

Penerapan *Algoritma Naive Bayes* Dalam Analisis Kenaikan Bahan Bakar Minyak Pada Twitter

Rahmad Aldi Saputra^{1*}, Sejati Waluyo²

^{1,2}Fakultas Teknologi Informasi, Teknik Informatika, Universitas Budi Luhur, Jakarta, Indonesia

Email: ^{1*}rahmadaldisaputra666@gmail.com, ²sejati.waluyo@budiluhur.ac.id
(* : corresponding author)

Abstrak- Awal tahun 2022 dunia sedang dilanda konflik perang antara Ukraina dan Rusia yang menyebabkan ketidak stabilan harga minyak dunia karena banyak responden masyarakat yang mengeluhkan atas kenaikanBBM dengan menggunakan *Tweets* yang disampaikan oleh user kemudian diklasifikasikan kategori *Tweets* tersebut apakah bersifat *Positive*, *Negative* atau *Neutral* dengan menggunakan program *metode Naive Bayes* untuk mendapatkan klasifikasi terbaik. dan dalam penelitian ini didapatkan responden positive, peneliti melakukan riset perbandingan antara data latih dan data uji yang saling terkait dengan perbandingan 80- 20 mendapatkan nilai Benar sebesar 81.00%, 70-30 mendapatkan nilai Benar sebesar 83.00% dan 60-40 mendapatkan nilai Benar sebesar 77.50%.

Kata Kunci: analisis sentimen, naive bayes classifier, preprocessing, tweets, kenaikan BBM

Appllication Of Naive Bayes Algorithm In Analysis Fuel Increase On Twitter

Abstract- At the beginning of 2022 the world was being hit by a war conflict between Ukraine and Russia which caused instability in world oil prices because many community respondents complained about the increase in fuel using Tweets submitted by users and then classified the categories of Tweets as Positive, Negative or Neutral using the program method. Naive Bayes to get the best classification. And in this study positive respondents were found, the researchers conducted comparative research between training data and test data which were interrelated with a comparison of 80-20 getting a True value of 81.00%, 70-30 getting a True value of 83.00% and 60-40 get a True value of 77.50%.

Keywords: sentiment analyst, naive bayes classifier, preprocessing, tweets, increase BBM

1. PENDAHULUAN

Pada tahun 2022 telah terjadi konflik antara rusia dan ukraina yang berdampak pada bahanbakar dunia. Yang membuat negara-negara di Asia Tenggara harus mengantisipasi perluasan konflik agar tidak berpindah ke kawasan Asia Tenggara. konsolidasi kekuatan di kawasan Indo-Pasifik sebelumnya dibuat melalui deklarasi pakta pertahanan Australia, United Kingdom dan United States of America (AUKUS) [1].

Seiring terjadinya konflik menyebabkan tidak kestabilan harga minyak dunia karena Russia adalah salah satu produsen dunia minyak bumi kalium karbonat (potash) bahan baku pupuk, dan industri pertambangan seperti nikel, alumunium dan palladium. pemulihan ekonomi global juga terancam akan lebih rendah dari prediksi awal. pemulihan ekonomi dunia post COVID-19 dengan ancaman inflasi sebelumnya telah terlihat di beberapa negara maju seperti Amerika Serikat hingga Indonesia. prediksi awal pertumbuhan ekonomi global diprediksi 4.4% pada 2022, 3.8% pada 2023, 3.9% pada negara maju di tahun 2022, dan 2.6% di negara berkembang pada tahun 2023 [2].

Oleh karena itu penulis mencoba melakukan sentimen analisis kenaikan bahan bakar minyak(BBM) pada twitter. Sebab Twitter merupakan salah satu media sosial yang bisa digunakan untuk mendapatkan opini masyarakat terhadap isu Kenaikan Bahan Bakar Minyak. Tweet memiliki sebuah teks yang panjangnya 280 karakter [3].

Menanggapi hal itu masih Banyaknya noise seperti (emote,tanda baca dan huruf kapital) pada suatu kalimat yang membuat dataset tersebut tidak bersih, Banyaknya kata yang belum diubah ke kata dasar pada kalimat disentimen analisis dan Berapa tingkat akurasi yang didapatkan dari hasil klasifikasi menggunakan algoritma Naive Bayes. maka dilakukan Menghapus atau menghilangkan noise seperti (emote,tanda baca dan huruf kapital) pada suatu kalimat dengan menggunakan Cleansing atau CaseFolding, Mengubah tiap kata pada suatu kalimat menjadi kata dasar menggunakan Stemming dan Mengetahui tingkat akurasi yang didapatkan dari hasil klasifikasi menggunakan algoritma Naive Bayes.

Analisis sentimen adalah suatu cara untuk menilai opini tertulis atau lisan untuk menentukan suatu opini bersifat positif, negatif atau netral [4]. Terdapat beberapa metode yang telah digunakan untuk analisis sentimen ini, diantaranya pendekatan Lexicon Based [5], Algoritma K-Nearest Neighbor (KNN) [6], metode Naive Bayes [7]. Salah satu metode atau algoritma yang populer dalam pengklasifikasian pada analisis sentimen adalah

Algoritma Naïve Bayes[8]. Mengapa menggunakan metode Naive bayes karna data penelitian ini memiliki jumlah data klasifikasi 1000 baris dan juga metode naïve bayes memiliki tingkat akurasi yang tinggi.

2. METODOLOGI PENELITIAN

2.1 Analisis Sentimen

Analisis sentimen adalah proses menentukan emosi dan *mengklasifikasikan polaritas* teks dalam sebuah dokumen. kalimat untuk memungkinkan kategori ditentukan sebagai emosi positif, negatif, atau netral.

Dinyatakan secara positif atau negatif Analisis sentimen melibatkan penambangan data untuk menganalisis, memproses, dan *mengekstrak data tekstual* dalam suatu *entitas* [9].

2.2 Text Mining

Text mining adalah Penambangan teks bertujuan untuk menghasilkan informasi dari sekumpulan dokumen. Penambangan teks dapat menghasilkan informasi dengan memproses, mengelompokkan, dan menganalisis sejumlah besar data tidak terstruktur [10].

Penambangan teks digunakan untuk mendapatkan informasi yang berguna dari sekumpulan dokumen, termasuk sumber data teks tidak terstruktur. dalam text mining dapat menghasilkan analisis sentimen suatu pernyataan sebagai positif atau negative[11].

2.3 Twitter

Twitter adalah jejaring sosial sering dikunjungi. Twitter sekarang adalah sosmed yang paling banyak digunakan oleh orang-orang. Jadi, Twitter dapat dianggap sebagai pesaing jejaring sosial Facebook [12].

Twitter banyak digunakan oleh *Generasi Z* sebagai jejaring sosial sarana yang digunakan untuk memperoleh informasi serta sarana untuk mengenal pengguna lain.

2.4 Naive Bayes Classifier

Naïve Bayes adalah salah satu algoritma pembelajaran *induktif* yang paling efektif dan efisien untuk machine learning dan data mining. Performa naïve bayes yang kompetitif dalam proses klasifikasi walaupun menggunakan asumsi *keidependenan* atribut (tidak ada kaitan antar atribut) [13].

Naïve Bayes merupakan pengklasifikasian dengan metode *probabilitas* dan *statistik* yang dikemukakan oleh *ilmuwan Inggris Thomas Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *teorema Bayes*. *Teorema* tersebut dikombinasikan dengan "naive" dimana diasumsikan kondisi antar atribut saling bebas.berikut merupakan rumus dari naïve bayess [14].berikut merupakan rumus dari naïve bayes :

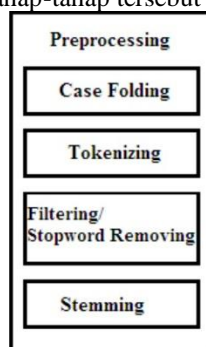
$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (1)$$

2.5 Data Penelitian

Teknik yang digunakan dalam pengumpulan data ini menggunakan *Crawling* pada Twitter, dengan judul dataset "Kenaikan BBM" dan jumlah data baris 1000 yang akan dibagi untuk data latih(*training*) dan data uji (*data testing*).

2.6 Preprocessing

Tahap Preprocessing adalah dataset yang dilakukan untuk proses preprocessing kemudian dijadikan data testing dan data training dalam 5 tahap [15], tahap-tahap tersebut adalah :



Gambar 1. *Preprocessing*

- a. *Case folding* adalah proses mengubah semua teks yang berhuruf kapital menjadi huruf kecil.

- b. *Tokenizing* adalah proses yang dilakukan penghilangan angka, tanda baca dan karakter lain yang dianggap tidak memiliki pengaruh terhadap pemrosesan *teks*.
- c. *Stopword removal* adalah tahap pemilihan kata-kata yang dianggap penting. Karakteristik utama dalam pemilihan *stopwords* biasanya adalah kata yang mempunyai frekuensi kemunculan yang tinggi misalnya kata penghubung seperti “dan”, “atau”, “tapi”, “akan” dan lainnya. Tujuan utama dalam penerapan proses *stopwords* ini adalah mengurangi jumlah kata dalam sebuah dokumen.
- d. *Stemming* merupakan suatu proses yang *mentransformasi* kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya (*root word*) dengan menggunakan aturan- aturan tertentu.

2.7 Rancangan Pemodelan

Pada tahapan ini peneliti melakukan Rancangan pemodelan terhadap dataset yang telah dilakukan *preprocessing*. Dataset tersebut lalu akan dipecah menjadi data latih dan data uji. Kemudian sudah dilakukan pengujian dengan *split presentase data* 60%-40%, 70%-30%, dan juga 80%-20%. Pada tahap ini peneliti menggunakan algoritma Naive Bayes.

2.8 Evaluasi

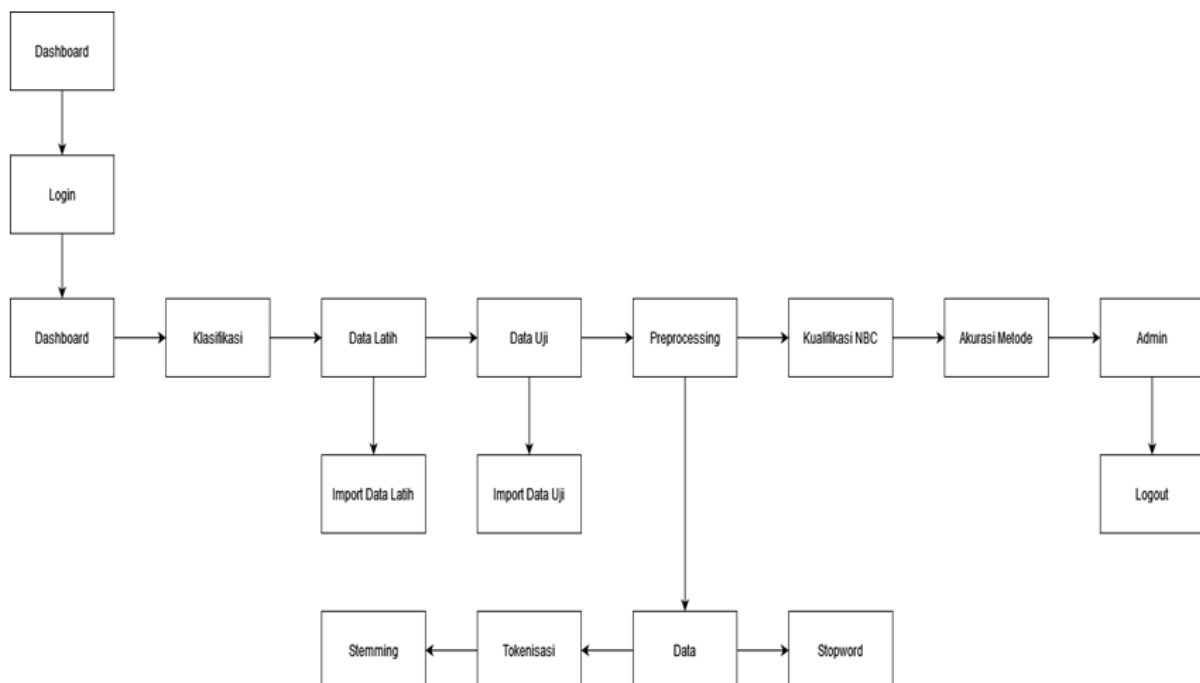
Pada tahap ini peneliti akan melakukan evaluasi metode klasifikasi dengan menghitung performa untuk mengukur seberapa besar presentase akurasi, presisi, dan *recall* yang didapatkan menggunakan algoritma Naive Bayes tersebut.

2.9 Rancangan Pengujian

Pada tahapan ini dilakukan pengujian terhadap aplikasi yang dibuat yakni *Naive Bayes Classifier* untuk mengukur akurasi perhitungan suatu data yaitu sentimen positive, negative dan neutral. pengujian dilakukan dengan menggunakan *data sample* dengan *split data* 60% data latih 40% data uji, 70% data latih 30% data uji, dan 80% data latih 20% data uji.

2.10 Rancangan Menu

Berikut merupakan rancangan menu dari aplikasi *Naive Bayes Classifier* yang terdiri dari Dashboard, Klasifikasi, Data Latih, Data Uji, Preprocessing, Kualifikasi NBC, Akurasi Metode dan Admin



Gambar 2. Rancangan Menu

3. HASIL DAN PEMBAHASAN

3.1 Data Collection

Data yang digunakan pada penelitian ini merupakan data tweet dari twitter. Data yang didapatkan pada penelitian ini berjumlah 6.002 tweet. Peneliti menggunakan data tweet dari tiap username.

Tabel 1. Total Dataset

Keyword	Tanggal	Jumlah
Kenaikan BBM	3 juni 2022 – 4 Agustus 2022	6.002

Berikut ini merupakan sampel dari data Kenaikan BBM.

Tabel 2. Sample Dataset

Tweet	Sentimen
dulu sebelum pdi penguasa harga bbm gk mahal seperti sekarang mengapa karena pdi p rajin demo kalau bbm naik	Negative
harga minyak dunia naek makanya bbm naik	Neutral
aman aja sih dengan naik nya bbm	Positive

3.2 Data Preparation

Pada tahapan *Data Preparation* ini penulis akan membagi menjadi 3 bagian sebagaiberikut :

1. Pemberian label sentimen pada setiap Tweet, di tahap ini berisi proses pemberian Klasifikasi sentimen positif, negatif, dan netral dalam dataset.
2. Tahap *preprocessing*, tahapan ini merupakan proses merapikan setiap Tweet yang sudah klasifikasi.
3. Menentukan data training dan data testing. disini peneliti melakukan perbandingan antara 70% data training dan 30% data testing

3.3 Proses Pembersihan data duplikat

Pada tahapan ini dilakukan pengurangan data untuk melakukan perhitungan indeks klasifikasi dengan menggunakan metode naïve bayes.

Tabel 3. remove duplicate

Proses	Jumlah Data
Jumlah data awal keyword Kenaikan BBM	6.002 tweet
Remove duplicate	1.000 tweet

3.4 Tahapan Preprocessing

Langkah – langkah pembersihan data merupakan preprocessing, yang terdiri dari *Case Folding*, *Tokenize*, *Stopword Removal*, dan *Stemming*. Dalam proses *Stopword removal* peneliti menggunakan kamus *stopwords* dengan library *nlTK* dan pada bagian *stemming* peneliti menggunakan bantuan kamus besar bahasa Indonesia dengan library *Sastrawi*. Berikut ini merupakan hasil dari tahapan *preprocessing* :

Tabel 6. Tahapan Preprocessing

Preprocessing	Sebelum	Sesudah
<i>Case folding</i>	Berharap Perang segera berhenti agar BBM juga tidak Naik Terus	berharap perang segera berhenti agar bbm juga tidak naik terus.
<i>Tokenize</i>	berharap perang segera berhenti agar bbm juga tidak naik terus	“berharap”, “perang”, “segera”, “berhenti”, “agar”, “bbm”, “juga”, “tidak”, “naik”, “t erus”.
<i>Stopword removal</i>	“berharap”, “perang”, “segera”, “berhenti”, “agar”, “bbm”, “juga”, “tidak”, “naik”, “terus”	“berharap”, “perang”, “berhenti”, “agar”, “bbm” “tidak”, “naik”.
<i>Stemming</i>	“berharap”, “perang”, “berhenti”, “agar”, “bbm” “tidak”, “naik”	“agar”, “bbml”, “tidak”, “naik”.

3.5 Naive Bayes

naïve bayes adalah mengklasifikasikan probabilitas berdasarkan pembelajaran mesin atas *probabilitas* lain. *Naive bayes* dapat dihitung secara manual untuk memprediksi masa yang akan datang berikut contoh perhitungan *naïve bayes*

Tabel 12. *Data Sample Naive Bayes*

Sample tweet	Sentimen
jangan khawatir pasti ada solusi supaya harga bbmgak naik	Positive
tenang yang naik hanya bbm non subsidi	Positive
aman aja sih dengan naiknya bbm	Neutral
bbm gas naik untuk kepentingan siapa	Neutral
iya anjir bener gw menikmati harga mahal pajak listrik bbm naik	Negative

Karena terdapat 2 Data Positif, 2 Data Neutral dan 1 Data Negative pada 5 dataset di atas. Maka, probabilitas pada data positive $P(\text{positif}) = \frac{2}{5} = 0,4$. Probabilitas pada data neutral $P(\text{neutral}) = \frac{2}{5} = 0,4$ Dan probabilitas pada Data Negative $P(\text{negative}) = \frac{1}{5} = 0,2$.

Dibawah ini ada beberapa kata yang dihitung pada table di atas, dilakukan perhitungan probabilitas pada Kelas Positive, Kelas Negative dan Kelas Neutral sebagai berikut:

$$P(\text{solusi}|\text{positif}) = \frac{1}{17} + \frac{1}{40} = \frac{2}{57} = 0.03508$$

$$P(\text{tenang}|\text{positif}) = \frac{1}{17} + \frac{1}{40} = \frac{2}{57} = 0.03508$$

$$P(\text{siapa}|\text{neutral}) = \frac{1}{12} + \frac{1}{40} = \frac{2}{52} = 0.03846$$

$$P(\text{sih}|\text{neutral}) = \frac{1}{12} + \frac{1}{40} = \frac{2}{52} = 0.03846$$

$$P(\text{anjing}|\text{negatif}) = \frac{1}{11} + \frac{1}{40} = \frac{2}{51} = 0.03921$$

$$P(\text{anjing}|\text{negatif}) = \frac{1}{11} + \frac{1}{40} = \frac{2}{51} = 0.03921$$

Tabel 13. *Probabilitas Kata*

Word	Nk	n	Vocabulary	P
Jangan	1	17	40	0.035
Khawatir	1	17	40	0.035
Pasti	1	17	40	0.035
Ada	1	17	40	0.035
Solusi	1	17	40	0.035
Supaya	1	17	40	0.035
Harga	1	17	40	0.035
Bbm	1	17	40	0.035
Gak	1	17	40	0.035
Naik	1	17	40	0.035
Tenang	1	17	40	0.035
Yang	1	17	40	0.035
Naik	1	17	40	0.035
Hanya	1	17	40	0.035
Bbm	1	17	40	0.035
Non	1	17	40	0.035
Subsidi	1	17	40	0.035
Aman	1	12	40	0.038
Aja	1	12	40	0.038
Sih	1	12	40	0.038
Dengan	1	12	40	0.038
Naiknya	1	12	40	0.038
Bbm	1	12	40	0.038

Word	Nk	n	Vocabulary	P
Bbm	1	12	40	0.038
Gas	1	12	40	0.038
Naik	1	12	40	0.038
Untuk	1	12	40	0.038
Kepentingan	1	12	40	0.038
Siapa	1	12	40	0.038
Iya	1	11	40	0.039
Anjing	1	11	40	0.039
Bener	1	11	40	0.039
Gw	1	11	40	0.039
Menikmati	1	11	40	0.039
Harga	1	11	40	0.039
Mahal	1	11	40	0.039
Pajak	1	11	40	0.039
Listrik	1	11	40	0.039
Bbm	1	11	40	0.039
Naik	1	11	40	0.039

Table diatas adalah perhitungan tiap kata untuk menentukan klasifikasi atau kelas,selanjutnya kita dapat menentukan klasifikasi atau kelas pada Tweet yang baru dengan menggunakan metode naïve bayes.

Sebagai contoh kita akan menentukan kelas dari sebuah data baru yaitu sebagai berikut :

Tabel 14.Menentukan Sentimen

Data	Sentimen
beratnya bbm minyak	?????

Tabel diatas dapat dihitung apakah Tweets tersebut memiliki Label kelaspositive,negative dan neutral:

Positive : $p(\text{Positive}) \cdot p(\text{beratnya}|\text{positive}) \cdot p(\text{bbm}|\text{positive}) \cdot p(\text{minyak}|\text{positive})$ $0.3 \times 0.01477 \times 0.01477 \times 0.01477 = 9,666355e - 7$

Neutral : $p(\text{Neutral}) \cdot p(\text{beratnya}|\text{neutral}) \cdot p(\text{beratnya}|\text{neutral}) \cdot p(\text{beratnya}|\text{neutral})$ $0.5 \times 0.02941 \times 0.02941 \times 0.02941 = 1,27190618e - 5$

Negative : $p(\text{Negative}) \cdot p(\text{beratnya}|\text{negative}) \cdot p(\text{bbm}|\text{negative}) \cdot p(\text{minyak}|\text{negative})$ $0.3 \times 0.01477 \times 0.01477 \times 0.01477 = 9,666355e - 7$

Dapat disimpulkan dari hasil perhitungan menggunakan metode naïve bayes bahwa probabilitas positive memiliki indeks $9,666355e-7$, Neutral memiliki indeks $1,27190618e-5$ dan Negative memiliki indeks $9,666355e-7$. Maka, dapat disimpulkan bahwa Tweet tersebut memiliki klasifikasi atau kelas label "Neutral"

Tabel 15.Hasil Prediksi Naive Bayes

Data	Sentimen
beratnya bbm minyak	Neutral

3.6 Hasil Pengujian Program

Dengan dataset Kenaikan BBM sebanyak 1.000 baris data dilakukan pengujian dan didapatkan hasil sebagai berikut, :

Tabel 16.Hasil Pengujian Program

Data Latih	Data Uji	Split Data	Akurasi	Presisi	Recall
800	200	80 – 20	81.00%	80.81 %	63.50%
700	300	70 – 30	83.00%	87.822 %	54.877%
600	400	60 – 40	77.50%	75.739 %	65.544%

4. KESIMPULAN

Penelitian dengan metode naïve bayes ini dapat menyimpulkan bahwa Opini masyarakat tentang kenaikan bahan bakar minyak mendapatkan responden positive, peneliti melakukan riset perbandingan antara data latih dan data uji yang saling terkait dengan perbandingan 80- 20 mendapatkan nilai Akurasi Benar sebesar 81.00%, 70-30 mendapatkan nilai Akurasi Benar sebesar 83.00% dan 60-40 mendapatkan nilai Akurasi Benar sebesar 77.50%

DAFTAR PUSTAKA

- [1] Ahmad Khoirul Umam. (2021) <https://www.bbc.com/indonesia/dunia-58592794>
- [2] Eisha. (2021) <https://www.detik.com/edu/detikpedia/d-5960883>
- [3] Munir, S. (2021). *Penggunaan Slang Pada Generasi Z Di Twitter. 1(69)*, 100.
- [4] Menurut Alsaedi & Khan. (2019) A Study On Sentiment Analysis Techniques Of Twitter Data (Ijaca) International Journal Of Advanced Computer Science And Applications, Vol. 10, No. 2, 2019
- [5] Effendy. (2017) Comparison Of Problems And Unmet Needs Of Patients With Advanced Cancer In A European Country And An Asian Country, Vol. 15, No. 5, 2017
- [6] Claudy Dkk. (2018) Analisis Sentimen Vaksin Covid-19 Di Indonesia Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization, Vol 3, No 2, 2019
- [7] Fanissa Dkk. (2018) Klasifikasi Berita Kriminal Menggunakan Naive Bayes Classifier (Nbc) Dengan Pengujian K-Fold Cross Validation, Vol 5, No 2, 2019
- [8] Amalya, A. A. (2020). Analisis Sentimen Kepuasan Pelanggan Terhadap Produk Dan Pelayanan Pt Pertamina Pada Twitter Menggunakan Algoritma Naive Bayes. In *Jurnal Aksara Komputer Terapan*. <https://jurnal.pcr.ac.id/index.php/jakt/article/view/4108>
- [9] Prabowo, W. A., & Wiguna, C. (2021). Sistem Informasi Umkm Bengkel Berbasis Web Menggunakan Metode Scrum. *Jurnal Media Informatika Budidarma*, 5(1), 149. <https://doi.org/10.30865/mib.v5i1.2604>
- [10] Denny Nathaniel Chandra Gede Indrawan I Nyoman Sukaraja (2017). Klasifikasi Berita Lokal Radar Malang Menggunakan Metode Naive Bayes Dengan Fitur N-Gram, Vol 10, No.1
- [11] Prabowo, W. A., & Wiguna, C. (2021). Sistem Informasi Umkm Bengkel Berbasis Web Menggunakan Metode Scrum. *Jurnal Media Informatika Budidarma*, 5(1), 149. <https://doi.org/10.30865/mib.v5i1.2604>
- [12] (Basri, 2017: 10) The Application Of Problem-Based Learning Strategy To Increase High Order Thinking Skills Of Senior Vocational School Student Sv10 N6 P123-129 2017
- [13] Rahmatullah, B. A. (2021). *Analisis Sentimen Mahasiswa Terhadap Perkuliahan Dalam Jaringan Menggunakan Metode Naive Bayes Classifier*.
- [14] Rahmatullah, B. A. (2021). *Analisis Sentimen Mahasiswa Terhadap Perkuliahan Dalam Jaringan Menggunakan Metode Naive Bayes Classifier*.
- [15] Resti. (2020) Text-Preprocessing Model Youtube Comments In Indonesian, Vol 4, No 4