

# Sentimen Analisis Masyarakat Tentang Penetapan Hari Raya Idul Adha Tahun 2023 Pada Video Youtube Menggunakan Algoritma Random Forest dan Support Vector Machine

Rahmad Firdaus<sup>1</sup>, Rafiq Al Hariri<sup>2</sup>, Hasanatul Fu'adah Amran<sup>3</sup>  
<sup>1,2,3</sup>Teknik Informatika, Ilmu Komputer, Universitas Muhammadiyah Riau  
[rahmadfirdaus@umri.ac.id](mailto:rahmadfirdaus@umri.ac.id), [180401139@student.umri.ac.id](mailto:180401139@student.umri.ac.id), [hasanatul@umri.ac.id](mailto:hasanatul@umri.ac.id)

## Abstract

The interaction of society on social media is reflected through the responses given in the comment section as feedback on various activities or phenomena that are happening, whether in the form of text, images, or videos. One example is the comments appearing on YouTube videos regarding the determination of Eid al-Adha. From the various responses in the comment section, there is information in the form of unstructured data. These comment data can be processed and analyzed by applying sentiment analysis using a classification approach. This study uses the Random Forest and Support Vector Machine algorithms to analyze the comment data. A total of 1049 comment data were obtained from 13 videos used in this study. This data is then divided into training and testing data with various percentage divisions, namely 60:40, 70:30, 80:20, and 90:10. The test results show that the Support Vector Machine algorithm achieves an accuracy of 96%, while the Random Forest algorithm achieves an accuracy of 94%. This indicates that the Support Vector Machine algorithm is better at classifying comments on videos, especially in the case of the determination of Eid al-Adha in 2023.

Keywords: sentiment analysis, idul adha, youtube, random forest, SVM

## Abstrak

Interaksi masyarakat di media sosial terlihat dari tanggapan yang diberikan pada kolom komentar sebagai umpan balik dari setiap kegiatan atau fenomena yang sedang terjadi. Mulai dari status yang berupa teks, gambar maupun video. Salah satunya adalah komentar pada video youtube tentang penetapan hari raya Idul Adha. Dari berbagai tanggapan pada kolom komentar tersebut diperoleh sebuah informasi dalam bentuk data tidak terstruktur. Data komentar tersebut dapat diolah dan dianalisis dengan menerapkan analisis sentimen menggunakan pendekatan klasifikasi. Penelitian ini menggunakan algoritma Random Forest dan Support Vector Machine. Penelitian ini menggunakan 1049 data komentar yang diperoleh dari 13 video. Penelitian ini menggunakan data *training* dan *testing* dengan presentase 60:40 70:30, 80:20, dan 90:10. Menghasilkan hasil pengujian menggunakan algoritma Support Vector Machine mendapatkan akurasi sebesar 96%, algoritma Random Forest mendapatkan akurasi 94%. Hasil tersebut menunjukkan bahwa algoritma Support Vector Machine lebih baik digunakan untuk mengklasifikasi komentar pada video terutama dalam kasus penetapan hari raya Idul Adha pada tahun 2023.

Kata kunci: analisis sentimen, idul adha, youtube, random forest, SVM

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International License

## 1. Pendahuluan

Saat ini, platform media sosial telah menjadi sarana komunikasi yang amat diminati di antara pengguna internet di Indonesia. Media sosial menjadi wadah untuk menyampaikan pendapat dan berekspresi mengenai berbagai topik, seperti halnya *youTube* [1]. YouTube merupakan sebuah *platform* media sosial yang menampilkan beragam video dengan durasi tertentu, dibuat oleh para pembuat konten (yang dikenal sebagai *youTubers*) dengan berbagai jenis kreativitas. Tingginya minat masyarakat dalam menggunakan *youTube* telah membuat *platform* ini semakin diminati oleh berbagai kalangan. *youTube* menempati posisi teratas sebagai *platform* media sosial yang paling populer di Indonesia. Sebagai salah satu *platform streaming* video terkemuka di dunia, *youtube* menawarkan berbagai macam konten hiburan dan pendidikan. Selain itu, *youtube* juga dimanfaatkan sebagai media untuk menyebarkan berita dan informasi. [2]. Berita terbaru mengenai perbedaan dalam penentuan hari raya Idul Adha tahun 2023

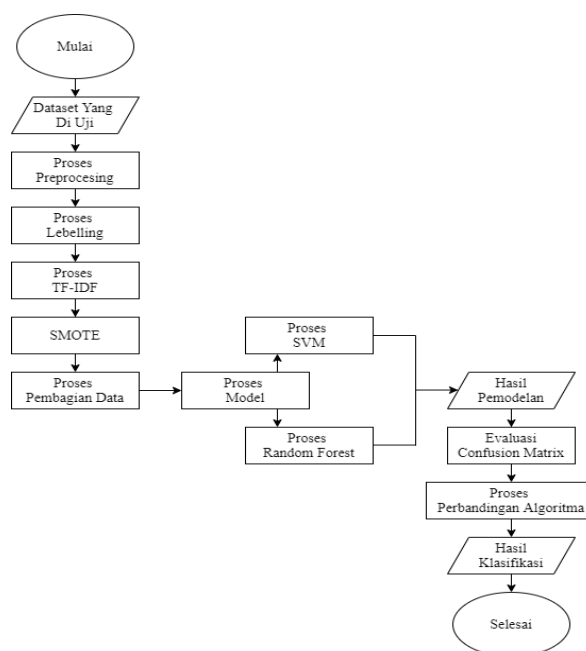
menjadi topik kontroversial di kalangan para ulama. Perbedaan ini kadang-kadang memicu ketegangan dan gangguan terhadap hubungan keislaman yang harmonis. Fenomena perbedaan seperti ini memang umum terjadi di Indonesia, di mana Muhammadiyah cenderung memilih metode *hisab* sebagai acuan, sementara Nahdlatul Ulama lebih condong kepada metode *rukyyatul hilal*. Pemerintah telah berusaha menyatukan kedua pandangan ini melalui metode *imkanurrukyah*. Meskipun telah diadakan sidang *Isbat* yang dihadiri oleh semua pihak terkait, termasuk organisasi Islam, untuk memfasilitasi kesatuan pendapat, namun perbedaan dalam penentuan tetap muncul, dengan setiap organisasi tetap mengeluarkan keputusan sesuai pandangannya masing-masing [3]. Salah satu cara untuk melihat tanggapan masyarakat terhadap perbedaan tersebut adalah melalui kolom komentar di video *youtube*. Di sana, kita dapat menemukan beragam pendapat, baik yang positif maupun negatif, yang mencerminkan respons yang beragam. Semakin banyak komentar yang muncul,

semakin banyak pula pengguna yang berinteraksi. Namun, situasi ini juga dapat menimbulkan interaksi yang tidak terduga, karena setiap individu memiliki sudut pandang dan pengalaman yang berbeda-beda. [4]. Analisis sentimen adalah proses untuk memahami data teks dengan mengelompokkan pendapat menjadi ulasan positif dan negatif. Ini merupakan bagian dari text mining yang bertujuan untuk mengidentifikasi pendapat atau kecenderungan opini seseorang atau kelompok terhadap suatu masalah atau objek. Analisis sentimen, atau opinion mining, melibatkan pemahaman, ekstraksi, dan pengolahan otomatis data teks untuk mendapatkan informasi tentang sentimen yang terdapat dalam suatu kalimat opini [5]. *Text mining* adalah analisis teks yang memanfaatkan dokumen sebagai sumber data, dengan tujuan mengidentifikasi kata-kata yang merepresentasikan isi dokumen tersebut. Hal ini memungkinkan untuk melakukan analisis keterhubungan, keterkaitan, dan klasifikasi antar dokumen. Teknik ini merupakan variasi dari data mining yang bertujuan untuk menemukan pola menarik dari kumpulan data teks besar. Salah satu aplikasi utamanya adalah klasifikasi, di mana text mining dapat digunakan untuk mengkategorikan dokumen berdasarkan pola-pola yang terdapat di dalamnya [5]. *Random Forest* adalah salah satu algoritma klasifikasi yang telah banyak digunakan untuk menganalisis polaritas komentar pelanggan secara otomatis. Algoritma ini merupakan teknik pembelajaran ensemble yang berbasis pada algoritma *Decision Tree* (Pohon Keputusan), di mana terdiri dari beberapa pohon keputusan sebagai *classifier*. Kelas yang ditentukan dalam proses klasifikasi diambil dari keputusan mayoritas yang dihasilkan oleh pohon-pohon keputusan yang ada. Melalui proses *voting* keputusan dari berbagai pohon keputusan, *Random Forest* mampu meningkatkan akurasi klasifikasinya. Dikarenakan performa kumpulan pohon keputusan yang dikombinasikan, *Random Forest* dikenal sebagai salah satu algoritma klasifikasi dengan tingkat akurasi yang baik. [6]. *Support Vector Machine* (SVM) adalah teknik *machine learning* yang populer untuk pengklasifikasi teks, dan sering kali menunjukkan performa yang baik di berbagai domain. SVM mampu mengidentifikasi *hyperplane* yang memisahkan dua kelas yang berbeda dengan optimal. Hal ini dilakukan dengan mencari garis pembatas (*decision boundary*) yang memisahkan antara dua kelas tersebut. SVM mencari nilai *hyperplane* dengan memanfaatkan *support vector* dan nilai margin. *Support vector* adalah titik data yang paling dekat dengan *hyperplane*, sedangkan margin adalah jarak antara *hyperplane* dan *support vector*. Dengan memaksimalkan margin, SVM dapat menemukan *hyperplane* yang optimal untuk memisahkan kelas-kelas tersebut [7]. Terdapat beberapa penelitian yang telah dilakukan oleh para peneliti sebelumnya dengan judul Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma *Naive Bayes*, *Random Forest* Dan *Support*

*Vector Machine* oleh [7]. Didapatkan hasil bahwa pada pengujian nilai akurasi tertinggi yaitu ada pada pengujian dengan menggunakan model algoritma *Random Forest* dengan jumlah akurasi sebesar 97,16% serta nilai AUC 0,996 kemudian disusul algoritma *Support Vector Machine* yang menghasilkan akurasi sebesar 96,01% dengan nilai AUC sebesar 0,543 dan nilai akurasi terendah ada pada pengujian menggunakan algoritma *Naive Bayes* dengan nilai akurasi sebesar 94,16% serta nilai AUC sebesar 0,999. Penelitian lain dengan judul Perbandingan Metode *Random Forest* dan *Support Vector Machine* Pada Analisis Sentimen *Twitter* oleh [8]. Hasil klasifikasi menunjukkan bahwa algoritma *Random Forest* memperoleh nilai akurasi sebesar 94%, sedangkan algoritma *Support Vector Machine* (SVM) mendapatkan nilai akurasi sebesar 93%. Tujuan dari penelitian ini adalah untuk mengevaluasi dan membandingkan performa antara algoritma *Random Forest* dan *Support Vector Machine* dalam mengklasifikasi sentimen masyarakat terhadap video di YouTube yang berkaitan dengan penetapan Hari Raya Idul Adha tahun 2023.

## 2. Metode Penelitian

Berikut tahapan-tahapan yang dilakukan pada penelitian ini :



Gambar 1. Metode Penelitian

### 2.1. Pengumpulan Data (*Dataset*)

Tahapan pengumpulan data dimulai dengan menggunakan teknik *Scraping* untuk mengambil data dari video-video youtube. Data ini biasanya disimpan dalam format file JSON. Selanjutnya, file JSON tersebut dikonversi menjadi format Excel untuk kemudian diolah lebih lanjut.

Setelah itu, data komentar dari berbagai video digabungkan menjadi satu file. Dalam proses penggabungan ini, beberapa kolom yang dianggap tidak diperlukan dapat dihapus untuk menyederhanakan struktur data. Setelah tahap tersebut selesai, file tersebut dikonversi menjadi format CSV, yang merupakan format yang umum digunakan untuk menyimpan data tabular.

## 2.2. Pemrosesan Data (*Data Preprocessing*)

Proses *preprocessing* data merupakan proses mengubah suatu dokumen yang tidak terstruktur menjadi menjadi data yang terstruktur untuk kebutuhan analisis agar tidak mengurangi performa dan pola data yang sesuai dengan kebutuhan sistem. Terdapat beberapa tahapan yang dilalui pada tahapan ini :

Pertama, tahapan *case folding* merupakan tahapan mengkonversikan keseluruhan teks menjadi bentuk standar, yaitu huruf kecil. Jadi, pada tahapan ini dilakukan untuk membuat kalimat menjadi huruf kecil semua.

Kedua, tahapan *cleansing* dilakukan untuk penghapusan karakter dalam *string* yang bukan bagian dari alfabet dan tidak memiliki arti tertentu guna mengurangi karakter yang tidak dibutuhkan. Jadi, tahapan ini dilakukan untuk membersihkan kalimat yang memiliki angka, hastag, emoji, link, tanda baca, dan lain sebagainya.

Ketiga, tahapan *tokenize* pada data teks merupakan tahap untuk memecah sekumpulan kalimat atau *string* menjadi potongan karakter atau kata.

Keempat, melakukan normalisasi dengan cara mengubah kalimat teks menjadi teks yang secara lengkap menampilkan cara pengucapannya.

Kelima, melakukan *stopword* yaitu kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak perlu. Seperti kata konjungsi “yang”, “dan”, “dari”, “di”.

Keenam, tahapan *stemming* merupakan proses menghilangkan kata yang memiliki imbuhan sesuai dengan KBBI sehingga membentuk suatu kata menjadi kata dasar.

## 2.3. Pelabelan Data (*Labelling*)

Tahapan pelabelan data dilakukan untuk membedakan ulasan yang mengandung nilai positif dan negatif. Ulasan positif dilambangkan dengan angka 1 dan ulasan negatif dilambangkan dengan angka -1.

## 2.4. Pembobotan Kata (TF-IDF)

Tahapan pembobotan kata dilakukan dengan menghitung berapa banyak frekuensi kemunculan setiap kata didalam dokumen, IDF (*word*) adalah nilai IDF dari setiap kata yang akan di cari, TF adalah jumlah keseluruhan dokumen yang ada, DF adalah jumlah kemunculan kata pada semua dokumen.

## 2.5. SMOTE

Teknik Smote digunakan untuk mencegah terjadinya *imbalance* data terhadap dataset, sehingga mempengaruhi hasil dari akurasi, maka pada penelitian ini menggunakan teknik Smote untuk menyeimbangkan *dataset* dari hasil *accuracy*, *precision*, *recall* dan *F1-score*.

## 2.6. Pembagian Data

Pada tahapan ini dilakukan pembagian data *training* dan *testing*. Data *testing* adalah data yang digunakan untuk melatih algoritma dalam mencari model yang sesuai. Data *training* adalah untuk menguji performa dari model.

## 2.7. Proses Model

Proses pemodelan dengan mengimplementasikan algoritma *Random Forest* dan algoritma *Support Vector Machine*.

## 2.8. Hasil Pemodelan

Tahap ini bertujuan untuk memperlihatkan hasil dari model algoritma *Random Forest* dan *Support Vector Machine* yang telah diperoleh dari proses sebelumnya.

## 2.9. Evaluasi Model

Tahapan ini dilakukan untuk menguji performa model dari mesin yang telah dibangun, untuk menghitung akurasi dan mengidentifikasi klasifikasi algoritma *Random Forest* dan *Support Vector Machine* menggunakan *confusion matrix* dengan menghitung *accuracy*, *precision*, *recall* dan *F1-Score*.

Akurasi adalah seberapa baik model dapat melakukan klasifikasi data dengan benar, baik itu dalam menentukan kelas positif (*true positive*) maupun kelas negatif (*true negative*).

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} \times 100\% \quad (1)$$

Keterangan :

Tp = *True Positive*

Fp = *False Positive*

Tn = *True Negative*

Fn = *False Negative*

*Precision* adalah ukuran yang menggambarkan tingkat keakuratan antara data yang diprediksi sebagai positif (positif yang diminta) dengan hasil prediksi yang diberikan oleh model.

$$Precision = \frac{TP}{FP+TP} \times 100\% \quad (2)$$

*Recall* adalah menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi.

$$Recall = \frac{TP}{FN+TP} \times 100\% \quad (3)$$

*F1-score* adalah perbandingan rata-rata *precision* dan *recall* yang dibobotkan.

$$F1 - score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (4)$$

### 3. Hasil dan Pembahasan

#### 3.1. Pengumpulan Data

Dataset diperoleh dari komentar video tentang penetapan hari raya Idul Adha tahun 2023 di channel youtube tvOneNews, Muhammadiyah channel, Kompastv, CNN Indonesia, Liputan6, Metrotv, Kompas.com, Harian surya, Tribun jateng dengan jumlah 13 video dan diperoleh dataset sebanyak 1049 data komentar.

Tabel 1. Menggabungkan Data Komentar Youtube

No	Text	Time	Author
1	Muhammadiyah pk metode hisab, itupun dibolehkan. Artinya jauh2 sebelum kalender di cetak mereka sdh tau kapan idul adha, makanya dlm kalender tanggal dibikin merah yaitu tgl 29.	2 minggu yang lalu	@harnofadli9120
2	Klo gak salah ya idul adha itu mengikuti jamaah haji nah kan ibadah haji itu di mekah mekah itu kan negara arab ya nah klo pemerintah arab lebaran tanggal 28 terus knpa di Indonesia tanggal 29 apakah ada pelaksanaan ibadah haji selain di arab sehingga pemerintah Indonesia bisa lebaran tangal 29 mohon pencerahanya.	2 minggu yang lalu	@aldiwaeto
3	Seharusnya mentri agama itu dipilih dari kalangan muhammadiyah, karena mayoritas agama di indonesia islam yang berkiblat pada makkah arab saudi, malah penjaga gereja di pilih jdi mentri agama. Yah salam	2 minggu yang lalu	@nuklir7277
....	....	...	...
1048	Bukankah mentaati pemerintah dalam hal yg ma'ruf itu sebuah kewajiban... Dan mendapat dosa jika tidak mentaatinya,dan lihatlah mudorotnya...	3 minggu yang lalu	@lintang.5166
1049	Urusan Agama, setiap kelompok punya	3 minggu	@qomarzaen2078

Patokan masing2 brow, Muhammadiyah selalu bersama dengan Mekkah, Idul adha ada hari arofah tgl 9 dan 10 solat idul adha brow,	yang lalu	
---	-----------	--

#### 3.2. Preprocessing Data

Pada tahapan ini akan dilakukan proses pembersihan data dengan menghilangkan karakter-karakter yang tidak perlu dan duplikasi data.

Tahapan *case folding* ini dilakukan untuk membuat kalimat menjadi huruf kecil semua. Berikut hasil data yang sudah melalui tahapan ini :

Tabel 2. Hasil Tahapan Case Folding

Text	Case Folding
Muhammadiyah pk metode hisab, itupun dibolehkan. Artinya jauh2 sebelum kalender di cetak mereka sdh tau kapan idul adha, makanya dlm kalender tanggal dibikin merah yaitu tgl 29.	muhammadiyah pk metode hisab, itupun dibolehkan. artinya jauh2 sebelum kalender di cetak mereka sdh tau kapan idul adha, makanya dlm kalender tanggal dibikin merah yaitu tgl 29.

Selanjutnya akan dilakukan tahapan *cleaning* dilakukan untuk membersihkan kalimat yang memiliki angka, hastag, emoji, link, tanda baca, dan lain sebagainya. Berikut hasilnya :

Tabel 3. Hasil Tahapan Cleaning

Text	Cleaning
Muhammadiyah pk metode hisab, itupun dibolehkan. Artinya jauh2 sebelum kalender di cetak mereka sdh tau kapan idul adha, makanya dlm kalender tanggal dibikin merah yaitu tgl 29.	muhammadiyah pk metode hisab itupun dibolehkan artinya jauh sebelum kalender di cetak mereka sdh tau kapan idul adha makanya dlm kalender tanggal dibikin merah yaitu tgl

Selanjutnya akan dilakukan tahapan *tokenize* pada data teks untuk memecah sekumpulan kalimat atau *string* menjadi potongan karakter atau kata. Berikut hasilnya:

Tabel 4. Hasil Tahapan Tokenize

Text	Tokenize
Muhammadiyah pk metode hisab, itupun dibolehkan. Artinya jauh2 sebelum kalender di cetak mereka sdh tau kapan idul adha, makanya dlm kalender tanggal dibikin merah yaitu tgl 29.	['muhammadiyah', 'pk', 'metode', 'hisab', 'itupun', 'dibolehkan', 'artinya', 'jauh', 'sebelum', 'kalender', 'di', 'cetak', 'mereka', 'sdh', 'tau', 'kapan', 'idul', 'adha', 'makanya', 'dlm', 'kalender', 'tanggal', 'dibikin', 'merah', 'yaitu', 'tgl']

Tahapan normalisasi dilakukan mengubah teks kalimat menjadi teks yang secara lengkap memperlihatkan cara pengucapannya. Berikut hasilnya :

Tabel 5. Hasil Tahapan Normalisasi

Text	Normalisasi
Muhammadiyah metode hisab, itupun dibolehkan. Artinya jauh2 sebelum kalender di cetak mereka sdh tau kapan idul adha, makanya dlm kalender tanggal dibikin merah yaitu tgl 29.	['muhammadiyah', 'metode', 'hisab', 'itupun', 'dibolehkan', 'artinya', 'jauh', 'sebelum', 'kalender', 'di', 'cetak', 'mereka', 'sudah', 'tahu', 'kapan', 'idul', 'adha', 'makanya', 'dalam', 'kalender', 'tanggal', 'dibikin', 'merah', 'yaitu', 'tanggal']

Tahapan *stopword* merupakan kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Seperti kata konjungsi “yang”, “dan”, “dari”, “di”. Berikut hasilnya :

Tabel 6. Hasil Tahapan *Stopward*

Text	Stopward
Muhammadiyah metode hisab, itupun dibolehkan. Artinya jauh2 sebelum kalender di cetak mereka sdh tau kapan idul adha, makanya dlm kalender tanggal dibikin merah yaitu tgl 29.	['muhammadiyah', 'metode', 'hisab', 'itupun', 'dibolehkan', 'kalender', 'cetak', 'idul', 'adha', 'kalender', 'tanggal', 'dibikin', 'merah', 'tanggal']

Tahapan *stemming* merupakan proses menghilangkan kata yang memiliki imbuhan sesuai dengan KBBI sehingga membentuk suatu kata menjadi kata dasar. Berikut hasilnya :

Tabel 7. Hasil Tahapan *Stemming*

Text	Stemming
Muhammadiyah metode hisab, itupun dibolehkan. Artinya jauh2 sebelum kalender di cetak mereka sdh tau kapan idul adha, makanya dlm kalender tanggal dibikin merah yaitu tgl 29.	['muhammadiyah', 'metode', 'hisab', 'itu', 'boleh', 'kalender', 'cetak', 'idul', 'adha', 'kalender', 'tanggal', 'bikin', 'merah', 'tanggal']

### 3.3. Labelling Data

Tabel 8. Hasil Labeling Data

Text	Count	Sentimen
apa iya lari dari pemerintah mati jahiliyah apa artinya religius tapi mati jahiliyah mari pikirkan itu sahal bin abdillah altustari w h berkata patuhilah pemerintah dalam hal pemberlakuan mata uang ukuran dan timbangan hukum haji salat jumat dua hari raya idul fitri dan idul adha dan jihad tafsir alqurthubi v dan abu hayyan dalam alba hr almuhith iii	-4	Negatif
klo gak salah ya idul adha itu mengikuti jamaah haji nah kan ibadah haji itu di mekah mekah itu kan negara arab ya nah klo memerintah arab lebaran tanggal terus knpa di indonesia tanggal apakah ada pelaksanaan ibadah haji selain di arab sehingga pemerintah indonesia bisa lebaran tanggal mohon pencerahanya.	1	Positif

### 3.4. Pembobotan TF-IDF

Berikut menampilkan hasil dari Perhitungan TF-IDF dengan contoh dokumen.

Dokumen 1 (D1): hari raya idul adha beda muhammadiyah dan pemerintah

Dokumen 2 (D2): idul adha beda pendapat

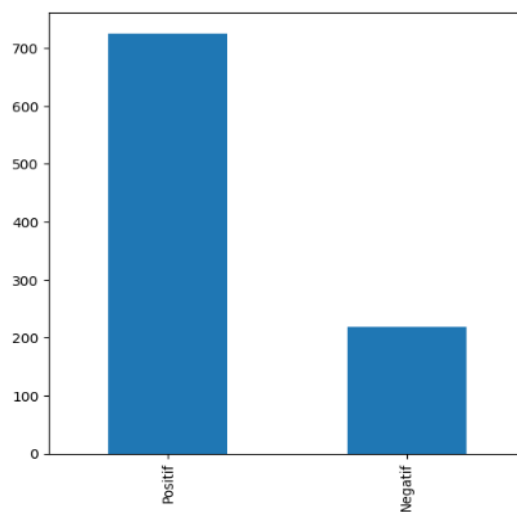
Dokumen 3 (D3): Muhammadiyah

Tabel 9. Hasil Perhitungan TF-IDF

No	Kata	TF			Df	D/d	IDF	w		
		D1	D2	D3				D1	D2	D3
1	hari	1	0	0	1	3	0.477	0.7	0	0
2	raya	1	0	0	1	3	0.477	0.7	0	0
3	idul	1	1	0	2	5	1.176	0.6	0	0
4	adha	1	1	0	2	5	1.176	0.6	0	0
5	beda	1	1	0	2	5	1.176	0.6	0	0
6	muhammadiyah	1	0	1	2	5	1.176	0.6	0	0.176
7	pemerintah	1	0	0	1	3	0.477	0.7	0	0
8	pendapat	0	1	0	1	3	0.477	0	0.176	0

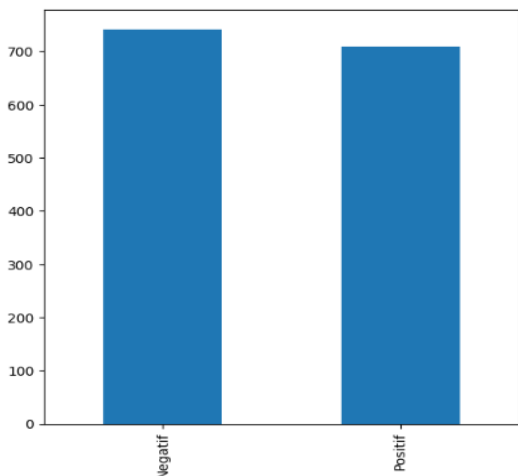
### 3.5. SMOTE

Tahapan *Synthetic Minority Oversampling Technique* (SMOTE) berguna untuk melakukan penyeimbangan kata minoritas pada dataset yang akan berpengaruh terhadap hasil *accuracy*, *precision*, *recall* dan *F1 score* nya. Berikut hasil dari sebelum melakukan SMOTE dan sesudah melakukan SMOTE dalam bentuk diagram batang :



Gambar 2. Sebelum Melakukan SMOTE

Sebelum melakukan tahapan SMOTE data yang didapatkan tidak seimbang dapat dilihat pada Gambar 2, dengan ulasan positif berjumlah 726 dan ulasan negatif berjumlah 218.



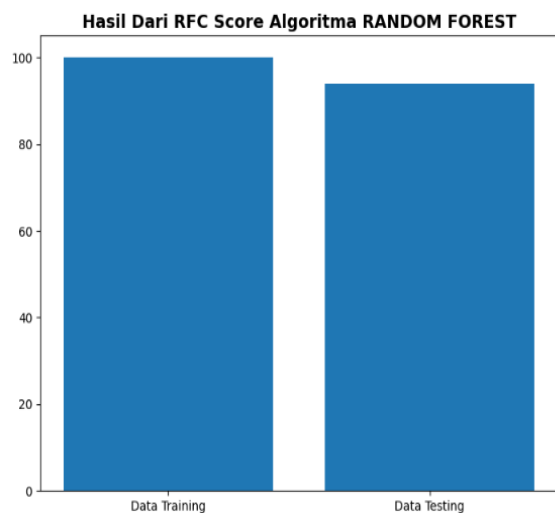
Gambar 3. Sesudah Melakukan SMOTE

Sesudah melakukan tahapan SMOTE data yang didapatkan sudah seimbang dapat dilihat pada Gambar 3, dengan ulasan positif berjumlah 710 dan ulasan negatif berjumlah 742.

### 3.6. Pembagian Data

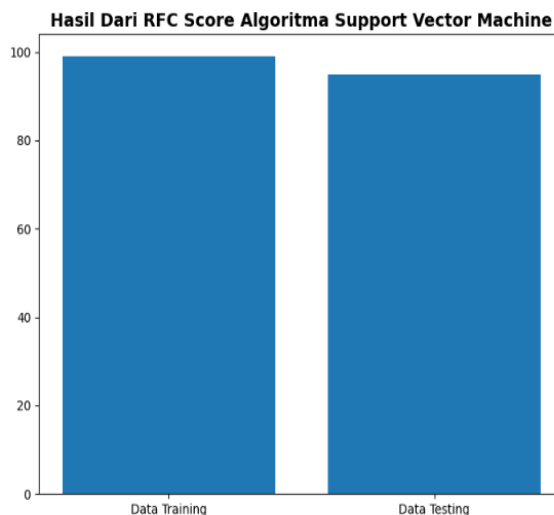
selanjutnya tahapan pembagian data dilakukan untuk mengevaluasi performa model dari algoritma *Random Forest* dan SVM, yang dibagi menjadi data *training* dan *testing*. Pembagian data dibagi menjadi data *training* dan *testing* dengan presentase 60:40 70:30, 80:20, dan 90:10.

### 3.7. Hasil Pemodelan



Gambar 4. Hasil Model Klasifikasi Algoritma *Random Forest*

Pada Gambar 4, hasil model klasifikasi algoritma *Random Forest* dalam bentuk diagram batang mendapatkan nilai untuk data *training* 100% dan data *testing* 94%.



Gambar 5. Hasil Model Klasifikasi Algoritma *Support Vector Machine*

Pada Gambar 5, hasil model klasifikasi algoritma *Support vector machine* dalam bentuk diagram batang mendapatkan nilai untuk data *training* 99% dan data *testing* 95%.

### 3.8. Hasil Pengujian

Berdasarkan proses model klasifikasi menggunakan algoritma *Random Forest* dan *Support Vector Machine*, dilakukan pengujian sebanyak 4 kali dengan presentase berbeda :

Tabel 10. Pengujian Algoritma *Random Forest*

No	Presentase Pengujian	Accurac y	Precisio n	Recall	F1- Score
1	90 : 10	94%	95%	96%	95%
2	80 : 20	93%	93%	95%	93%
3	70 : 30	91%	89%	94%	92%
4	60 : 40	91%	90%	94%	92%

Dari Tabel 10 maka penelitian ini menggunakan pembagian data menjadi 90% data *training* dan 10% data *testing* karena mendapatkan hasil paling tinggi dalam menghasilkan nilai Akurasi, Presisi, *Recall*, F1 Score dari algoritma *Random Forest*.

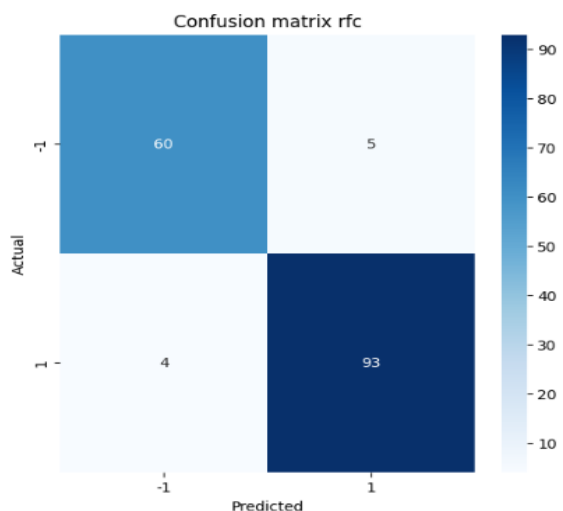
Tabel 11. Pengujian Algoritma *Support Vector Machine*

No	Presentase Pengujian	Accura cy	Precisi on	Reca ll	F1- Scor e
1	90 : 10	96%	97%	96%	96%
2	80 : 20	94%	94%	96%	95%
3	70 : 30	94%	92%	97%	94%
4	60 : 40	93%	92%	92%	94%

Dari Tabel 11 maka penelitian ini menggunakan pembagian data menjadi 90% data *training* dan 10% data *testing* karena mendapatkan hasil paling tinggi dalam menghasilkan nilai *Accurasy*, Presisi, *Recall*, F1 Score dari algoritma SVM.

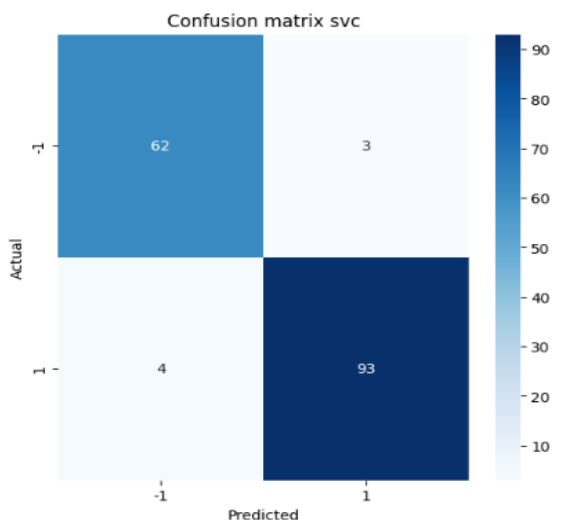
### 3.7. Evaluasi Model

Terdapat dua kelas yang menjadi *output* penelitian ini sehingga pengujian dengan confusion matrix menggunakan skala 2x2. Berikut ini merupakan hasil perhitungan confusion matrix dari nilai *accuracy*, *recall*, *precision*, f1-score dan tampilan grafik confusion matrix



Gambar 6. Grafik Confusion Matrix Algoritma Random Forest

Dari hasil evaluasi pada algoritma Random Forest, dalam menganalisis sentimen komentar video di Youtube tentang penetapan hari raya idul adha tahun 2023. Random Forest menghasilkan tingkat *accuracy* 94% dengan hasil *precision* positif 95% negatif 94%, *recall* positif 96% negatif 92%, dan *F1 score* positif 95% negatif 93%.



Gambar 7. Grafik Confusion Matrix Algoritma Support Vector Machine

Dari hasil evaluasi pada algoritma Support Vector Machine, dalam menganalisis sentimen komentar video di Youtube tentang penetapan hari raya idul adha tahun 2023. *Support Vector Machine* menghasilkan tingkat *accuracy* 96% dengan hasil *precision* positif

97% negatif 94%, *recall* positif 96% negatif 95%, dan *F1 score* positif 96% negatif 95%.

Berikut ini adalah perbandingan hasil evaluasi model kedua algoritma :

Tabel 12. Perbandingan Performa Algoritma

Algoritma	Accuracy	Recall	Precision	F1 Score
Random Forest	94%	95%	96%	95%
SVM	96%	97%	96%	96%

## 4. Kesimpulan

1. Hasil klasifikasi sentimen tertinggi terdapat pada ulasan positif berjumlah 807, dan untuk ulasan negatif berjumlah 242.
2. Pembagian data *training* dan data *testing* dengan akurasi tertinggi adalah 90:10, dengan menerapkan algoritma Support Vector Machine menghasilkan nilai *Accuracy* 96%, *Recall* 97%, *Precision* 96% dan *F1 - Score* 96% dan algoritma Random Forest menghasilkan nilai *Accuracy* 94%, *Recall* 95 %, *Precision* 96% dan *F1 - Score* 95%.
3. Algoritma Support Vector Machine lebih baik digunakan untuk mengklasifikasi komentar pada video youtube terutama dalam kasus penetapan hari raya Idul Adha pada tahun 2023 karena menghasilkan nilai *accuracy* sebesar 96%.

## Daftar Rujukan

- [1] M. S. P. Putra, Sri Dharma Wati, and Imam Solikin, "Analisis Sentimen Serikat Pekerja Pertamina Tolak Ahok Pada Media Sosial Youtube Menggunakan Algoritma Naive Bayes," *J. Ilm. Betrik*, vol. 12, no. 2, pp. 99–105, 2021, doi: 10.36050/betrik.v12i2.219.
- [2] H. A. R. Harpizon, R. Kurniawan, Iwan Iskandar, R. Salambue, E. Budianita, and F. Syafria, "Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naive Bayes," *JNKTI (Jurnal Nas. Komputasi dan Teknol. Informasi)*, vol. 5, no. 1, pp. 131–140, 2022, [Online]. Available: <http://repository.uin-suska.ac.id/59746/>.
- [3] I. Kamlludin, "ELFALAKY : Pandangan Mui terkait perbedaan penetapan 1 syawal 1444 H DI INDONESIA," *J. Ilmu Falak*, vol. 3, no. 2, p. 96, 2019.
- [4] F. Dona, I. Maulida, and B. Nugraha, "Klasifikasi Dan Analisis Dataset Komentar Video Youtube Menggunakan Supervised Learning," *Seminastika*, vol. 3, no. 1, pp. 86–90, 2021, doi: 10.47002/seminastika.v3i1.232.
- [5] M. Rezki, D. N. Kholifah, M. Faisal, P. Priyono, and R. Suryadithia, "Analisis Review Pengguna Google Meet dan Zoom Cloud Meeting Menggunakan Algoritma Naive Bayes," *J. Infortech*, vol. 2, no. 2, pp. 264–270, 2020, doi: 10.31294/infortech.v2i2.9286.
- [6] B. B. Baskoro, I. Susanto, and S. Khomsah, "Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR)," *INISTA (Journal Informatics Inf. Syst. Softw. Eng. Appl.)*, vol. 3, no. 2, pp. 21–29, 2021, doi: 10.20895/INISTA.V3.
- [7] E. Fitri, "Sentiment Analysis of the Ruanguru Application Using Naive Bayes, Random Forest and Support Vector Machine Algorithms," *J. Transform.*, vol. 18, no. 1, p. 71, 2020.

[8] Sudianto, P. Wahyuningtias, H. W. Utami, U. A. Raihan, and H. N. Hanifah, "Comparison Of Random Forest And Support Vector Machine Methods On Twitter Sentiment Analysis ( Case Study : Internet Selebgram Rachel Vennya Escape From

Quarantine ) Perbandingan Metode Random Forest Dan Support Vector Machine Pada Analisis Sentimen Twitt," *Jutif*, vol. 3, no. 1, pp. 141–145, 2022.