

# Analisis Sentimen Masyarakat Terhadap Penghapusan Honorer Berdasarkan Opini Dari Twitter Menggunakan Naïve Bayes Classifier

Dwi Ratna Andriyani\*, M Afdal, Siti Monalisa

Sains dan Teknologi, Sistem Informasi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: <sup>1\*</sup>11950324645@students.uin-suska.ac.id, <sup>2</sup>m.afdal@uin-suska.ac.id, <sup>3</sup>siti.monalisa@uin-suska.ac.id

Email Penulis Korespondensi: 11950324645@students.uin-suska.ac.id

Submitted: 29/05/2023; Accepted: 12/06/2023; Published: 29/06/2023

**Abstrak**—Penghapusan honorer saat ini menjadi perbincangan hangat yang ada diseluruh Indonesia. Berbagai cara yang dilakukan tenaga honorer agar kebijakan penghapusan honorer tidak terlaksana. Tenaga honorer kebanyakan telah menjabat selama beberapa tahun, tetapi pemerintah telah mengeluarkan surat edaran tentang penghapusan honorer. Berbagai pro dan kontra masyarakat tentang penghapusan honorer, seperti tenaga honorer dapat kehilangan pekerjaan, tidak mendapatkan penghasilan, dan pengangguran semakin banyak. Tujuan penelitian ini adalah pemerintah dapat memberikan strategi yang harus dilakukan jika terjadi penghapusan honorer, seperti mengangkat seluruh honorer menjadi Pegawai Negeri Sipil atau Pegawai Pemerintah dengan Perjanjian Kerja. Maka penghapusan honorer tersebut menjadi salah satu trending topik pada sosial media Twitter pada tahun 2022. Dari hasil analisis yang dilakukan, opini masyarakat yang memanfaatkan Twitter sangat berpengaruh untuk tenaga honorer dengan mengelompokkan opini menjadi tiga kategori yaitu opini positif, opini netral, dan opini negatif. Maka penelitian dengan *text mining* tersebut menggunakan algoritma *Naïve Bayes Classifier* dengan data dari *tweets* Twitter dibulan Januari 2022 hingga Desember 2022 dengan 2.705 data. Hasil penelitian ini didapatkan akurasi dengan perhitungan 10 *K-fold Cross Validation* terdapat pada K-10 yaitu sebesar 73,01%. Dan didapatkan polaritas sentimen terhadap penghapusan honorer pada sentimen kelas positif sebesar 10% dengan 285 data tweet, sentimen kelas netral sebesar 67% dengan 1.801 data tweet, serta sentimen kelas negatif sebesar 23% dengan 619 data tweet.

**Kata Kunci:** Analisis Sentimen; Honorer; K-Fold Cross Validation; Naïve Bayes Classifier; Text Mining; Twitter

**Abstract**—The removal of honorees is currently a hot topic throughout Indonesia. Sharing how honorary personnel do so that the honorary removal policy is not implemented. Most honorary personnel have served for several years, but the government has issued a circular on the abolition of honorees. Various pros and cons of society regarding the abolition of honorees, such as honorary workers can lose their jobs, not get income, and unemployment is increasing. The purpose of the study is that the government can provide strategies that must be carried out in the event of the removal of honorees, such as appointing all honorees to become Civil Servants or Government Employees with Work Agreements. So the removal of the honoree became one of the trending topics on Twitter social media in 2022. From the results of the analysis conducted, public opinion that uses Twitter is very influential for honorary workers by grouping opinions into three categories, namely positive opinions, neutral opinions, and negative opinions. So the study with text mining used the Naïve Bayes Classifier algorithm with data from Twitter tweets from January 2022 to December 2022 with 2,705 data. The results of this study obtained accuracy with 10 K-fold Cross Validation on K-10, which was 73.01%. And it was found that sentiment polarity against the removal of honorees on positive class sentiment by 10% against agreeing to remove honorees with 285 data tweets, neutral class sentiment by 67% against agreeing and disagreeing with the removal of honorees with 1,801 data tweets, and negative class sentiment by 23% against disagreeing with the removal of honorees with 619 data tweets.

**Keywords:** Sentimen Analysis; Honorary; K-Fold Cross Validation; Naïve Bayes Classifier; Text Mining; Twitter

## 1. PENDAHULUAN

Media sosial adalah media *online* yang memiliki berbagai aplikasi berbasis internet yang didalamnya dirancang untuk mendorong interaksi sosial antar manusia dan memiliki fungsi yang berbeda-beda untuk masing-masing sehingga siapa pun yang menggunakan dapat memperoleh manfaat darinya [1]. Manfaat dari media sosial sangat berdampak besar bagi masyarakat Indonesia untuk aktif dimedia sosial [2]. Media sosial Twitter paling banyak diminati masyarakat Indonesia. Twitter merupakan sosial media dengan pengguna terbanyak di Indonesia [3]. Dengan terus meningkatnya penggunaan Twitter, maka masyarakat memanfaatkan keadaan untuk memberikan kritik dan sarannya tentang Penghapusan Honorer [4]. Berdasarkan surat edaran B/185/M.SM.02.03/2022 yang dikeluarkan oleh Menteri PAN-RB terdapat tiga alasan penghapusan tenaga honorer pada tahun 2023 yaitu, (1)pemerintah ingin memperjelas status tenaga honorer, (2)pemerintah ingin memberikan gaji yang layak bagi tenaga honorer yang telah mengabdikan, dan (3)pemerintah ingin memberikan tunjangan untuk kejelasan dan penghasilan yang lebih tinggi.

Dinas Pendidikan adalah salah satu unit pemangku negara dipimpin oleh pemerintah daerah (perda) [5]. Pada Dinas Pendidikan Provinsi Riau memiliki tenaga kerja PNS dan Honorer. Pada instansi pemerintah, sulit mendapatkan pegawai karena sulitnya persetujuan dari Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Republik Indonesia (KemenPAN-RB). Hal inilah yang memicu munculnya pegawai tidak tetap atau honorer. Namun, setelah puluhan tahun menjadi honorer pemerintah mengumumkan akan menghapus honorer. Dari wawancara yang dilakukan kepada pegawai honorer di Dinas Pendidikan Provinsi Riau ada yang setuju dan tidak setuju dihapusnya honorer tahun 2023. Dimana seluruh pegawai honorer banyak menjadi pengangguran, tidak mendapatkan pekerjaan, tidak mendapatkan penghasilan, dan pemerintah pun tidak memberikan solusi terhadap Penghapusan Honorer. Maka, pegawai honorer meminta kepada pihak pemerintah untuk memberikan strategi yang harus dilakukan dengan dihapusnya honorer. Strateginya adalah seluruh pegawai



honorar langsung diangkat menjadi Pegawai Negeri Sipil (PNS) dan Pegawai Pemerintah Perjanjian Kerja (P3K). Selain itu, pemerintah harus menyediakan lapangan pekerjaan bagi para honorar yang terdampak. Dalam proses Penghapusan Honorar ini, berbagai pendapat dari pengguna Twitter juga telah diterima, ada berpendapat sepakat dan tidak sepakat, dan ada pula melontarkan candaan berupa *tweets* komentar yang negatif.

Melihat hal tersebut diperlukan proses untuk menentukan *tweets* berupa sentimen positif, sentimen netral, dan sentimen negatif. Analisis sentimen melibatkan pemrosesan data tekstual seperti ulasan, tweet, dan postingan media sosial untuk mendapatkan informasi sentimen dari dokumen [6]. Analisis sentimen berguna untuk mengklasifikasikan data dengan langsung, sehingga data sentimen dapat mengandung kalimat opini [7]. Data opini tersebut dapat diolah menggunakan konsep *text mining*. *Text mining* dapat memberikan penjelasan yang bermanfaat dari sekumpulan data. Oleh dari itu, file teks dalam format yang sistematis atau tidak sistematis adalah data sumber yang digunakan dalam *text mining* [8]. Maka, terdapat beberapa aspek jika penelitian ini tidak dilakukan akan berdampak besar bagi politik, sosial dan ekonomi, hukum, serta pelayanan publik. Pada 3 aspek tersebut, ada beberapa pihak yang menguntungkan jika penghapusan honorar tidak terjadi yaitu pada (1) aspek politik: calon kepala daerah dan wakil kepala daerah dapat mencalonkan diri pada pilkada, (2) aspek sosial dan ekonomi: Badan Pusat Statistik (BPS) tidak akan terjadi peningkatan pengangguran, dan (3) aspek hukum: pemerintah serius dalam penataan sistem kepegawaian sehingga akan menghasilkan pegawai ASN yang profesional serta memiliki kejelasan status secara hukum dan administrasi.

Proses dari penelitian ini dimulai dari melakukan observasi dan wawancara pada Dinas Pendidikan Provinsi Riau untuk mewawancarai tenaga honorar tentang pendapatnya jika honorar dihapus. Pengumpulan data pada penelitian melalui *crawling* data Twitter melalui metode *snsrape* pada bulan Januari 2022-Desember 2022 pada akun Twitter resmi pemerintah serta dengan kata kunci “Hapus Honorar”. Data *crawling* tersebut belum terstruktur sistematis, maka proses data perlu dilakukan *Pre-processing* dengan menggunakan *case folding, cleaning, filtering, dan stemming*. Data twitter belum memiliki label opini positif, negatif, dan netral, dimana pelabelan dilakukan dengan tiga orang pakar bahasa. Kemudian data diubah menjadi nilai *numerik* dengan menggunakan pembobotan kata TF-IDF. Pada tahap klasifikasi NBC, pembagian data dibagi menjadi 2 yaitu, *data training dan data testing* dengan perhitungan 10 *K-Fold Cross Validation*. Proses akhir dilakukan visualisasi pada data Twitter Penghapusan Honorar.

Pada penelitian yang dilakukan oleh Nanang (2019) tentang Sentimen Twitter Menggunakan Klasifikasi *Naive Bayes* Untuk Penerapan Sistem Plat Ganjil Genap. Hasil penelitian tersebut berupa bentuk positif dan negatif dari klasifikasi teks, dimana menghasilkan nilai akurasi 86%, presisi 71%, dan recall 80% [9]. Selanjutnya penelitian Rilinka dkk (2021) menggunakan DFD dan *Multinomial NBC* tentang Penghapusan Ujian Nasional mendapatkan hasil akurasi bahwa 600 *data training* dan 200 *data testing* sebanyak 72% dengan pengujian *feature selection* DFD sebesar 73.13% [10]. Sementara itu, penelitian lain dilakukan oleh Safitri (2020) tentang Sentimen Masyarakat Terhadap Pemilu 2019 Pada Twitter dimana hasil penelitian *data training* sebesar 81% dan *data testing* sebesar 76% [11]. Dan terakhir penelitian Akhmad dkk (2022) tentang Klasifikasi Sentimen Twitter Penghapusan Tenaga Honorar Dengan Perbandingan Algoritma NBC, SVM, dan *Random Forest* didapatkan hasil dengan algoritma terbaik adalah *Random Forest* dengan akurasi 66,67% [12]. Sedangkan pada penelitian ini dilakukan Analisis Sentimen Masyarakat Terhadap Penghapusan Honorar Berdasarkan Opini Dari Twitter Menggunakan *Naive Bayes Classifier*. Pada penelitian ini didapatkan perbedaan dengan penelitian lain yaitu dengan menggunakan *crawling* data Twitter pada bulan Januari 2022-Desember 2022 metode perhitungan 10 *K-Fold Cross Validation* dengan nilai akurasi tertinggi terletak pada *split K-10* yaitu sebesar 73.01%.

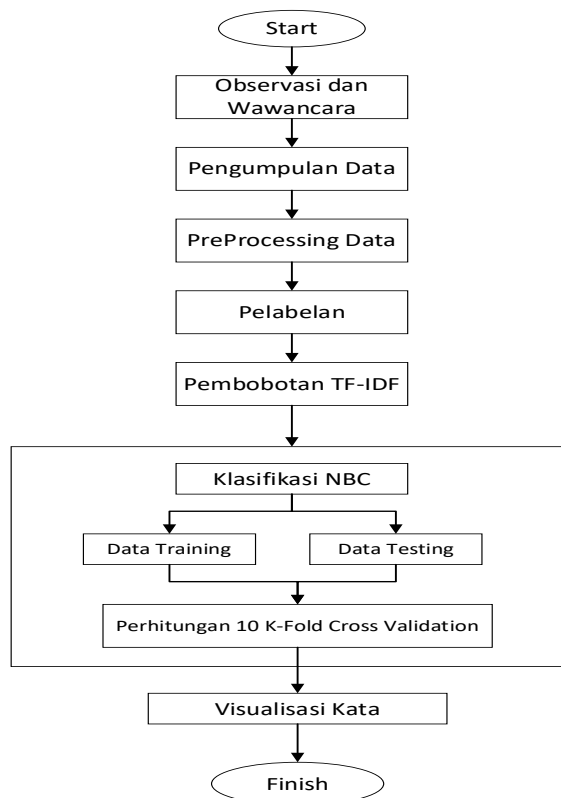
Pada penelitian analisis sentimen ini, metode NBC digunakan untuk menganalisis sentimen terhadap objek yang berisi kata netral, positif dan negatif. NBC adalah algoritma yang memprediksi masa depan berdasarkan data historis, atau disebut *Teorema Bayes*. Algoritma yang memiliki nilai akurasi yang tinggi sehingga hasil dari prediksinya akurat disebut sebagai *Naive Bayes Classifier*[13]. Klasifikasi NBC menggunakan asumsi bahwa pengaruh nilai numerik atribut kelas yang diberikan tidak tergantung pada kelas atau tidak dipengaruhi oleh atribut lainnya[14].

Tujuan yang akan dicapai pada penelitian ini adalah untuk memberikan manfaat bagi tenaga honorar, agar pemerintah ikut serta dalam memberikan solusi atau strategi yang harus dilakukan jika tenaga honorar berdampak menjadi pengangguran, serta didapatkan pula akurasi algoritma NBC pada *split K-10* mencapai sebesar 73.01% dengan *crawling* data *tweets* pada akun Twitter resmi pemerintahan yaitu @kempnrb, @BKNgoid, dan @DPR\_RI, serta dengan kata kunci “Hapus Honorar” selama setahun terakhir pada bulan Januari 2022 hingga Desember 2022.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Pada penelitian ini terdapat metodologi penelitian yang digunakan secara garis besar terdiri dari 7 langkah penyelesaian. Berikut langkah-langkah penelitian pada Gambar 1.



**Gambar 1.** Metodologi Penelitian

Berikut dalam menyelesaikan penelitian ini dilakukan secara sistematis dengan tahapan-tahapan metodologi penelitian seperti berikut:

- a. **Observasi dan Wawancara**  
Observasi dilakukan di Dinas Pendidikan Provinsi Riau dengan melakukan wawancara kepada 10 orang pegawai honorer tentang pendapatnya terhadap Penghapusan Honorer.
- b. **Pengumpulan Data**  
Pada tahap pengumpulan data ini dilakukan pengambilan data melalui metode *crawling* data Twitter menggunakan metode *snsrape*. Penelitian ini mengumpulkan data melalui *crawling* di akun @kempnrb, @BKNgoid, dan @DPR\_RI serta dengan kata kunci “Hapus Honorer” pada twitter, dengan total 2.705 data.
- c. **Pre-Processing**  
Pada tahap ini dilakukan *Pre-processing* data karena data twitter yang dihasilkan tidak sistematis. Tahap *pre-processing* yang dilakukan adalah *case folding*, *cleaning*, *filtering*, dan *stemming*.
- d. **Pelabelan**  
Pada tahap ini pelabelan kata dilakukan untuk mendapatkan sentimen positif, sentimen netral dan sentimen negatif yang dikerjakan secara manual dengan tiga pakar bahasa yakni Elvina S.Pd., M.Pd., Eka Novalia S.Pd., beserta Nabila Umi Afifah, S.Pd.
- e. **Pembobotan TF-IDF**  
Digunakan teknik TF-IDF untuk melakukan pemrosesan kata dan dapatkan bobot nilai disetiap kata.
- f. **Klasifikasi NBC**  
Pada tahap klasifikasi NBC dibagi menjadi dua pembagian data yaitu *data training* dan *data testing* dengan menggunakan metode perhitungan *10 K-fold Cross Validation*. Dan Algoritma *Naïve Bayes Classifier* digunakan dalam proses klasifikasi penelitian ini.
- g. **Visualisasi Kata**  
Tahap akhir dari penelitian ini adalah memperoleh visualisasi kata dengan website *wordclouds*.

## 2.2 Analisis Sentimen

Analisis sentimen didefinisikan sebagai proses memperoleh informasi yang bermakna dari teks menggunakan teknik pemrosesan dan menentukan sikap penulis yang bersifat positif, negatif, atau netral [15]. Analisis sentimen merupakan salah satu cabang penelitian dari text mining. Fokus analisis sentimen adalah menganalisis opini dari sebuah dokumen teks [16].

## 2.3 Twitter

Sosial media gratis dapat dimanfaatkan pengguna umum untuk memberikan informasi dengan real-time dinamakan



Twitter [17]. Menurut [18] twitter adalah platform yang dirancang untuk saling berkomunikasi melalui pertukaran pesan.

### 2.4 Data Mining

*Data mining* berfokus dengan mengidentifikasi tren atau pola yang dimaksudkan untuk membantu prediksi peristiwa masa depan dalam dokumen data. Dan Langkah untuk mendapatkan bagian terpenting dari informasi dalam sekumpulan data. Hal ini berbeda dengan *text mining* yang terus menerus mencari informasi yang berhubungan dengan teks, sedangkan *data mining* lebih cenderung mencari informasi yang berhubungan dengan angka atau simbol [19].

### 2.5 Text Mining

*Text mining* merupakan teknik yang mengharuskan seseorang memperoleh informasi dengan file *manuskrip* yang mempunyai teks tidak sistematis dan menyimpan ringkasan data [20]. Tujuan dari *text mining* adalah untuk memperoleh pengetahuan implisit yang tersembunyi dalam literatur dan menyajikannya dalam bentuk eksplisit [21].

### 2.6 Honorer

Honorer adalah tenaga kerja dengan gaji setiap bulan berdasarkan banyaknya jam mengajar [22]. Honorer merupakan orang-orang yang telah ditunjuk oleh pejabat tinggi pemerintah untuk melakukan tugasnya, yang gajinya ditegaskan oleh APBN [23]. Menurut Mentri PAN-RB pada nomor B/185/M.SM.02.03/2022 mengenai status pegawai di sekitar pusat pemerintah dan pemerintah daerah disebutkan bahwa "Mentiadakan jenis kepegawaian selain PNS dan PPPK di lingkungan masing-masing instansi dan tidak melakukan rehabilitasi pegawai non-ASN" [24].

### 2.7 Algoritma Naïve Bayes Classifier (NBC)

Algoritma NBC adalah algoritma khusus untuk mengklasifikasikan variabel tertentu dengan memanfaatkan metode *probabilitas* dan *statistik* [25]. NBC menggunakan alat matematika yang dikenal sebagai konsep *probabilitas* untuk mendeteksi barangkali menjadi klasifikasi terbaik dengan memeriksa frekuensi setiap klasifikasi pada data *training* [26]. Berikut rumus untuk dasar *teorema Naïve Baiyes* [27].

$$P(h | e) = \frac{P(e|h).P(h)}{P(e)} \quad (1)$$

Dimana:

$P(h | e)$  = Probabilitas akhir dimana hipotesis  $h$  terjadi jika diberikan bukti  $e$ .

$P(e / h)$  = Probabilitas bukti  $e$  terjadi akan mempengaruhi hipotesis  $h$ .

$P(h)$  = Probabilitas awal  $h$  tanpa memandang bukti.

$P(e)$  = Probabilitas awal  $e$  terjadi tanpa memandang bukti lain.

## 3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data secara *crawling* pada bahasa pemrograman *python* 3.8.9 dari data Twitter sebanyak 2.705 data dengan metode *snsrape*. Data penelitian ini diambil pada data *tweets* dari akun Twitter resmi pemerintahan yaitu @kempamb, @BKNGoid, dan @DPR\_RI, serta dengan kata kunci "Hapus Honorer" dibulan Januari 2022 hingga Desember 2022. Kemudian data akan dilakukan *Pre-processing data* sebelum dioperasikan dengan algoritma klasifikasi *Naïve Bayes Classifier* (NBC), dikarenakan data ini masih terkandung data yang tidak sistematis. Kemudian dilakukan tahap pelabelan serta TF-IDF buat menentukan nilai bobot pada data. Data yang telah diberi pelabelan dan TF-IDF akan dilakukan tahap data uji dan data latih memakai teknik perhitungan 10 *K-Fold Cross Validation* dan kemudian menerapkan algoritma NBC untuk klasifikasi.

### 3.1 Data Awal

Penelitian menggunakan data awal dari *crawling* pada bulan Januari 2022 hingga Desember 2022 dari Twitter. Berikut merupakan data awal yang dilihat pada Tabel 1.

**Tabel 1.** Data Awal

No	Data Tweet
1	@detikfinance Guru honorer gakak melihat negara yg dibelanya menghambur2kan bonus kepada pegawai resminya... Kebijakan penghapusan honorer tahun 2023 adalah teror yg dilakukan pemerintah kepada org2 yg sukarela bekerja demi negara dan bangsa Jahat sekali bukan?
2	@madHink @MardaniAliSera @PNS_Ababil @kpertiwi29 prioritas draf usulan revisi UU ASN krg lbh terkait : otomastisasi pengangkatan honorer jd PNS, penghapusan Komisi ASN, dan kebijakan kepegawaian harus atas persetujuan DPR... letak transformasi progresifnya dmn ya utk birokrasi lebih baik?
...	...
2704	Baru juga awal 2022 tp pikiran udh ngebayangin 2023 bakal kaya gimna kalo honorer setiap instansi pd



di hapusin...Alhamdulillah aja klo honorer di hapus tp di ganti jd PPPK mah 😞😞😞  
 2705 Siap-siap! Tahun Depan MenPAN-RB Hapus Semua Tenaga Honorer dari Instansi Pemerintahan  
<https://t.co/VswOR95hpk>

Pada Tabel 1 dapat dilihat bahwa data awal penelitian dilakukan dengan *crawling data* yang didapatkan sebanyak 2.705 data dengan berbagai pendapat masyarakat tentang penghapusan honorer.

### 3.2 Preprocessing

Data dari bagian komentar atau *tweets* tentang Penghapusan Honorer yang diperoleh melalui *crawling* situs jejaring sosial media Twitter tidak dapat digunakan, karena sebagian besar terdiri dari teks atau komentar yang tidak terstruktur dengan banyak *noise*. Tahap *Pre-processing* adalah prosedur untuk menyiapkan data mentah sebelum memulai proses selanjutnya. Tujuan *Preprocessing* lainnya adalah menyiapkan data sehingga dapat digunakan dengan lebih mudah.

#### 3.2.1 Case Folding

Tahap ini adalah tahap pertama dalam *Pre-Processing*. *Case folding* merupakan langkah pertama pada persiapan dokumen. Dalam proses ini, setiap kata dalam teks dokumen diubah menjadi versi yang lebih kecil. Berikut hasil data tweet case folding di lihat pada Tabel 2.

Tabel 2. Hasil Case Folding

No	Data Tweet
1	@detikfinance guru honorer ngakak melihat negara yg dibelanya menghambur2kan bonus kepada pegawai resminya... kebijakan penghapusan honorer tahun 2023 adalah teror yg dilakukan pemerintah kepada org2 yg sukarela bekerja demi negara dan bangsa jahat sekali bukan?
2	madhink @mardanielisera @pns_ababil @kptiwi29 prioritas draf usulan revisi uu asn krg lbh terkait : otomastisasi pengangkatan honorer jd pns, penghapusan komisi asn, dan kebijakan kepegawaian harus atas persetujuan dpr... letak transformasi progresifnya dmn ya utk birokrasi lebih baik?
...	...
2704	baru juga awal 2022 tp pikiran udh ngebayangin 2023 bakal kaya gimna kalo honorer setiap instansi pd di hapusin...alhamdulillah aja klo honorer di hapus tp di ganti jd pppk mah 😞😞😞
2705	siap-siap! tahun depan menpan-rb hapus semua tenaga honorer dari instansi pemerintahan <a href="http://t.co/vswor95hpk">http://t.co/vswor95hpk</a>

#### 3.2.2 Cleaning

Kumpulan dokumen berbasis teks dimana akan dilakukan pembersihan data untuk menghapus kata *duplikat*, tautan URL, nama pengguna, dan pemformatan lainnya yang biasa disebut dengan *cleaning*. Sehingga hasilnya jumlah datanya sekarang berjumlah di lihat pada Tabel 3.

Tabel 3. Hasil Cleaning

No	Data Tweet
1	guru honorer ngakak melihat negara yg dibelanya menghamburkan bonus kepada pegawai resminya kebijakan penghapusan honorer tahun adalah teror yg dilakukan pemerintah kepada org yg sukarela bekerja demi negara dan bangsa jahat sekali bukan
2	madhink @mardanielisera @pns_ababil @kptiwi29 prioritas draf usulan revisi uu asn krg lbh terkait : otomastisasi pengangkatan honorer jd pns, penghapusan komisi asn, dan kebijakan kepegawaian harus atas persetujuan dpr... letak transformasi progresifnya dmn ya utk birokrasi lebih baik?
...	...
2704	baru juga awal tp pikiran udh ngebayangin bakal kaya gimna kalo honorer setiap instansi pd di hapusin alhamdulillah aja klo honorer di hapus tp di ganti jd pppk mah
2705	siap siap tahun depan menpan rb hapus semua tenaga honorer dari instansi pemerinta

#### 3.2.3 Filtering

Setelah tahap *cleaning* selesai, proses *filtering* dilanjutkan dengan menggunakan *stopword* Tala untuk Bahasa Indonesia, yaitu menghapus kata-kata yang tidak diperlukan sehingga hanya kata-kata yang dapat mengubah nilai kata menjadi netral, negatif, dan positif yang digunakan. Berikut jumlah *stopword* yang digunakan di lihat pada Tabel 4.

Tabel 4. Data Kata Stopword

No	Stopword	No	Stopword	No	Stopword	No	Stopword
1	ada	6	akhiri	11	antaranya	16	atas



2	adalah	7	akhirnya	12	apa	17	atau
3	adanya	8	aku	13	apaan	18	ataukah
4	adapun	9	akulah	14	apabila	...	...
5	akhir	10	antara	15	asalkan	758	yang

Pada Tabel 4 dapat dilihat bahwa jumlah kata-kata stopword berjumlah 758 kata, seperti kata ada, adalah, adanya, adapun, antara, yang, dan lain sebagainya.

Berikut hasil filtering dilihat pada Tabel 5

**Tabel 5.** Hasil Filtering

No	Data Tweet
1	guru honorer ngakak negara dibelanya menghamburkan bonus pegawai resminya kebijakan penghapusan honorer teror pemerintah org sukarela negara bangsa jahat
2	babel prioritas draf usulan revisi uu asn krg lbh terkait otomastisasi pengangkatan honorer jd pns penghapusan komisi asn kebijakan kepegawaian persetujuan dpr letak transformasi progresifnya dmn ya birokrasi
...	...
2704	tp pikiran udh ngebayangin kaya gimna kalo honorer instansi pd hapusin alhamdulillah aja klo honorer hapus tp ganti jd pppk mah
2705	menpan rb hapus tenaga honorer instansi pemerintahan

### 3.2.4 Stemming

Sesudah tahap *filtering* selesai, Langkah berikutnya ialah tahap *stemming* yang melibatkan pembuatan kata imbuhan bagian depan dan belakang setiap hurufnya. Hasil *stemming* di lihat pada Tabel 6.

**Tabel 6.** Hasil Stemming

No	Data Tweet
1	guru honorer ngakak negara bela hambur bonus pegawai resmi bijak hapus honorer teror perintah org sukarela negara bangsa jahat
2	prioritas draf usul revisi uu asn krg lbh kait otomastisasi angkat honorer jd pns hapus komisi asn bijak pegawai tuju dpr letak transformasi progresif dmn ya birokrasi
...	...
2704	tp pikir udh ngebayangin kaya gimna kalo honorer instansi pd hapusin alhamdulillah aja klo honorer hapus tp ganti jd pppk mah
2705	menpan rb hapus tenaga honorer instansi pemerintah

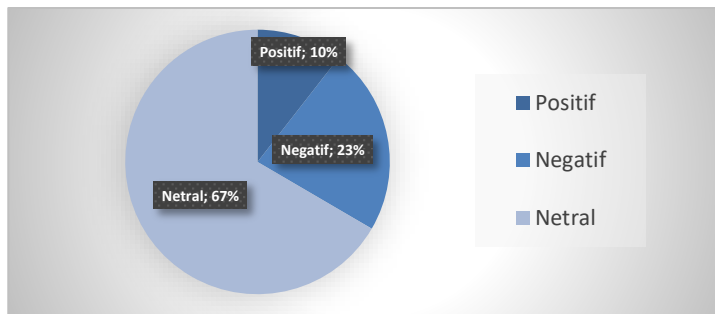
### 3.3 Pelabelan

Data Twitter yang ada pada penghapusan honorer merupakan data yang tidak memiliki label netral, positif dan negatif. Oleh karena itu, dibutuhkan metode khusus dalam menganalisis data *tweets* tersebut. Dalam penelitian ini, pelabelan data honorer ditangani dengan manual oleh tiga ahli pakar bahasa yaitu Elvina, M.Pd., S.Pd., Eka Novalia, S.Pd., dan Nabila Umi Afifah, S.Pd. Berikut merupakan hasil pelabelan *tweet* untuk Penghapusan Honorer seperti di Tabel 7 dan Gambar 3.

**Tabel 7.** Hasil Pelabelan

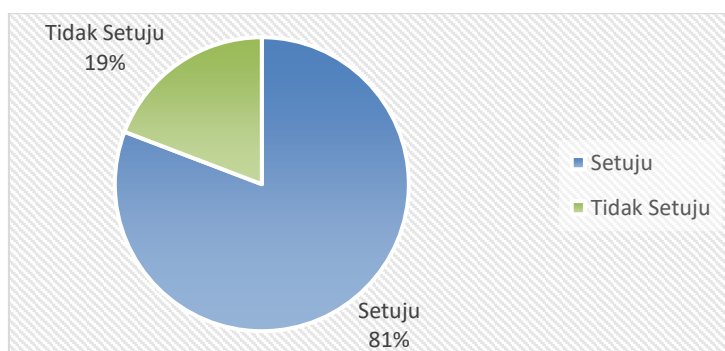
No	Data Tweet	Sentimen
1	guru honorer ngakak negara bela hambur bonus pegawai resmi bijak hapus honorer teror	Negatif
2	prioritas draf usul revisi uu asn krg lbh kait otomastisasi angkat honorer jd pns hapus komisi asn bijak pegawai tuju dpr letak transformasi progresif dmn ya birokrasi	Negatif
...	...	...
2704	tp pikir udh ngebayangin kaya gimna kalo honorer instansi pd hapusin alhamdulillah aja klo honorer hapus tp ganti jd pppk mah	Positif
2705	menpan rb hapus tenaga honorer instansi pemerintah	Netral

Pada Tabel 7 dapat dilihat bahwa hasil data *tweet* yang telah dilakukan pelabelan manual berada disentimen positif adalah sebanyak 285 tweet, disentimen netral adalah sebanyak 1801 tweet, dan disentimen negatif adalah sebanyak 619 tweet.



Gambar 2. Pelabelan Sentimen

Berdasarkan Gambar 2 didapatkan data tweet tertinggi pada kelas netral yaitu sebesar 67%, data tweet menengah kelas negatif yaitu sebesar 23%, dan data tweet terendah kelas positif yaitu sebesar 10%.



Gambar 3. Diagram Setuju dan Tidak Setuju

Pada Gambar 3 dataset sebanyak 2.705 data dibagi menjadi dua bagian yaitu data yang tidak setuju dihapusnya honorer dan setuju dihapusnya honorer. Maka didapatkan hasil akurasi dari tidak setuju dihapusnya honorer sebesar 81% dan setuju dihapusnya honorer sebesar 19%.

### 3.4 Term Frequency-Inverse Document Frequency (TF-IDF)

Setelah semua tahap *pre-processing* dan pelabelan selesai, data tersebut harus dalam bentuk *numerik*. Maka diperlukan mengubah data menjadi nilai *numerik* dengan menggunakan pembobotan TF-IDF [28]. Setiap kata dalam dokumen diberi bobot berdasarkan *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) [29]. Pada tahapan TF-IDF, perhitungan dilakukan dengan menggunakan *Python* dan model *Skicit-Learn*. TF-IDF disimpan sesuai dengan klasifikasi data latih dan uji. Berikut hasil TF-IDF dilihat pada Tabel 8.

Tabel 8. Hasil TF-IDF

No	Terms									
	angkat	asn	batal	bijak	bkn	tenaga	tolak	tunda	...	wacana
1	0,00	0,00	0,00	0,27	0,00	0,00	0,00	0,00	...	0,00
2	0,27	0,53	0,00	0,23	0,00	0,00	0,62	0,31	...	0,00
3	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	...	0,00
...	...	...	...	...	...	...	...	...	...	...
2705	0,00	0,00	0,00	0,00	0,00	0,23	0,00	0,00	...	0,00

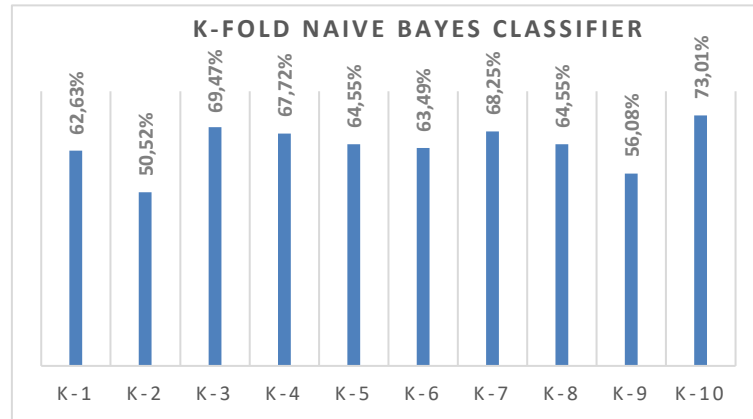
### 3.5 Algoritma Naïve Bayes Classifier (NBC)

Dalam penelitian ini, NBC menggunakan 2.705 set data dari tweet dengan 10 *K-Fold Cross Validation*. Data dibagi dengan dua kategori yaitu data pelatihan (*data training*) dan data pengujian (*data testing*). Dimana disisakan satu himpunan untuk uji data untuk setiap pengulangan dan sisanya untuk data latih dengan mengulang k-kali secara acak hingga bagian nomor k yang paling tidak rumit. Berikut hasil data latih dan data uji dengan 10 *K-fold Cross Validation* dilihat pada Tabel 9.

Tabel 9. Hasil Data Latih dan Data Uji Pada 10 K-fold

10 K-fold	Akurasi
K-1	62,63%
K-2	50,52%
K-3	69,47%
K-4	67,72%

K-5	64,55%
K-6	63,49%
K-7	68,25%
K-8	64,55%
K-9	56,08%
K-10	73,01%
<b>Nilai Max</b>	<b>73,01%</b>



**Gambar 4.** Data Latih dan Data Uji Dengan *K-fold Cross Validation* Pada NBC

Berdasarkan Tabel 8 dan Gambar 4 disimpulkan bahwa hasil pengujian dengan akurasi paling tinggi berada pada split K-10 dengan akurasi sebesar 73,01%. Lalu pengujian dengan nilai akurasi paling rendah terletak pada K-2 dengan nilai akurasi 50,52%. Bahwa pada K-10 digunakan sebagai hasil prediksi kelas sebab mendapatkan akurasi yang paling tinggi.

### 3.6 Analisa dan Hasil Klasifikasi

#### 3.6.1 Hasil Prediksi Kelas Data Tweet Menggunakan NBC

Dengan menggunakan metode 10 *K-Fold Cross Validation* pada pengolahan data diperoleh hasil prediksi data kelas uji (*testing*) dengan tingkat akurasi tertinggi yaitu K-10. Hasil prediksi kelas dapat dilihat Tabel 10.

**Tabel 10.** Hasil Prediksi Kelas Data Testing Pada K-fold

No	Negatif	Netral	Positif	Prediksi Kelas
1	0.133307	0.805703	0.060989	Netral
2	0.272214	0.672568	0.055218	Netral
3	0.131416	0.402422	0.057287	Negatif
...	...	...	...	...
809	0,443044	0,508948	0,048009	Netral
810	0,157113	0,763708	0,079179	Netral
811	0,225285	0,578927	0,195787	Netral

#### 3.6.2 Visualisasi

##### a. Visualisasi Semua Kata dan Visualisasi Sentimen Positif

Berikut hasil visualisasi semua kata menggunakan website *wordclouds* yaitu <https://wordclouds.com/>. Hasil visualisasi semua kata pada *tweets* data penghapusan honorer ditunjukkan pada Gambar 5. Dan hasil visualisasi kata sentimen positif pada data *tweets* Penghapusan Honorer ditunjukkan pada Gambar 6.



**Gambar 5.** Visualisasi Semua Kata



**Gambar 6.** Visualisasi Sentimen Positif

Berdasarkan Gambar 5 terdapat tiga kata yang paling sering muncul dalam data tweets Penghapusan Honorer adalah "hapus", "honorer", dan "tenaga". Sedangkan pada Gambar 6 dalam data *tweets* tentang Penghapusan Honorer, muncul tiga kata dengan frekuensi tertinggi pada sentimen positif, yaitu "pemerintah", "honorer", dan "bijak".

#### b. Visualisasi Sentimen Negatif dan Visualisasi Sentimen Netral

Hasil visualisasi kata sentimen negatif pada data *tweets* Penghapusan Honorer ditunjukkan pada Gambar 7. Hasil visualisasi kata sentimen netral pada data *tweets* Penghapusan Honorer ditunjukkan pada Gambar 8.



Gambar 7. Visualisasi Sentimen Negatif



Gambar 8. Visualisasi Sentimen Netral

Berdasarkan Gambar 7 pada data *tweets* tentang Penghapusan Honorer, muncul tiga kata dengan frekuensi tertinggi pada sentimen negatif, yaitu "ancam", "hapus", dan "konflik". Sedangkan pada Gambar 8 dalam data *tweets* tentang Penghapusan Honorer, muncul tiga kata dengan frekuensi tertinggi pada sentimen netral, yaitu "bijak", "rencana", dan "guru".

## 4. KESIMPULAN

Berdasarkan setiap tahapan data *tweets* di Twitter menggunakan bahasa *python* diperoleh 2.705 data mengenai opini masyarakat terhadap Penghapusan Honorer. Maka memperoleh hasil akurasi dengan Algoritma Naïve Bayes Classifier (NBC) adalah sebesar 73,01% pada *split* K-10 dengan pengklasifikasian data *tweets* menggunakan 10 *K-Fold Cross Validation*. Selain itu, diperoleh *polaritas* sentimen terhadap Penghapusan Honorer pada sentimen kelas positif 10% pada 285 data tweet, sentimen kelas netral 67% pada 1.801 data tweet, serta sentimen kelas negatif 23% pada 619 data tweet. Didapatkan juga hasil akurasi masyarakat yang setuju dihapusnya honorer sebesar 19% dan tidak setuju dihapusnya honorer sebesar 81%. Ditemukan pula keberadaan semua kata pada tiga frekuensi paling tinggi yaitu "hapus", "honorer", dan "tenaga". Dikelas sentimen positif, tiga kata pada frekuensi paling banyak yaitu "pemerintah", "honorer" dan "bijak". Dikelas sentimen negatif, tiga kata pada frekuensi paling banyak yaitu "ancam", "hapus" dan "konflik". Dan dikelas sentimen netral, tiga kata pada frekuensi paling banyak yaitu "bijak", "rencana" dan "guru". Selain itu tujuan penelitian ini untuk mengetahui seberapa besar contribution knowledge pemerintah tentang adanya penghapusan honorer terhadap tenaga honorer. Informasi ini diharapkan dapat dijadikan acuan pemerintah untuk mempertimbangkan keputusannya pada tenaga honorer yang berdampak menjadi pengganguran dan tenaga kerja di Indonesia akan menjadi semakin banyak. Untuk penelitian selanjutnya, harus dicoba untuk mengembangkan metode algoritma lain dan menggunakan data yang lebih real-time untuk memperoleh hasil akurasi yang lebih baik.

## REFERENCES

- [1] Y. H. Pranyoto and S. Geli, "Pengaruh Penggunaan Media Sosial Sebagai Media Pembelajaran Terhadap Hasil Belajar Kognitif Mahasiswa Sekolah Tinggi Katolik Santo Yakobus Marauke," *Jumpa*, vol. VIII, no. 1, pp. 32–33, 2020.
- [2] M. A. Harahap and S. Adeni, "Tren Penggunaan Media Sosial Selama Pandemi Di Indonesia," *J. Prof. FIS UNIVED*, vol. 7, no. 2, pp. 13–23, 2020.
- [3] F. A. Girnanfa and A. Susilo, "Studi Dramaturgi Pengelolaan Kesan Melalui Twitter Sebagai Sarana Eksistensi Diri Mahasiswa di Jakarta," *J. New Media Commun.*, vol. 1, no. 1, pp. 58–73, 2022, doi: 10.55985/jnmc.v1i1.2.
- [4] D. Darwis, N. Siskawati, and Z. Abidin, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional," *J. TeknoKompak*, vol. 15, no. 1, pp. 131–145, 2021, doi: <https://doi.org/10.33365/jtk.v15i1.744>.
- [5] Eka Viandari, "Dinas Pendidikan Provinsi se-Indonesia," *Quipper Blog*, 2021. <https://www.quipper.com/id/blog/info-furu/dinas-pendidikan-provinsi-se-indonesia/> (accessed Apr. 12, 2023).
- [6] S. Khairunnisa, A. Adiwijaya, and S. Al Faraby, "Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 406, 2021, doi: 10.30865/mib.v5i2.2835.
- [7] I. Susianti et al., "Prosiding Seminar Edusainstech FMIPA UNIMUS 2020 ISBN : 978-602-5614-35-4, "Analisis Sentimen Pada Twitter Terkait New Normal," Prosiding Seminar Edusainstech FMIPA UNIMUS 2020 ISBN : 978-602-5614-35-4," pp. 354–363, 2020.
- [8] G. E. I. Kamby and Dkk, "Penerapan Clustering pada Aplikasi Pendeteksi Kemiripan Dokumen Teks Bahasa Indonesia," Penerapan Clust. pada Apl. Pendeteksi Kemiripan Dok. teks Bhs. Indones., vol. 15, no. 2, pp. 75–82, 2020. <https://doi.org/10.35793/jti.15.2.2020.28907>.
- [9] S. N. Mandiri, "Analisis Sentimen Terhadap Penerapan Sistem Plat Nomor Ganjil / Genap Pada Twitter Dengan Metode



- Klasifikasi Naïve Bayes,” vol. 3, no. 1, pp. 94–99, 2019.
- [10] N. Yudistira, “Analisis Sentimen Penghapusan Ujian Nasional pada Twitter menggunakan Document Frequency Difference dan Multinomial Naïve Bayes,” *J. Pengemb. Teknol. Inf. dan Ilmu ...*, vol. 5, no. 3, pp. 876–883, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/8659/3981>
- [11] S. Juanita, “Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes,” *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 552, 2020, doi: 10.30865/mib.v4i3.2140.
- [12] A. Miftahusalam, A. F. Nuraini, A. A. Khoirunisa, and H. Pratiwi, “Perbandingan Algoritma Random Forest, Naïve Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Penghapusan Tenaga Honorer,” *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 563–572, 2022, doi: 10.34123/semnasoffstat.v2022i1.1410.
- [13] S. Lestari and S. Saepudin, “Analisis Sentimen Vaksin Sinovac Pada Twitter Menggunakan Algoritma Naive Bayes,” *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.*, vol. 1, no. 1, pp. 163–170, 2021.
- [14] N. Herlinawati, Y. Yuliani, S. Faizah, W. Gata, and S. Samudi, “Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine,” *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 5, no. 2, p. 293, 2020, doi: 10.24114/cess.v5i2.18186.
- [15] P. Nandwani and R. Verma, “A review on sentiment analysis and emotion detection from text,” *Soc. Netw. Anal. Min.*, vol. 11, no. 1, pp. 1–19, 2021, doi: 10.1007/s13278-021-00776-6.
- [16] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, “Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm,” *Procedia Comput. Sci.*, vol. 161, pp. 765–772, 2019, doi: 10.1016/j.procs.2019.11.181.
- [17] N. L. P. C. Savitri, R. A. Rahman, R. Venyutzky, and N. A. Rakhmawati, “Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning,” *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 47–58, 2021, doi: 10.28932/jutisi.v7i1.3216.
- [18] A. Yahyadi and F. Latifah, “Analisis Sentimen Twitter Terhadap Kebijakan PPKM di Tengah Pandemi COVID-19 Menggunakan Mode LSTM,” *J. Inf. Syst. Applied, Manag. Account. Res.*, vol. 6, no. 2, pp. 464–471, 2022, doi: 10.52362/jisamar.v6i2.791.
- [19] M. W. Pertiwi, “Analisis Sentimen Opini Publik Mengenai Sarana dan Transportasi Mudik Tahun 2019 Pada Twitter Menggunakan Algoritma Naïve Bayes, Neural Network, K-NN dan SVM,” *Inti Nusa Mandiri*, vol. 14, no. 1, pp. 27–32, 2019.
- [20] C. Prianto, N. H. Harani, and I. Firmansyah, “Analisis Sentimen Terhadap Kandidat Presiden Republik Indonesia Pada Pemilu 2019 di Media Sosial Twitter,” *J. Media Inform. Budidarma*, vol. 3, no. 4, p. 405, 2019, doi: 10.30865/mib.v3i4.1549.
- [21] S. Reddy et al., “Use and validation of text mining and cluster algorithms to derive insights from Corona Virus Disease-2019 (COVID-19) medical literature,” *Comput. Methods Programs Biomed. Updat.*, vol. 1, no. April, p. 100010, 2021, doi: 10.1016/j.cmpbup.2021.100010.
- [22] T. Sariwulan, I. Agung, U. Sudrajat, and G. Atmadiredja, “The influence of job expectation, job satisfaction, and government policy towards the work stress, job enthusiasm and continuance commitment of the honorarium teacher,” *Cakrawala Pendidik.*, vol. 38, no. 2, pp. 305–319, 2019, doi: 10.21831/cp.v38i2.24380.
- [23] S. Fitria and S. R. Nio, “Kepuasan Kerja Dan Loyalitas Kerja Pada Guru Honorer Sma Swasta Bukittinggi,” *Proyeksi*, vol. 15, no. 2, p. 131, 2020, doi: 10.30659/jp.15.2.131-140.
- [24] kepegawaian unmul, “Surat Edaran Mempan RB No. B/185/M.SM.02.03/2022 Tentang Status Kepegawaian Di Lingkungan Instansi Pemerintah Pusat Dan Pemerintah Daerah,” *SIDAK*, 2022. <https://kepegawaian.unmul.ac.id/post/surat-edaran-mempan-rb-no-b-185-m-sm-02-03-2022-tentang-status-kepegawaian-di-lingkungan-instansi-pemerintah-pusat> (accessed Apr. 02, 2023).
- [25] I. Kurniawan and A. Susanto, “Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden ( Pilpres ) 2019,” pp. 1–10, 2019, doi: 10.30864/eksplora.v9i1.237.
- [26] F. Syarifuddin, M. Misdrum, A. A. Widodo, P. S. Informatika, and U. M. Pasuruan, “Klasifikasi Data Set Virus Corona Menggunakan Metode Naïve Bayes Classifier,” *J. SPIRIT*, vol. 12, no. 2, pp. 46–52, 2020, doi: <http://dx.doi.org/10.53567/spirit.v12i2.184>.
- [27] A. Nugroho, R. Hidayatillah, S. Sumpeno, and M. H. Purnomo, “Klasifikasi Interaksi Kampanye di Media Sosial Menggunakan Naïve Bayes Kernel Estimator,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 8, no. 2, p. 107, 2019, doi: 10.22146/jnteti.v8i2.499.
- [28] M. P. Simatupang and D. P. Utomo, “Analisa Testimonial Dengan Menggunakan Algoritma Text Mining Dan Term Frequency- Inverse Document Frequence (Tf-Idf) Pada Toko Allmeeart,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 808–814, 2019, doi: 10.30865/komik.v3i1.1697.
- [29] I. Widaningrum, D. Mustikasari, R. Arifin, S. L. Tsaqila, and D. Fatmawati, “Algoritma Term Frequency – Inverse Document Frequency ( TF-IDF ) dan K-Means Clustering Untuk Menentukan Kategori Dokumen,” *J. SISFOTEK.*, vol. 6, no. 1, pp. 145–149, 2022.