

Analisis Sentimen Terhadap Kebijakan Pemerintah Terkait Pandemi Covid-19 Pada Twitter

Medi Taruk*¹, Anindita Septiarini², Fandi Alief Al Akbar³

^{1,2,3}Program Studi Informatika, Fakultas Teknik, Universitas Mulawarman
e-mail: *¹meditaruk@ft.unmul.ac.id, ²anindita@unmul.ac.id, ³fandiakbar@gmail.com

Abstrak

Twitter adalah salah satu media pertukaran informasi yang mudah dan populer. Twitter merupakan media sosial yang banyak digunakan di Indonesia dan memiliki persebaran serta distribusi informasi yang sangat cepat. Covid-19 menjadi topik yang hangat pada awal 2020. Virus yang bermula dari Wuhan China ini telah menyebar secara cepat ke hampir seluruh dunia. Sejak adanya kasus pertama dengan dua orang positif di Indonesia, topik covid-19 ini selalu dibahas dalam berbagai media berita, dan tentu saja media sosial. Pada penelitian ini, tweet yang mengandung opini masyarakat terhadap kebijakan pemerintah tentang covid-19 diklasifikasikan menggunakan metode klasifikasi Naïve Bayes. Tweet yang berisi kebijakan pemerintah tentang covid-19 diambil dengan menggunakan webscraping. Namun, data yang telah dikumpulkan masih belum dapat digunakan untuk klasifikasi karena terdapat kata-kata yang tidak baku dan terdapat banyak noise di dalam data tersebut sehingga perlu dilakukan preprocessing. Tahapan preprocessing dilakukan untuk menghilangkan hal yang tidak dibutuhkan (url, mention), tokenization, stopword removal, dan stemming. Penelitian ini telah menghasilkan sistem klasifikasi sentiment analisis masyarakat terhadap kebijakan pemerintah tentang covid-19 dengan metode Naïve Bayes. Berdasarkan hasil penelitian tingkat akurasi dari metode Naïve Bayes menggunakan perbandingan data pelatihan sebesar 70 dan data testing sebesar 30 mendapatkan Precision sebesar 38%, Recall sebesar 42%, F-Measure sebesar 40% dan tingkat akurasi sebesar 76%.

Kata kunci— Analisis Sentimen, Twitter, Naive Bayes, Klasifikasi

1. PENDAHULUAN

Twitter merupakan media sosial yang banyak digunakan di Indonesia dan memiliki persebaran serta distribusi informasi yang sangat cepat. Pengguna Twitter yang berasal dari berbagai kalangan dan disiplin ilmu menyebabkan aliran komunikasi dan opini di media ini menjadi lebih beragam, salah satunya adalah kritikan dan komentar mengenai kebijakan pemerintah.

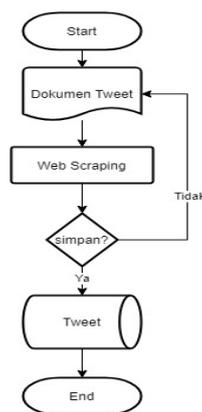
Covid-19 menjadi topik yang hangat pada awal 2020. Virus yang bermula dari Wuhan China ini telah menyebar secara cepat ke hampir seluruh dunia. Sejak adanya kasus pertama dengan dua orang positif di Indonesia, topik covid-19 ini selalu dibahas dalam berbagai media berita, dan tentu saja media social.

Pada penelitian ini, tweet yang mengandung opini masyarakat terhadap kebijakan pemerintah tentang covid-19 diklasifikasikan menggunakan metode klasifikasi Naïve Bayes. Tweet yang berisi kebijakan pemerintah tentang covid-19 diambil dengan menggunakan webscraping. Namun, data yang telah dikumpulkan masih belum dapat digunakan untuk

klasifikasi karena terdapat kata-kata yang tidak baku dan terdapat banyak noise di dalam data tersebut sehingga perlu dilakukan preprocessing. Tahapan preprocessing dilakukan untuk menghilangkan hal yang tidak dibutuhkan (*url*, *mention*), *tokenization*, *stopword removal*, dan *stemming*.

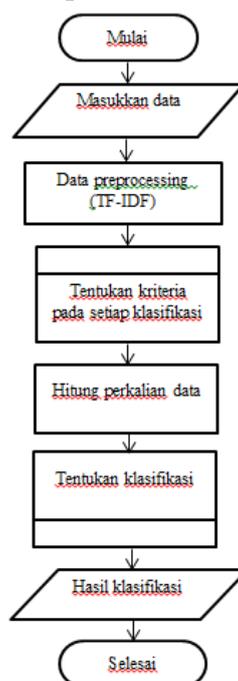
2. METODE PENELITIAN

Data yang digunakan dalam penelitian ini diambil dari sosial media Twitter dalam rentan waktu april s.d juni 2020. Pengambilan data menggunakan *Web Scrapping* dengan kata kunci “*covid-19*”, “*kebijakan pemerintah*”. Berikut ini merupakan tahapan dalam pengumpulan data pada penelitian ini.



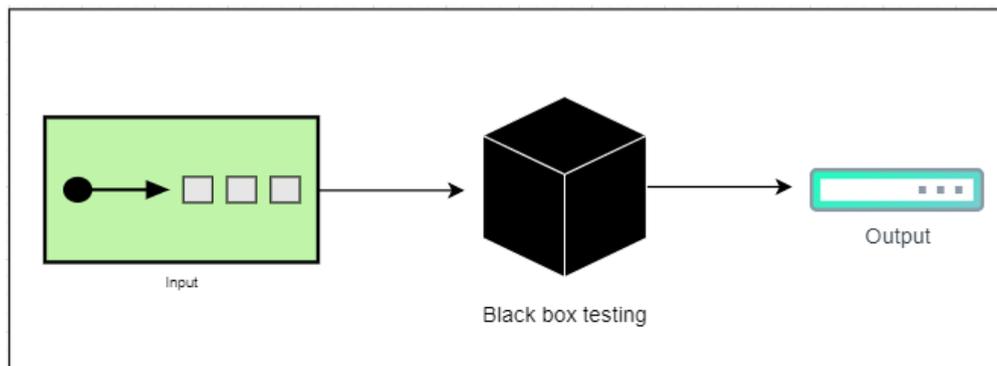
Gambar 1. Flowchart Pengumpulan Data

Data preprocessing akan memproses data sehingga memiliki bobot dari setiap kata yang dihasilkan dan algoritma *naive bayes* memiliki fungsi perbandingan dari data yang digunakan dengan klasifikasi yang ada, nilai terbesar dari pada setiap klasifikasi, maka data tersebut telah dinyatakan ke dalam klasifikasi tertentu berdasarkan nilai terbesar. Berikut adalah alur kerja dari penerapan algoritma yang ada dapat dilihat pada Gambar 2.



Gambar 2. Alur Kerja Penerapan Algoritma

Pengujian dalam penelitian ini menggunakan *black box testing* dimana fokus pada apa yang dihasilkan oleh sistem sesuai atau tidak dengan perancangan awal. Berikut adalah alur atau perancangan dari *black box testing* dapat dilihat pada Gambar 3.



Gambar 3. Alur Perancangan *Black Box Testing*

Skenario pengujian yang dilakukan sudah ditentukan berdasarkan rancangan-rancangan yang ditetapkan. Diharapkan dengan adanya skenario ini, sistem ini dapat dieksekusi dan memberikan hasil-hasil yang sesuai dengan rancangan, namun jika sistem mengeluarkan hasil yang tidak sesuai rancangan, maka skenario pengujian ini pun tergolong berhasil karena mampu memberikan hasil di luar skenario dan dapat dilakukan evaluasi untuk pengembangan dan perbaikan sistem selanjutnya.

Evaluasi bertujuan untuk menguji model klasifikasi data mining untuk mengetahui kinerja sistem. Metode yang digunakan adalah *confusion matrix* dengan menghitung *precision*, *recall*, *accuracy* dan *f1-measure*. Ada 100 data yang terbagi menjadi 3 sentimen yaitu positif, netral, dan negatif. Data yang telah dinormalisasi sebelum dimasukkan ke dalam mesin klasifikasi, data dibagi menjadi dua yaitu data pelatihan sebanyak 70 dan data pengujian sebanyak 30 menggunakan metode *confusion matrix*.

Tabel 1. Pengujian

Actual class	Predicted class		
	Positif	Netral	Negatif
Positif	0	1	4
Netral	0	1	1
Negatif	0	1	22

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini menggunakan 100 tweets. Data yang digunakan sebagai data input berasal dari data mentah berupa data tweet yang dikumpulkan menggunakan *web scraping*. Data yang digunakan untuk proses pembentukan model prediksi ini adalah data 100 tweet. Kemudian data tersebut dibagi menjadi dua bagian yaitu *Training Set* dan *Testing Set*.

Training Set merupakan data yang digunakan dalam proses pelatihan sistem dan terdiri dari pasangan data input dan data target. Dari keseluruhan data yang didapatkan tersebut, 70% bagian akan digunakan sebagai *data training*. *Testing Set* merupakan data yang digunakan untuk menguji kemampuan sistem dan juga terdiri dari pasangan data input dan data target. *Data testing* yang digunakan berjumlah 30% dari data yang diperoleh.

3.1 Case Folding

Tidak semua dokumen teks konsisten dalam menggunakan huruf kapital. Oleh karena itu

peran *case folding* dibutuhkan untuk mengkonversikan keseluruhan teks dalam dokumen menjadi suatu bentuk standar (huruf kecil atau lowercase). Sebagai contoh, user yang ingin mendapatkan informasi tentang “SELEBGRAM” dan mengetik “Selebgram”, “SELEBGRAM”, atau “selebgram” tetap diberikan hasil retrieval yang sama yakni “selebgram”. *Case folding* adalah mengubah semua huruf dokumen menjadi huruf kecil. Hanya huruf ‘a’ sampai dengan ‘z’ yang diterima. Karakter selain huruf akan dihilangkan.

3.2 Tokenizing

Tahap tokenizing digunakan untuk memisahkan kalimat yang ada dalam string menjadi potongan kata tunggal. Untuk lebih jelasnya, dapat dilihat contoh penerapan tokenizing pada 3 caption yang berbeda. Contoh dari tahap tokenizing dapat dilihat pada Tabel 2.

Tabel 2. Tokenizing

Teks Input	Teks Output
korona bikin orang kesel menghambat jalan	korona bikin orang kesel menghambat jalan
hari ini dikantor pos dengan keadaan yang riuh pencairan dana bagi warga yang mendapat bantuan	hari ini dikantor pos dengan keadaan yang riuh pencairan dana bagi warga yang mendapat bantuan
mentan saya lagi fokus ke kayu putih buat bikin sovenir kalung anti korona untuk warga	mentan saya lagi fokus ke kayu putih buat bikin sovenir kalung anti korona untuk warga

3.3 Stopword

Pada tahap ini dilakukan pembuangan kata-kata yang kurang penting atau kata-kata yang sering muncul (*Stopword*), seperti kata penghubung dan kata keterangan yang bukan merupakan kata unik misalnya “sebuah”, “oleh”, “pada”, dan sebagainya. Contoh dari tahap *stopword* dapat dilihat pada Tabel 3.

Tabel 3. *Stopword*

Teks Input	Teks Output
korona bikin orang kesel menghambat jalan	korona bikin orang kesel menghambat jalan
hari ini dikantor pos dengan keadaan yang riuh pencairan dana bagi warga yang mendapat bantuan	dikantor pos riuh pencairan dana warga bantuan

mentan saya lagi fokus ke kayu putih buat bikin sovenir kalung anti korona untuk warga	mentan fokus kayu putih bikin sovenir kalung anti korona warga
--------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------

3.4 Stemming

Tahap *Stemming* adalah proses menghapus imbuhan, awalan, akhiran yang bertujuan untuk mengubah kata-kata sesuai dengan kata dasarnya. Contoh dari tahap *stemming* dapat dilihat pada Tabel 4.

Tabel 4. *Stemming*

Hasil Filtering	Hasil Stemming
korona bikin orang kesel menghambat jalankorona dikantor pos riuh pencairan dana warga bantuan	bikin orang kesel hambat jalan kantor pos riuh cair dana warga bantu
mentan fokus kayu putih bikin sovenir kalung anti korona warga	mentan fokus kayu putih bikin sovenir kalung anti korona warga

3.5 Pembobotan Kata

Dalam klasifikasi berita, pembobotan kata digunakan untuk mendapatkan suatu kategori. Salah satu metode pembobotan adalah TF-IDF (*Term Frequency – Inverse Document Frequency*). Nilai bobot suatu kata (term) menyatakan kepentingan bobot tersebut dalam merepresentasikan judul. Pada pembobotan TF-IDF, bobot akan semakin besar jika frekuensi kemunculan kata semakin tinggi, tetapi bobot akan berkurang jika kata tersebut semakin sering muncul pada berita lainnya.

$$idf = \log\left(\frac{N}{df}\right) \quad (1)$$

Dari persamaan (1) diatas diperoleh hasil pembobotan kata pada tabel 5.

Tabel 5. Perhitungan TF-IDF

No	Kata	Doc 1	Doc 2	Doc 3	df	Idf	Tf.idf		
							Doc 1	Doc 2	Doc 3
1	korona	1	0	1	2	0.18	0.18	0	0.18
2	bikin	1	0	0	1	0.48	0.48	0	0
3	orang	1	0	0	1	0.48	0.48	0	0
4	kesel	1	0	0	1	0.48	0.48	0	0
5	hambat	1	0	0	1	0.48	0.48	0	0
6	jalan	1	0	0	1	0.48	0.48	0	0
7	kantor	0	1	0	1	0.48	0	0.48	0
8	pos	0	1	0	1	0.48	0	0.48	0
9	riuh	0	1	0	1	0.48	0	0.48	0
10	cair	0	1	0	1	0.48	0	0.48	0
11	dana	0	1	0	1	0.48	0	0.48	0
12	warga	0	1	1	2	0.18	0	0.18	0.18
13	bantu	0	1	0	1	0.48	0	0.48	
14	mentan	0	0	1	1	0.48	0	0	0.48
15	fokus	0	0	1	1	0.48	0	0	0.48

16	kayu	0	0	1	1	0.48	0	0	0.48
17	putih	0	0	1	1	0.48	0	0	0.48
18	bikin	0	0	1	1	0.48	0	0	0.48
19	sovenir	0	0	1	1	0.48	0	0	0.48
20	kalung	0	0	1	1	0.48	0	0	0.48
21	anti	0	0	1	1	0.48	0	0	0.48

3.6 Penerapan Proses

Tahap ini merupakan tahap penentuan keterhubungan antara kata-kata pada data. Tahap ini menggunakan sebuah algoritma *Naive Bayes Classifier*. *Naive Bayes Classifier* terdiri dari dua proses dalam proses klasifikasi datanya. Kedua proses itu adalah proses pembelajaran *Naive Bayes Classifier* dan proses klasifikasi *Naive Bayes Classifier*.

Tahap proses prior probabilitas melakukan perhitungan pada kata yang terdapat *data testing* dengan persamaan (2) dan (3).

$$P(v_j) = \frac{docs_j}{|D|} \quad (2)$$

$$P(a_i|V_j) = \frac{n_{k+1}}{n+|kata|} \quad (3)$$

Dari persamaan (2) dan (3) diatas diperoleh hasil pada tabel 6.

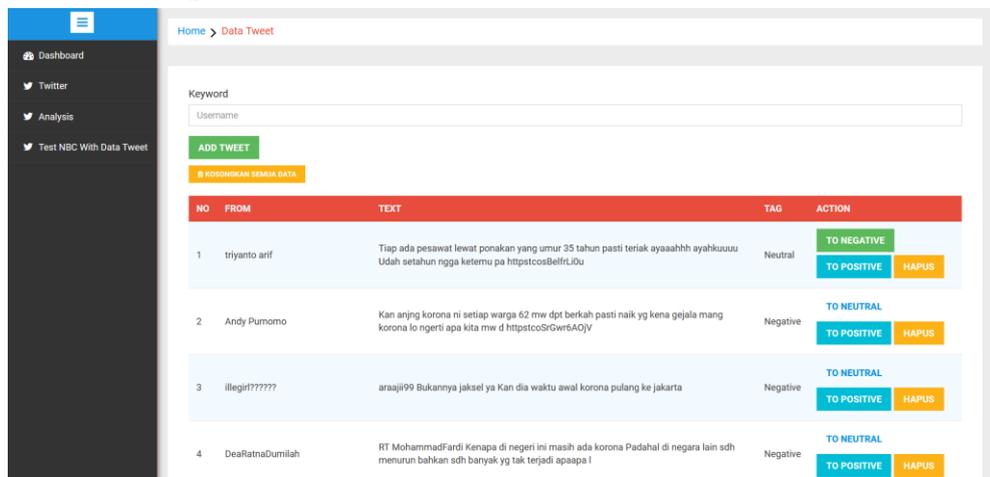
Tabel 6. Perhitungan Manual

No	Kata	Tf.idf			df	Idf	Tf.idf		
		Doc 1	Doc 2	Doc 3			Doc 1	Doc 2	Doc 3
1	korona	1	0	1	2	0.18	0.18	0	0.18
2	bikin	1	0	0	1	0.48	0.48	0	0
3	orang	1	0	0	1	0.48	0.48	0	0
4	kesel	1	0	0	1	0.48	0.48	0	0
5	hambat	1	0	0	1	0.48	0.48	0	0
6	jalan	1	0	0	1	0.48	0.48	0	0
7	kantor	0	1	0	1	0.48	0	0.48	0
8	pos	0	1	0	1	0.48	0	0.48	0
9	riuh	0	1	0	1	0.48	0	0.48	0
10	cair	0	1	0	1	0.48	0	0.48	0
11	dana	0	1	0	1	0.48	0	0.48	0
12	warga	0	1	1	2	0.18	0	0.18	0.18
13	bantu	0	1	0	1	0.48	0	0.48	
14	mentan	0	0	1	1	0.48	0	0	0.48
15	fokus	0	0	1	1	0.48	0	0	0.48
16	kayu	0	0	1	1	0.48	0	0	0.48
17	putih	0	0	1	1	0.48	0	0	0.48

18	bikin	0	0	1	1	0.48	0	0	0.48
19	sovenir	0	0	1	1	0.48	0	0	0.48
20	kalung	0	0	1	1	0.48	0	0	0.48
21	anti	0	0	1	1	0.48	0	0	0.48
total						8.46	2.58	3.06	3.24

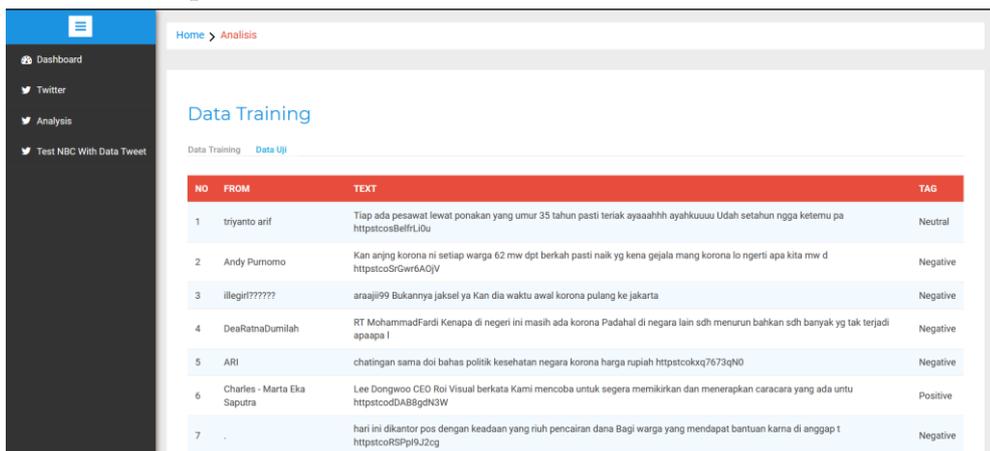
3.7 Penerapan Tampilan

Halaman Data tweet berfungsi untuk melihat *data training* dimana *user* dapat menambahkan data tweet baru, memberikan tag pada data tweet, atau menghapus data tweet. Gambar 4 adalah hasil implementasi halaman data tweet.



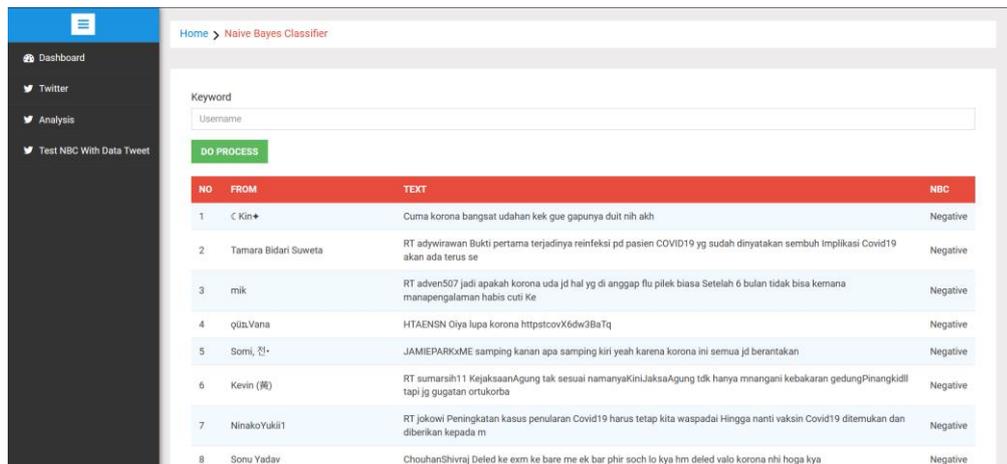
Gambar 4. Halaman Data Tweet

Halaman analisis berfungsi untuk melihat *data training* dan data uji yang digunakan. Gambar 5 adalah hasil implementasi halaman analisis.



Gambar 5. Halaman Analisis

Halaman test NBC berfungsi untuk menguji data testing menggunakan data tweet, user dapat menginputkan keyword yang ingin ditest. Gambar 6 adalah hasil implementasi halaman test NBC.



Gambar 6. Halaman Test NBC

4. KESIMPULAN

Berdasarkan hasil implementasi dan uji coba yang telah dilakukan, maka dapat disimpulkan hasil penelitian ini adalah sebagai berikut :

1. Penelitian ini telah menghasilkan sistem klasifikasi sentiment analisis masyarakat terhadap kebijakan pemerintah tentang covid-19 dengan metode Naïve Bayes.
2. Berdasarkan hasil penelitian tingkat akurasi dari metode Naïve Bayes menggunakan perbandingan data pelatihan sebesar 70 dan data testing sebesar 30 mendapatkan Precision sebesar 38%, Recall sebesar 42%, F-Measure sebesar 40% dan tingkat akurasi sebesar 76%

5. SARAN

Untuk penelitian selanjutnya dapat menerapkan algoritma klasifikasi seperti SVM dan K-Means dan menerapkan jumlah data uji lebih banyak sehingga hasil yang diperoleh untuk prediksi klasifikasi lebih akurat

DAFTAR PUSTAKA

- [1] Anam, C., & Santoso, H. B. (2018). Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa. *Jurnal UPM, Vol.8 No.1, ISSN: 2088-4591*.
- [2] Antinasari, P., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Vol. 1, No. 12, e-ISSN: 2548-964X, 1733-1741*.
- [3] Darmawan, A., Kustian, N., & Rahayu, W. (2018). Implementasi Data Mining menggunakan Model SVM untuk Prediksi Kepuasan Pengunjung Taman Tabebuaya. *Jurnal String, Vol. 2 No. 3, ISSN: 2549 -2837*.

- [4] Ebizmba. (2015). Top 15 Most Popular Social Networking Sites.
- [5] Fadli, d. R. (2020). Coronavirus.
- [6] Hikmawan, S., Pardamean, A., & Khasanah, S. N. (2020). Sentimen Analisis Publik Terhadap Joko Widodo Terhadap Wabah Covid-19 Menggunakan Metode Machine Learning. *Jurnal Kajian Ilmiah, Vol. 20, No. 2, ISSN: 1410-9794.*
- [7] Kompas. (2020). *Fakta Terbaru soal Virus Corona dan Covid-19.*
- [8] Manalu, E., Sianturi, F. A., & Manalu, M. R. (2017). Penerapan Algoritma Naive Bayes untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan dan Jumlah Pemesanan pada CV. Papadan Mama Pastries. *Jurnal Mantik Penusa, Vol.1, No. 2, ISSN 2088-3943.*
- [9] Nugroho, D. G., Chrisnanto, Y. H., & Wahana, A. (2016). Analisis Sentimen pada Jasa Ojek Online menggunakan Metode Naiva Bayes. *Jurnal Nasional Sains dan Teknologi, ISBN 978-602-99334-5-1.*
- [10] Nurhuda, F., Sihwi, S. W., & Doewes, A. (2013). *Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier.*
- [11] Pressman, R. (2015). *Rekayasa Perangkat Lunak: Pendekatan Praktisi Buku I.* Yogyakarta: Andi.
- [12] Puspitadewi, I., Erwina, W., & Kurniasih, N. (2016). Pemanfaatan Twitter TMCPOOLDAMETRO dalam Memenuhi Kebutuhan Informasi Para Pengguna Jalan Raya. *Jurnal Kajian Informasi dan Perpustakaan, Vol.4 No.1, ISSN: 2303-2677, 21-28.*
- [13] Rustiana, D. (2017). Analisis Sentimen Pasar Otomotif Mobil: Tweet Twitter menggunakan Naive Bayes. *Jurnal SIMETRIS, Vol. 8, No. 1, ISSN: 2252-4983.*
- [14] Turland, M. (2019). *PHP Architect's Guide to Web Scraping with PHP.* University Dr: Muskeeteers.me.
- [15] Witten, I. H., & Frank, E. (2016). *Data Mining Practical Machine Learning Tools and Techniques (3rd ed).* Elsevier.