



P-ISSN : 2339-1871

E-ISSN : 2715-7369

# JURNAL ILMIAH BETRIK

*Besemah Teknologi Informasi dan Komputer*

**VOLUME 12 ISSUE 2 Agustus 2021**



Published by :  
LPPM Sekolah Tinggi Teknologi Pagar Alam

# ANALISIS SENTIMEN SERIKAT PEKERJA PERTAMINA TOLAK AHOK PADA MEDIA SOSIAL YOUTUBE MENGGUNAKAN ALGORITMA NAIVE BAYES



PDF

Published: Aug 12, 2021

DOI:  
<https://doi.org/10.36050/betrik.v12i2.219>

**Keywords:**  
Sentiment analysis, Naive bayes, Youtube, Rstudio, Rapidminer Studio

Muhammad Soekarno Putra Putra  
Universitas Bina Dharma  
Sri Dharma Wati  
Universitas Bina Dharma  
Imam Solikin  
Universitas Bina Dharma

## Abstract

*In this study, an analysis of public sentiment will be carried out on the news of the Pertamina workers union rejecting Ahok which was revealed through the social media network youtube on the KompasTV channel. On the news there were about 25000 comments containing harsh words and even hate speech, this news attracted a lot of comments, there were those who praised, criticized and insulted. Sentiment analysis is the process of classifying textual documents into two classes, namely positive and negative sentiment classes. The data used in this study amounted to 1000 comments consisting of 1000 training data or training data and 300 test data or testing data, the classification method used in this study was naive bayes. In this study, data collection using Rstudio tools and processed using the rapidminer studio tools, and get the accuracy of 98.00%. From some of the data that has been tested, there are some data that are predicted to be neutral, positive data and data that are predicted to be negative. From the precision class, the neutral prediction has a class value of 100.00%, the positive prediction has a class value of 80.00% and the negative prediction has a precision class value of 66.67%.*

How to Cite

TERAKREDITASI



SINTA INDONESIA



### Additional Menu

Focus & Scope

Editorial Team

Reviewer Acknowledgement



ISSN : 2339 - 1871

## JURNAL ILMIAH BETRIK

Besemah Teknologi Informasi dan Komputer

**Editor Office :** LPPM Sekolah Tinggi Teknologi Pagar Alam, Jln. Masik Siagim No. 75  
Simpang Mbacang, Pagar Alam, SUM-SEL, Indonesia  
Phone : +62 852-7901-1390.  
Email : [betrik@sttpagaralam.ac.id](mailto:betrik@sttpagaralam.ac.id) | [admin.jurnal@sttpagaralam.ac.id](mailto:admin.jurnal@sttpagaralam.ac.id)  
Website : <https://ejournal.sttpagaralam.ac.id/index.php/betrik/index>

# ANALISIS SENTIMEN SERIKAT PEKERJA PERTAMINA TOLAK AHOK PADA MEDIA SOSIAL YOUTUBE MENGGUNAKAN ALGORITMA NAIVE BAYES

Muhammad Soekarno Putra<sup>1</sup>, Sri Dharma Wati<sup>2</sup>, Imam Solikin<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Bina Darma<sup>12</sup>

Program Studi Manajemen Informatika, Fakultas Vokasi, Universitas Bina Darma<sup>3</sup>

Jl. Jend. A.Yani No.3 Palembang 30251

Sur-el : [soekarno@binadarma.ac.id](mailto:soekarno@binadarma.ac.id)<sup>1</sup>, [sridharmawati.yi@gmail.com](mailto:sridharmawati.yi@gmail.com)<sup>2</sup>, [imamsolikin@binadarma.ac.id](mailto:imamsolikin@binadarma.ac.id)<sup>3</sup>

**Abstrak:** Penelitian ini melakukan analisis sentimen publik terhadap berita serikat pekerja pertamina tolak Ahok yang diungkapkan melalui jejaring sosial media youtube pada channel KompasTV. Pada berita tersebut terdapat sekitar 25000 komentar yang berisi kata-kata kasar bahkan ujaran kebencian, berita ini menuai banyak komentar ada yang memuji, mengkritik dan menghina. Analisis sentimen merupakan proses klasifikasi dokumen tekstual ke dalam dua kelas yaitu kelas sentimen positif dan negatif. Data yang digunakan pada penelitian ini berjumlah 1.000 komentar terdiri dari data latih atau data training 1000 dan data uji atau data testing 300, metode klasifikasi yang digunakan dalam penelitian ini adalah naive bayes. Pada penelitian ini pengambilan data menggunakan tools Rstudio dan diproses dengan menggunakan tools rapidminer studio, dan mendapatkan hasil akurasi nya sebesar 98.00%. Dari beberapa data yang telah di uji, terdapat beberapa data yang diprediksi neutral, data positif dan data yang diprediksi negatif. Dari kelas presisi, prediksi neutral memiliki nilai kelas 100.00%, prediksi positif memiliki nilai kelas sebesar 80.00% dan prediksi negatif memiliki nilai kelas presisi sebesar 66.67%.

**Kunci Utama :** Analisis sentimen; Naive bayes; Youtube; Rstudio; Rapidminer studio

**Abstract :** This research analyzes public sentiment on the news of the Pertamina workers union rejecting Ahok which was revealed through the social media network youtube on the KompasTV channel. On the news there were about 25000 comments containing harsh words and even hate speech, this news attracted a lot of comments, there were those who praised, criticized and insulted. Sentiment analysis is the process of classifying textual documents into two classes, namely positive and negative sentiment classes. The data used in this study amounted to 1000 comments consisting of 1000 training data or training data and 300 test data or testing data, the classification method used in this study was naive bayes. In this study, data collection using Rstudio tools and processed using the rapidminer studio tools, and get the accuracy of 98.00%. From some of the data that has been tested, there are some data that are predicted to be neutral, positive data and data that are predicted to be negative. From the precision class, the neutral prediction has a class value of 100.00%, the positive prediction has a class value of 80.00% and the negative prediction has a precision class value of 66.67%.

**Keywords :** Sentiment analysis; Naive bayes; Youtube; Rstudio; Rapidminer Studio

## 1. PENDAHULUAN

Pada saat ini media sosial telah menjadi alat komunikasi yang sangat populer di kalangan pengguna internet di

Indonesia, media sosial merupakan salah satu tempat atau media untuk berekspresi dan berpendapat tentang berbagai macam topik salah satunya *youtube*. (Hartoko, 2019) Menyatakan bahwa “Saat ini siapa yang tak mengenal *Youtube* situs berbagi video ini makin “*booming*” di Indonesia, situs yang didirikan pada bulan Februari 2005 ini terus menanjak penggunaannya di seluruh dunia hingga membuat pemilik *google* tergoda untuk membelinya pada tanggal 9 Oktober 2006”.

Pada saat sekarang ini *youtube* sudah mempunyai banyak pengguna, salah satu pengguna yang memanfaatkan *youtube* adalah KompasTv. *Channel* KompasTv adalah media sosial yang bergabung pada tanggal 23 Agustus 2013 yang mempunyai konten tentang berita, dalam penelitian ini *Channel* KompasTv yang membawakan berita tentang pekerja serikat Pertamina Tolak Ahok dengan jumlah komentar 25.000 yang diambil mulai dari tanggal 19 November 2019 sampai 15 Januari 2020, Dari 25.000 komentar tersebut muncullah permasalahan nya untuk menentukan komentar positif dan negatif dan salah satu cara untuk menganalisis data komentar tersebut adalah dengan menggunakan *Tools Rapidminer Studio*. Sistem klasifikasi dapat dilakukan secara manual, tetapi seiring bertambahnya sumber opini menjadi semakin banyak dan banyak pula waktu terbuang, besarnya polaritas komentar yang ditunjukkan oleh masyarakat bisa dijadikan parameter untuk melihat setuju atau tidaknya masyarakat terhadap Pak Ahok. Dari ribuan komentar yang ada dapat dimanfaatkan untuk melihat sentimen yang di munculkan polaritas positif atau negatifnya suatu opini yang dapat mewakili sebuah perasaan atau keadaan dari pengguna media sosial.

Ada beragam teknik klasifikasi dokumen, diantaranya adalah *Naive Bayes*, *Decision trees*, dan *Support vektor Machines (SVM)*. (Mahbubah & Zuliarso, 2019) menyatakan bahwa “Salah satu

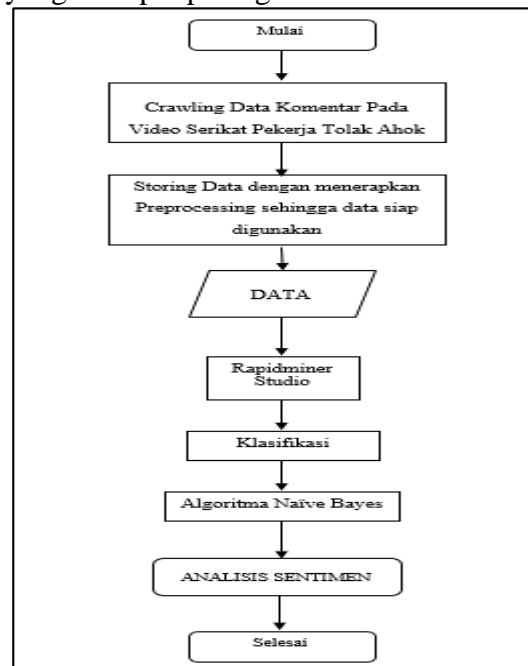
metode yang paling populer digunakan dalam pengklasifikasian dokumen sekarang ini adalah metode *Naive Bayes*. Metode *Naive Bayes* mempunyai kecepatan, sederhana dan akurasi yang tinggi ketika di aplikasikan dalam basis data besar dan data yang beragam”.

Tujuan dari penelitian ini adalah untuk mengetahui sejauh mana tingkat akurasi algoritma *naive Bayes* pada tahap analisis sentimen pada berita serikat pekerja Pertamina Tolak Ahok di media sosial *youtube channel* KompasTv.

## 2. METODE PENELITIAN

Metode penelitian yang digunakan adalah metode kuantitatif, merupakan suatu cara yang digunakan untuk menjawab masalah penelitian yang berkaitan dengan data berupa angka dan program statistik (Wahidmurni, 2017). Dan data kuantitatif yang di perlukan dalam penelitian ini adalah jumlah data komentar pada *channel youtube* KompasTv *hashtag* serikat pekerja Pertamina Tolak Ahok.

Adapun, proses-proses yang akan dilakukan pada penelitian ini dapat dilihat pada diagram alur atau *flowchart* yang terdapat pada gambar 1:



Gambar 1. Flow Chart Proses Penelitian

## 2.1 Analisis Sentimen

(Suyanto, 2017) Menyatakan “Analisis sentimen (*sentiment analysis*) dikenal juga sebagai ekstraksi opini (*opinion extraction*), penambangan sentimen (*sentiment mining*), dan analisis subjektivitas (*subjectivity analysis*), Secara istilah analisis sentimen adalah deteksi sikap-sikap (*attitude*) terhadap objek atau orang”.

(Arief & Imanuel, 2019) Menyatakan “Analisis sentimen atau penambangan opini dapat didefinisikan sebagai tugas mendeteksi, mengekstraksi, dan mengklasifikasikan pendapat tentang sesuatu hal, Tugas ini merupakan jenis pemrosesan bahasa alami (NLP) untuk melacak mood publik untuk hukum, kebijakan, atau pemasaran tertentu, dll”.

## 2.2 Text Mining

(Purbo, 2019) Menyatakan *text mining* merupakan “sebuah set metodologi dan *tool* untuk menemukan (*discovering*), memvisualisasikan (*presenting*), mengevaluasi pengetahuan dari kumpulan besar dari teks dokumen. *Text mining* juga disebut sebagai data text mining kira-kira setara dengan analisis teks adalah proses untuk memperoleh informasi berkualitas tinggi dari teks”.

(Han & Kamber, 2006) Menyatakan “*Text mining* bertujuan untuk mencari kata-kata yang dapat mewakili isi dari dokumen. Pada dasarnya proses kerja dari text mining banyak mengadopsi penelitian data mining, namun yang menjadi perbedaan adalah pola yang digunakan oleh text mining diambil dari sekumpulan bahasa alami yang tidak terstruktur, sedangkan dalam data mining pola yang diambil adalah dari data yang terstruktur”.

## 2.3 Algoritma Naive bayes

(Suntoro, 2019) menyatakan Algoritma *Naive Bayes* adalah salah satu algoritma klasifikasi berdasarkan teorema bayesian pada statistika, algoritma *Naive Bayes* dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Tahapan proses naive bayes yaitu menghitung jumlah kelas atau label, menghitung jumlah kasus perkelas, kalikan semua variable kelas, bandingkan hasil perkelas. Penggunaan metode *naive bayes classifier* pada penelitian ini didasarkan banyak data *set* yang dipakai sehingga dibutuhkan suatu metode yang mempunyai performansi yang cepat dalam pengklasifikasian serta keakuratan yang cukup tinggi. Dalam proses klasifikasi teks, yaitu tahap pelatihan dilakukan proses terhadap sampel data yang sedapat mungkin dapat menjadi representasi data, selanjutnya penentuan probabilitas *prior* bagi tiap kategori dari suatu data berdasarkan sampel data dan tahap klasifikasi ditentukan nilai kategori dari suatu data berdasarkan *term* yang muncul dalam data yang diklasifikasikan.

Teorema *naive bayes* dapat dinyatakan dalam persamaan dibawah ini.

$$P(X_k | Y) = \frac{P(Y | X_k)}{\sum_i P(Y | X_i)}$$

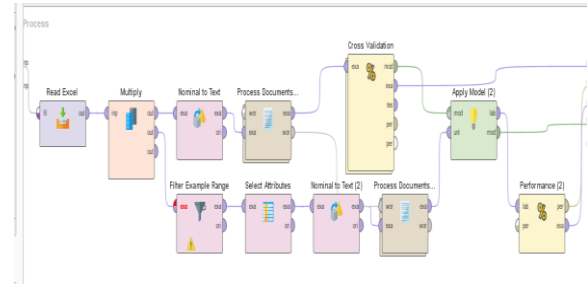
Dimana keadaan posterioir (probabilitas  $X_k$  di dalam  $Y$ ) dapat dihitung dalam keadaan prior (probabilitas  $Y$  didalam  $X_k$  dibagi dengan jumlah dari semua probabilitas  $Y$  di dalam semua  $X_i$ ).

## 2.4 Analisa Masalah

Dalam penelitian ini perlu dilakukan pengamatan data untuk proses analisis agar bisa dimengerti. Sehingga menjadi sebuah data yang siap digunakan dengan melalui sebuah proses analisis. Agar bisa diterapkan

pada analisis sentimen, penulis melakukan beberapa tahap seperti *crawling data*, *storing data* kemudian menggunakan algoritma *Naive Bayes*. Setelah melalui tahapan tersebut maka sumber data tersebut telah siap diolah pada proses analisis sentimen.

Sumber data yang didapatkan dari objek penelitian adalah data murni, yaitu data-data tersebut asli didapatkan dari objek penelitian dan tidak dimanipulasi. Sumber data yang digunakan adalah data komentar *YouTube* pada video yang berjudul “Serikat Pekerja Pertamina Tolak Ahok”. Video ini diupload pada *channel YouTube* KOMPASTV dan memiliki total 27,844 komentar pada video tersebut.



**Gambar 2. Rangkaian Utama Proses Klasifikasi**

### 3. HASIL DAN PEMBAHASAN

Proses klasifikasi dilakukan dengan menggunakan *software Rapidminer Studio*. Proses klasifikasi pada penelitian ini menggunakan algoritma *naive bayes classifier* dan data yang di dapat dari *crawling* data komentar *youtube* menggunakan *Rstudio* yang akan di proses menggunakan *tools rapidminer studio* untuk dilihat tingkat akurasi.

Tahapan selanjutnya yaitu mengkonversi data menjadi dokumen dengan menggunakan operator proses *document from data*, kemudian menentukan kolom mana yang akan menjadi *label* dengan memanfaatkan operator *set role*. Kemudian proses klasifikasi sentimen dilakukan menggunakan operator *naive bayes* dan *apply model* guna untuk pembentukan model data training yang akan di implementasikan ke data testing. Rangkaian proses utama klasifikasi dapat dilihat pada gambar 2.

### 3.1 Perhitungan Klasifikasi

Berikut contoh perhitungan yang diambil dari satu komentar *Channel YouTube KompasTV* sebagai berikut: “semangatt pak ahok!!!!!!!!!!” oleh *user Sheelma Iyennaurah* yang dipost pada 22 Januari 2020. Hasil pengolahan kata dapat dilihat pada tabel 1:

**Tabel 1. Contoh Hasil Preprocessing Komentar**

Kategori	Teks	Jumlah kata	Jumlah huruf
Kotor	Semangatt pak ahok!!!!!!!!!!	3	24
Bersih	semangat, pak, ahok	3	15

Tahap selanjutnya yaitu melakukan proses perhitungan menggunakan algoritma *Naive Bayes Classifier* dengan menggunakan data komentar pada *YouTube Channel KompasTV #Serikat Pekerja Pertamina Tolak Ahok*. Data terbagi menjadi data *training* yang terdiri dari 8856 kata yang telah diketahui kategorinya masing masing terdiri dari 2866 kata negatif, 2094 kata positif, dan 3896 kata netral. Berupa 859 komentar dengan kategori positif 203 komentar, negatif 284 komentar, netral 370 komentar. Data uji merupakan sebuah komentar yang belum diketahui kategorinya. Berikutnya akan diberikan contoh penggunaan algoritma *Naive bayes* untuk klasifikasi sentimen. Misalkan untuk komentar berikut

[“semangatt pak ahok!!!!!!!!!!”], yang sudah melalui analisis data.

Hasil perhitungan *Naive Bayes* untuk komentar tersebut seperti terlihat pada tabel 2:

**Tabel 2. Contoh Frekuensi Kata Uji Pada Data Latih**

No	Kata	Training negatif	Training positif	Training netral
1	semangat	0	8	0
2	Pak	0	2	0
3	Ahok	0	3	0

Dari tabel di atas diketahui: Jumlah *Term Frekuensi (TF)* keseluruhan dari  $N_k$  di setiap kategori. Dan dari data latih didapatkan ( $nc | \text{Positif}$ ) = 2099, ( $nc | \text{Netral}$ ) = 3896), ( $nc | \text{Negatif}$ ) = 2866) dan Jumlah Kosa kata Training = 2023.

Dari nilai-nilai tersebut, dapat dicari nilai nilai probabilitas keyword dengan menggunakan rumus  $P(X_i|V_j)$  dan probabilitas kategori dokumen  $P(V_j)$  yaitu:

$$P(V_j | \text{Positif}) = \frac{203}{859} = 0.23632130384167638$$

$$P(V_j | \text{Negatif}) = \frac{284}{859} = 0.33061699650756693$$

$$P(V_j | \text{Netral}) = \frac{370}{859} = 0.4307334109429569$$

$$P(\text{semangat} | \text{positif}) = \frac{(8+1)}{(2094+2023)} = 0.002183406113537118$$

$$P(\text{semangat} | \text{negatif}) = \frac{(0+1)}{(2866+2023)} = 2.0454080589077522E-4$$

$$P(\text{semangat} | \text{netral}) = \frac{(0+1)}{(3896+2023)} = 1.6894745734076703E-4$$

Hasil dari perhitungan  $P(X_i|V_j)$  kategori positif dapat dilihat pada tabel 3:

**Tabel 3. Contoh Hasil Perhitungan Probabilitas Kata**

Kata	$N_k$	$P(X_i V_j)$ positif
Semangat	8	0.002183406113537118
Pak	2	7.27802037845706E-4
ahok	3	9.715812484819043E-4

Berdasarkan nilai probabilitas pada tabel dapat dihitung nilai  $V_{map}$ :

$$V_{MAP} = \arg \max_{v_j \in V} P(O) \prod (I) = (0.23632130384167638)^*$$

$$(0.002183406113537118) * (7.27802037845706E-4) = 3.755352107474374E-7$$

Hasil akhir perhitungan *Naive bayes* untuk contoh komentar di atas dapat dilihat pada tabel 4.

**Tabel 4. Hasil Akhir Klasifikasi**

Komentar	Semangatt
Training Positif	3.755352107474374E-7
Training Negatif	1.3832003867221386E-8
Training Netral	1.2294528564035867E-8
Klasifikasi	positif

Perhitungan pada Tabel 4 memperlihatkan proses penentuan klasifikasi sentimen didasari oleh hasil perhitungan probabilitas dengan nilai tertinggi yaitu  $3.755352107474374E-7$ , sehingga komentar diatas dikategorikan sebagai komentar “Positif”. Langkah-langkah perhitungan di atas kemudian diterapkan pada pengembangan aplikasi sentimen analisis sederhana berbasis desktop.

### 3.2 Hasil Klasifikasi *Naive Bayes*

Tahapan klasifikasi memanfaatkan algoritma *naive bayes* dengan menggunakan 300 data *testing* sementara data training yaitu sebanyak 1000. Berikut hasil klasifikasi memanfaatkan *software rapiminder*.

1. Hasil *apply model*

Perubahan terjadi ketika permodelan pada data *training* di implementasi pada data *testing*. Perubahan pada kolom sentimen dilihat dengan membandingkan nilai *confidence* tertinggi antara kolom *confidence* negatif dan positif.

2. Hasil akurasi dengan *rapidminer*

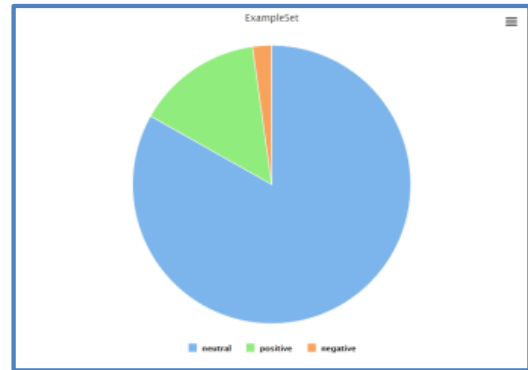
accuracy: 98.00%				
	true neutral	true negative	true positive	class precision
pred. neutral	92	0	0	100.00%
pred. negative	1	2	0	66.67%
pred. positive	1	0	4	80.00%
class recall	97.87%	100.00%	100.00%	

Gambar 3. Hasil akurasi dengan *Rapidminer*

Dari gambar 3 bisa diketahui tingkat akurasi yaitu sebesar 98.00%. Dari beberapa data yang telah di uji, terdapat beberapa data yang diprediksi positif dan data yang diprediksi negatif. Dari kelas presisi, prediksi positif memiliki nilai kelas sebesar 80.00% dan prediksi negatif memiliki nilai kelas presisi sebesar 66.67%. Lalu dari kelas *recall*, prediksi positif memiliki nilai kelas sebesar 100.00% dan prediksi negatif juga memiliki nilai kelas *recall* sebesar 100.00%.

3.3 Visualisasi Data

Visualisasi pada data yaitu penyajian dalam data yaitu bentuk yang lebih menarik dan sederhana. Hal ini dilakukan agar memudahkan pembaca dalam memahami secara langsung informasi dari data yang sudah di analisis. Pada penelitian visualisasi ini data dilakukan dengan penyajian yaitu data dalam bentuk *pie chart* pada gambar 4.



Gambar 3.11 Visualisasi Data

Adapun sentimen yang diberikan oleh pengguna merupakan penyampaian pendapat terhadap serikat pekerja Pertamina tolak Ahok. Hal ini dapat dilihat dari pemberian pujian ataupun dorongan (Seperti 'Semangat', 'Pintar'), penyampaian kelebihan ('Paling baik, 'Anti korupsi'), maupun ulasan secara langsung ('Sangat Puas', 'Bangga'). Analisis yang dikembangkan dapat melakukan pengklasifikasian atau klasifikasi data komentar yang mengandung kata sentimen yang bersifat positif, negatif, maupun netral.

4. SIMPULAN

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

1. Sentimen yang menjadi fokus penelitian yaitu sentimen yang terbagi menjadi positif dan negatif. Adapun, sentimen positif dapat ditentukan berdasarkan kata-kata yang memiliki konotasi baik dalam Bahasa Indonesia. Demikian pula pada sentimen negatif yang didapat berdasarkan kata-kata yang berkonotasi buruk pada bahasa Indonesia.
2. Metode analisis *naive bayes classifier* dapat digunakan sebagai metode analisis sentimen yang menghasilkan tiga *class* sentimen



yaitu positif, negatif, dan neutral dari data awal yang telah proses *transform case* (perubahan huruf *capital* menjadi huruf kecil), *tokenize* (pemisahan kalimat menjadi kata-kata), pemfilteran kata-kata berdasarkan panjang kata, pemfilteran kata *stopword*, serta *stem* (menjadikan kata berimbuhan sebagai kata dasar).

3. Penelitian yang dibuat berhasil digunakan dalam menganalisis sentimen serikat pekerja pertamina tolak ahok pada media sosial *YouTube* melalui metode *Naive Bayes Classifier*, baik sentimen yang positif maupun negatif.

Wahidmurni W. (2017). *PEMAPARAN METODE PENELITIAN KUANTITATIF*.

#### DAFTAR RUJUKAN

- Arief, R., & Imanuel, K. (2019). ANALISIS SENTIMEN TOPIK VIRAL DESA PENARI PADA MEDIA SOSIAL TWITTER DENGAN METODE LEXICON BASED. *Jurnal Ilmiah Matrik*, 21(3), 242–250. <https://doi.org/10.33557/jurnalmatrik.v21i3.727>
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques* (2nd ed.). Morgan Kaufmann Publishers.
- Hartoko, A. (2019). *Rahasia cepat tenar dan dapat duit lewat youtube*. PT Elex Media Komputindo.
- Mahbubah, L., & Zuliarso, E. (2019). ANALISA SENTIMEN TWITTER PADA PILPRES 2019 MENGGUNAKAN ALGORITMA NAIVE BAYES. *SINTAK*, 3. <https://www.unisbank.ac.id/ojs/index.php/sintak/article/view/7585>
- Purbo, O. W. (2019). *TEXT MINING ANALISIS MEDSOS KEKUATAN BRAND & INTELEJEN DI INTERNET* (A. A. Christian, Ed.; I,1st Publ). ANDI (Anggota IKAPI).
- Suntoro, J. (2019). *Data mining algoritma dan implementasi dengan pemrograman php*. PT Elex Media Komputindo.
- Suyanto. (2017). *DATA MINING UNTUK KLASIFIKASI DAN KLASTERISASI DATA*. INFORMATIKA Bandung.