	abselyono	12	12	10	2	0
acara	acara	3	3	0	3	0
access	access	1	1	0	1	0
aceh	aceh	2	2	1	1	0
acprtae	acprtae	1	1	1	0	0
actor	actor	1	1	0	0	1
ada	ada	134	129	53	43	38
adadikompas	adadikompas	2	2	2	0	0
adalah	adalah	21	21	3	14	4
adanya	adanya	14	14	8	4	2
adaptasi	adaptasi	2	2	2	0	0
adat	adat	1	1	1	0	0
adpw	adpw	1	1	1	0	0
adik	adik	1	1	0	0	1
adikarto	adikarto	1	1	0	1	0

Gambar 14. Hasil Uji *Tokenize*



Gambar 15. Tahapan Uji *Transformasi*



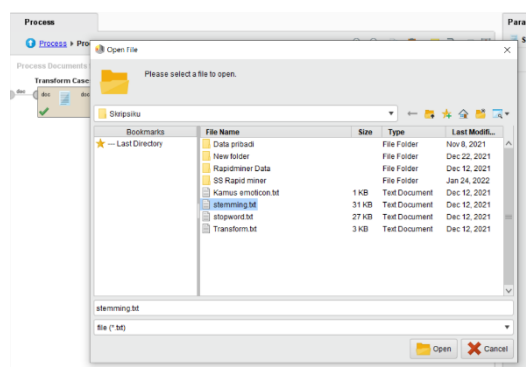
Gambar 16. Uji Transformasi

Word	Attribute Name	Total Occurrences	Document Oc...	Positif	Netral	Negatif
akibat	akibat	31	30	15	5	11
akibatnya	akibatnya	1	1	0	0	1
akselerasi	akselerasi	3	3	3	0	0
akses	akses	2	2	2	0	0
aiat	aiat	66	66	65	1	0
akun	akun	3	1	0	0	3
akurasi	akurasi	4	4	4	0	0
alami	alami	1	1	1	0	0
alamin	alamin	1	1	0	0	1
alasan	alasan	4	3	0	0	4
albertsolo	albertsolo	1	1	0	1	0
alhasil	alhasil	1	1	0	0	1
alias	alias	1	1	0	1	0
alihkan	alihkan	1	1	0	0	1
aliran	aliran	1	1	1	0	0

Gambar 17. Hasil Uji Transformasi



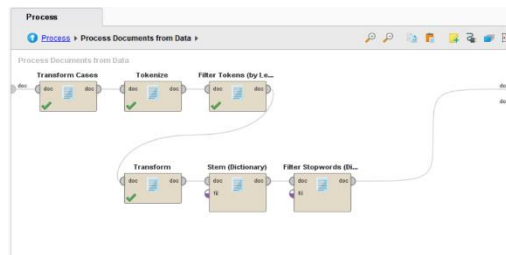
Gambar 18. Tahapan Uji Stemming



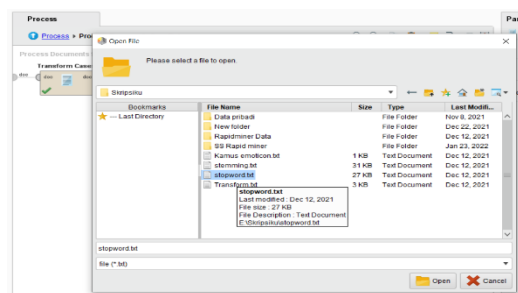
Gambar 19. Uji Stemming

Word	Attribute Name	Total Occurrences	Document Occu...	Positif	Netral	Negatif
abai	abai	1	1	0	0	1
abis	abis	3	2	0	1	2
about	about	4	4	0	4	0
absetyono	absetyono	12	12	10	2	0
acara	acara	3	3	0	3	0
access	access	1	1	0	1	0
aceh	aceh	2	2	1	1	0
acptae	acptae	1	1	1	0	0
actor	actor	1	1	0	0	1
ada	ada	170	160	54	62	44
adadkompas	adadkompas	2	2	2	0	0
adaptasi	adaptasi	2	2	2	0	0
adat	adat	1	1	1	0	0

Gambar 20. Hasil Uji Stemming



Gambar 21. Tahapan Uji Stopword Removal



Gambar 23. Uji Stopword Removal

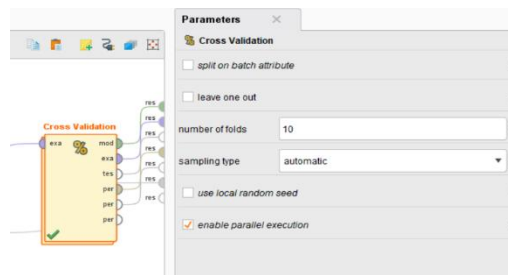
Word	Attribute Name	Total Occure...	Document Oc...	Positif	Netral	Negatif
aku	aku	10	6	1	1	8
akun	akun	3	1	0	0	3
akurasi	akurasi	4	4	4	0	0
alam	alam	1	1	1	0	0
alamin	alamin	1	1	0	0	1
alasan	alasan	4	3	0	0	4
albertsola	albertsola	1	1	0	1	0
alhasil	alhasil	1	1	0	0	1
alias	alias	1	1	0	1	0
alihan	alihan	1	1	0	0	1
aliran	aliran	1	1	1	0	0

Gambar 24. Hasil Uji Stopword Removal

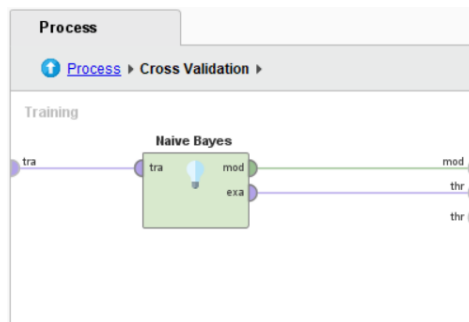
### 3.4 Cross Validation Menggunakan Naïve Bayes Classifier

Penelitian ini menggunakan Operator *Cross Validation* yang berguna untuk menguji atau mengevaluasi data training dengan baik. Selanjutnya pada *operator cross validation* akan dilakukan parameter yang digunakan pada *operator cross validation* adalah *number of folds*. *Number of folds* sendiri digunakan untuk melakukan pengujian jumlah data validasi yang ingin diujikan sebanyak 1-10 kali. Hal ini dilakukan untuk

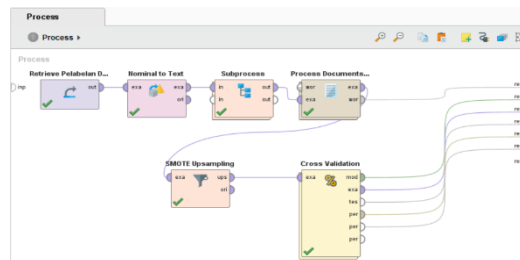
mengetahui nilai *performance* dan akurasi dari algoritma klasifikasi yang digunakan. Dapat dilihat pada gambar dibawah ini :



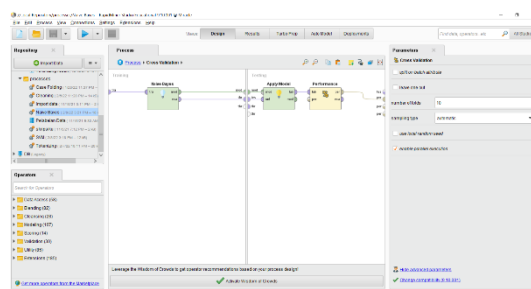
Gambar 25. Operator Cross Validation dan Parameter



Gambar 26. Proses Naive Bayes Classifier



Gambar 27. Proses utama klasifikasi



Gambar 28. Proses Cross Validation Naive Bayes Clasiffier

Table View Plot View

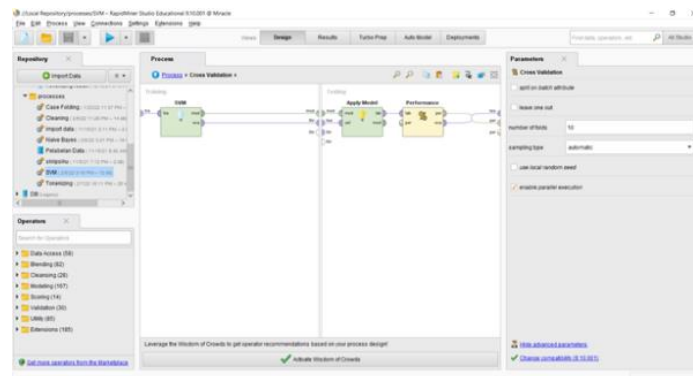
accuracy: 81.67% +/- 3.87% (micro average: 81.67%)

	true Positif	true Netral	true Negatif	class precision
pred. Positif	429	44	2	90.30%
pred. Netral	129	231	1	64.52%
pred. Negatif	53	69	584	83.25%
class recall	76.51%	67.10%	88.51%	

Gambar 29. Hasil Uji Cross Validation Naïve Bayes Clasiffier

### 3.5 Pengujian Cross Validation Menggunakan Support Vector Machine

Metode *Support Vector Machine* pada penelitian ini digunakan untuk membandingkan akurasi terbaik dari kedua algoritma tersebut. Pengujian metode *Support Vector Machine* dilakukan didalam operator *cross validation*, dan proses dapat dilihat pada gambar berikut :



Gambar 30. Proses Cross Validation Support Vector Machine

Setelah tahap *Cross Validation* maka akan didapat nilai akurasi, *recall*, *Precision* pada proses pengujian *Support Vector Machine* menggunakan *rapidminer*. Berikut hasil proses dalam klasifikasi, yang dapat dilihat pada gambar berikut

Table View Plot View

accuracy: 79.96% +/- 2.65% (micro average: 79.96%)

	true Positif	true Netral	true Negatif	class precision
pred. Positif	451	108	48	74.30%
pred. Netral	57	260	9	79.75%
pred. Negatif	23	52	474	86.34%
class recall	84.93%	61.90%	89.27%	

Gambar 31. Hasil Uji Cross Validation Support Vector Machine

### 3.6 Interpretasi Hasil

Berdasarkan hasil klasifikasi menggunakan kedua algoritma di atas, diperoleh bahwa nilai akurasi algoritma *Naïve Bayes Classifier* lebih akurat dibandingkan dengan *Support Vector Machine*. Dapat dikatakan penggunaan algoritma *Naïve Bayes Classifier* lebih baik daripada *Support Vector Machine* untuk jumlah data yang terbatas. Berikut akan dijabarkan masing-masing hasil klasifikasi kedua algoritma tersebut :

**Tabel 4.** Perbandingan *NBC* dan *SVM*

No.	<i>Naïve Bayes Classifier</i>	<i>Support Vector Machine</i>
1	Dalam proses klasifikasi diperlukan pembagian data menjadi dua bagian, yaitu data training dan data testing	Dalam proses klasifikasi diperlukan pembagian data menjadi dua bagian, yaitu data training dan data testing
2	Text preprocessing dengan beberapa metode yaitu, tokenizing, filtering, dan case folding.	Text preprocessing dengan beberapa metode yaitu, tokenizing, filtering, dan case folding.
3	Hasil akhir menyatakan bahwa penggunaan algoritma <i>naïve bayes</i> dalam pengklasifikasian data twitter pada penelitian ini mencapai tingkat akurasi 81.07%.	Hasil akhir menyatakan bahwa penggunaan algoritma <i>Support Vector Machine</i> dalam pengklasifikasian data twitter pada penelitian ini mencapai tingkat akurasi 79.96%.
4	Menghasilkan nilai Accuracy sebesar 81.07%.	Menghasilkan nilai Accuracy sebesar 79.96%.

### 3.7 Analisis Error Cross Validation

Analisis ini dilakukan untuk mengetahui prediksi yang salah atau tidak sesuai dengan data yang diujikan sebagai berikut:

1. Analisis error yang pertama yaitu diketahui banyaknya prediksi yang salah pada hasil pengujian yaitu dimana kategori sebagai label positif padahal kata tersebut sebenarnya terdapat pada label netral maupun pada label negatif pada saat diprediksi oleh Naive Bayes Classifier. Contoh kata alasan, kekerasan, dll)
2. Analisis error yang kedua yaitu diketahui banyaknya prediksi yang salah pada hasil pengujian yaitu dimana kategori sebagai label netral padahal kata tersebut sebenarnya terdapat pada label positif maupun pada label negatif pada saat diprediksi oleh Naive Bayes Classifier. 57 Contoh kata : bebas, menolak, dll)

Analisis error yang pertama yaitu diketahui banyaknya prediksi yang salah pada hasil pengujian dimana kategori sebagai label negatif padahal kata tersebut sebenarnya terdapat pada label positif maupun pada label netral pada saat diprediksi oleh Naive Bayes Classifier. Contoh kata : bakal, dampak, dll).

## SIMPULAN

Berdasarkan hasil pembahasan yang telah dilakukan, maka dapat disimpulkan sebagai berikut:

1. COVID-19 merupakan penyakit yang tidak sedikit meresahkan masyarakat umum, apalagi ditambah dengan kicauan di media sosial tentang COVID-19 yang semakin merabak, salah satunya adalah *twitter* yang menambah masyarakat ketakutan.
2. Pada penelitian ini, data yang digunakan sebanyak 1104 data yang di ambil dari *tweet* di twitter dengan menggunakan 3 sentimen yaitu positif, negatif, dan netral.
3. Pada pengujian menggunakan aplikasi *rapidminer* dengan algoritma *Naive Bayes Classifier* menghasilkan nilai akurasi data sebesar 81.07%.
4. Untuk membandingkan hasil nilai akurasi penelitian ini, dilakukan juga pengujian dengan algoritma lain yaitu *Support Vector Machine* yang menghasilkan nilai akurasi data sebesar 79,96%

## REFERENSI/DAFTAR PUSTAKA

- [1] Widodo dan Massus. S., 2006. Requirements Management Pada Extreme Programing. Jakarta: Seminar Nasional Aplikasi Teknologi informasi.
- [2] Burhanuddin dan Abdul S., 2020. Analisis Pengaruh Pandemi Covid-19 Pada Sentimen Pariwisata dan Analisis Jaringan Semantik Pada Media Sosial, *Vol.1*, Hal. 1-6.

- [3] Azuma, M. (2008) 'Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) – Software and quality in use models'.
- [4] Sugiyono (2013) *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- [5] Anisah, N., Anton, & Radiyah, U. (2016). Rancangan Sistem Informasi E-Recuitment Berbasis Web pada PT. Geoservices. *PROSISKO, Vol.3*, Hal. 1-7.
- [6] Desanti, R. I., Supri, C. F., & Widjaja, A. (2017). Aplikasi Perekrutan dan Penilaian Karyawan Berbasis Web pada PT. XYZ. *ULTIMA InfoSys, Vol. 8*, Hal. 1-8.
- [7] Mashudi. (2014). Pemodelan Sistem E-Recruitment Karyawan Baru Berbasis Web Studi Kasus PT. XYZ. *LENTERA ICT*, Hal. 1-7.
- [8] Permana, D. A., & Dewantara, R. Y. (2018). Analisis dan Rancangan Sistem Informasi Perekrutan Karyawan Berbasis Web (Studi pada PT Sumber Abadi Bersama, Gondanglegi, Kabupaten Malang). *Jurnal Administrasi Bisnis, Vol. 56*, Hal. 1-9.
- [9] Syastra, T. M., Herman, & Adam, S. (2017). Rancang Bangun Sistem E-Recuitment Berbasis Android. *IPTEKS TERAPAN*, Hal. 1-12.
- [10] Destiningrum, M., & Adrian, Q. J. (2017). Sistem Informasi Penjadwalan Dokter Berbasis Web Dengan Menggunakan Framework Codeigniter (Studi Kasus : Rumah Sakit Yukum Medical Centre). *Teknoinfo, Vol. 11*, Hal. 1 - 8.
- [11] Wati, H. D., Rahmanto, Y., & Fernando, Y. (2019). Rancang Bangun Sistem Informasi Manajemen Kegiatan Esktrakurikuler Berbasis Web (Studi Kasus : SMK Ma'arif Kalirejo Lampung Tengah). *Teknokompak, Vol. 13*, Hal. 1 - 5.
- [12] Asri, S. A., Bawa, I. B., & Suhendar, F. R. (2016). Sistem Informasi Pengelolaan Kegiatan UKM Berbasis Web di Lingkungan Politeknik Negeri Bali. *Jurnal Logic, Vol. 16*, Hal. 1 - 6.
- [13] Fatoni, A., & Dwi, D. (2016). Rancang Bangun Sistem Extreme Programming Sebagai Metodologi Pengembangan Sistem. *Jurnal Prosisko, Vol. 3*, Hal. 1 - 4.
- [14] Pressman, R. (2012). *Rekayasa Perangkat Lunak*. Yogyakarta: Penerbit Andi.
- [15] Raharjo, B. (2015). *Belajar Otodidak Framework Codeigniter*. Bandung: Penerbit Informatika.
- [16] Rossa, A. S., & Salahuddin, M. (2013). *Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek*. Bandung: Penerbit Informatika.
- [17] Witriani, N. S., & Janah, N. Z. (2017). Aplikasi Pengelolaan Data Organisasi Mahasiswa. *Industrial Research Workshop and National Seminar, Vol. 8*, Hal. 1 - 9.