

# Analisis Sentimen Terkait RUU Perampasan Aset dengan Support Vector Machine

Lutfi Rofiqi\*<sup>1</sup>, Mutaqin Akbar<sup>2</sup>

<sup>1,2</sup>Fakultas Teknologi Informasi, Universitas Mercu Buana Yogyakarta  
e-mail: \*<sup>1</sup>rofiqilutfi50@gmail.com, <sup>2</sup>mutaqin@mercubuana-yogya.ac.id

*Abstract - This research aims to classify public sentiment towards the Asset Confiscation Bill using the Support Vector Machine (SVM) method. Data was collected from social media Twitter and YouTube with Google Colab's Python tools and libraries, including nodejs and Tweet Harvest for Twitter and Google API Client for YouTube. Comment data in Indonesian was manually labeled as positive or negative using Microsoft Excel. Preprocessing stages include cleansing, case folding, tokenizing, filtering, stemming, and data normalization. The Term Frequency-Inverse Document Frequency (TF-IDF) method is used for word weighting, while SVM is used for classification with evaluation using a confusion matrix which measures accuracy, precision, recall and F1 score. SMOTE is used to balance data classes. The research results show that SVM is effective with an accuracy of 79.8%, F1 score 79.96%, precision 80.58%, and recall 79.8%. Sentiment analysis reveals that the majority of public sentiment towards the Asset Confiscation Bill is negative, reflecting dissatisfaction or distrust. This research proves that the SVM method provides a comprehensive and valid evaluation of processed data, providing reliable results for further analysis.*

*Keywords: Sentiment analysis, Support Vector Machine (SVM), Twitter, YouTube, TF-IDF, SMOTE, Asset Confiscation Bill, Confusion Matrix.*

**Abstrak - Penelitian ini bertujuan untuk mengklasifikasikan sentimen publik terhadap RUU Perampasan Aset menggunakan metode Support Vector Machine (SVM). Data dikumpulkan dari media sosial Twitter dan YouTube dengan alat dan library Python Google Colab, termasuk nodejs dan Tweet Harvest untuk Twitter serta Google API Client untuk YouTube. Data komentar dalam bahasa Indonesia dilabeli secara manual sebagai positif atau negatif menggunakan Microsoft Excel. Tahapan pra-proses mencakup *cleansing, case folding, tokenizing, filtering, stemming*, dan normalisasi data. Metode *Term Frequency-Inverse Document Frequency (TF-IDF)* digunakan untuk pembobotan kata, sementara SVM digunakan untuk klasifikasi dengan evaluasi menggunakan *confusion matrix* yang mengukur akurasi, presisi, *recall*, dan *F1 score*. SMOTE digunakan untuk menyeimbangkan kelas data. Hasil penelitian menunjukkan bahwa SVM efektif dengan akurasi 79.8%, *F1 score* 79.96%, presisi 80.58%, dan *recall* 79.8%. Analisis sentimen mengungkap mayoritas sentimen publik terhadap RUU Perampasan Aset adalah negatif, mencerminkan ketidakpuasan atau ketidakpercayaan. Penelitian ini membuktikan bahwa metode SVM memberikan evaluasi komprehensif dan valid terhadap data yang telah diproses, memberikan hasil yang dapat dipercaya untuk analisis lebih lanjut.**

**Kata Kunci: Sentimen analisis, Support Vector Machine (SVM), Twitter, YouTube, TF-IDF, SMOTE, RUU Perampasan Aset, Confusion Matrix.**

## I. PENDAHULUAN

Penyitaan dan perampasan aset hasil tindak pidana sudah lama diatur dalam sistem hukum Indonesia, termasuk dalam KUHP (Kitab Undang-Undang Hukum Pidana) tentang pidana tambahan [1]. Indonesia telah meratifikasi Konvensi PBB Melawan Korupsi 2003 menjadi UU Nomor 7 Tahun 2006. Isu perampasan aset telah menjadi fokus PPATK (Pusat Pelaporan Transaksi Keuangan) sejak 2008. Baru-baru ini, Mahfud MD mengungkap kasus korupsi besar yang kembali menyoroti isu ini [2]. Secara formal, RUU perampasan aset telah dimasukkan ke dalam 189 RUU Prolegnas 2015-2019 dan 248 RUU Prolegnas 2020-2024 [3]. Namun, pelaksanaannya tidak dapat dilakukan langsung di setiap negara, termasuk Indonesia [4].

RUU Perampasan Aset mendapat perhatian luas di media sosial, termasuk Twitter, yang menjadi platform penting untuk diskusi publik. Twitter, yang didirikan oleh Jack Dorsey, adalah jejaring sosial yang digunakan untuk mengirim pesan singkat yang disebut "tweet" [5]. Setelah diakuisisi oleh Elon Musk, terjadi banyak perubahan pada sistem, termasuk perubahan nama aplikasi dari Twitter menjadi X [6]. Perubahan ini

mempengaruhi dinamika diskusi dan interaksi pengguna di platform tersebut, termasuk bagaimana isu-isu hukum dan kebijakan publik dibicarakan.

Sejak peralihan kepemilikan, penggunaan Twitter sebagai alat komunikasi untuk isu-isu penting seperti RUU Perampasan Aset semakin meningkat. Pengguna Twitter aktif membahas pro dan kontra dari RUU ini, membagikan pandangan mereka dalam tweet-tweet yang tersebar luas. Analisis sentimen terhadap tweet-tweet ini dapat memberikan wawasan tentang bagaimana publik merespon kebijakan tersebut.

YouTube, platform daring lainnya, juga memainkan peran penting dalam diskusi publik. Didirikan pada Februari 2005 oleh tiga mantan karyawan PayPal, YouTube memungkinkan pengguna menyimpan, menonton, dan membagikan berbagai jenis video. Pengguna YouTube dapat mengunggah, menonton, mencari video, dan berinteraksi melalui kolom komentar [7], [8]. Video-video tentang RUU Perampasan Aset, baik yang mendukung maupun yang menentang, telah mengumpulkan banyak penonton dan komentar, menunjukkan tingginya minat publik terhadap isu ini.

Dalam menangani isu ini dibuatlah analisis sentimen. Analisis sentimen mempelajari cara menganalisis opini dan penilaian terhadap objek tertentu, seperti layanan, produk, atau isu. Implementasinya dapat dilakukan pada berbagai tingkatan, baik dalam dokumen maupun kalimat [9]. Untuk memahami sentimen publik secara lebih mendalam, teknik machine learning seperti Support Vector Machine (SVM) digunakan. SVM, yang berlandaskan pada teori pembelajaran statistik struktural, menunjukkan performa yang lebih unggul dibandingkan metode machine learning lainnya [10]. Dengan menggunakan ekstraksi fitur tf-idf dan algoritma SVM, sentimen publik terhadap RUU Perampasan Aset di Twitter dan YouTube dapat dianalisis secara sistematis.

Berdasarkan sejumlah artikel yang relevan dengan penelitian ini, metode Support Vector Machine (SVM) merupakan metode yang memiliki tingkat akurasi yang cukup baik dalam memprediksi. Penelitian ini bertujuan untuk merancang dan menerapkan sistem analisis sentimen terhadap RUU Perampasan Aset menggunakan algoritma SVM, serta mengevaluasi akurasinya. Manfaat dari penelitian ini termasuk memberikan informasi sentimen kepada pemangku kepentingan dan menjadi referensi bagi peneliti di masa mendatang. Batasan penelitian ini mencakup penggunaan Python, data dari Twitter dan YouTube dalam bahasa Indonesia, serta penggunaan algoritma SVM.

## II. PENELITIAN YANG TERKAIT

Setelah meninjau literatur, ditemukan beberapa artikel yang relevan dengan topik penelitian ini. Penelitian ini mengeksplorasi dampak media sosial terhadap perubahan perilaku dan pandangan masyarakat terhadap budaya, etika, dan norma, khususnya terkait opini seputar Covid-19. Studi ini menilai sentimen publik terhadap kebijakan pemerintah seperti lockdown dan pembatasan sosial. Menggunakan Support Vector Machine dan ekstraksi fitur tf-idf, hasilnya menunjukkan sentimen dapat diklasifikasikan sebagai positif (68,75%) dan negatif (31,25%), dengan akurasi 74%, presisi 75%, recall 92%, dan F1-Score 83%. Penelitian ini membantu memahami peran media sosial dalam membentuk opini dan sentimen masyarakat terhadap kebijakan pemerintah selama pandemi Covid-19 [11].

Twitter telah menjadi platform populer untuk diskusi mengenai pemindahan ibu kota Indonesia, yang sering kali kontroversial. Analisis sentimen menggunakan Support Vector Machine (SVM) diusulkan untuk mengklasifikasikan sentimen tweet-tweet terkait. Pengujian terhadap 1.236 tweet menunjukkan hasil yang baik, dengan akurasi 96,68%, presisi 95,82%, recall 94,04%, dan AUC 0,979 [12].

Program Kartu Prakerja yang diperkenalkan pada tahun 2020 menarik perhatian besar di Twitter. Penelitian ini menganalisis sentimen masyarakat terhadap program tersebut menggunakan Support Vector Machine (SVM). Data diperoleh dari Twitter dan dievaluasi menggunakan Confusion Matrix dengan dua kernel: linear dan RBF. Hasilnya menunjukkan bahwa kernel linear memiliki akurasi 98,67%, presisi 98%, recall 99%, dan F1-Score 98%, sedangkan kernel RBF memiliki akurasi 98,34%, presisi 97%, recall 98%, dan F1-Score 98%. Sentimen masyarakat terhadap Program Kartu Prakerja cenderung netral, dengan persentase netral sebesar 98,34% [13].

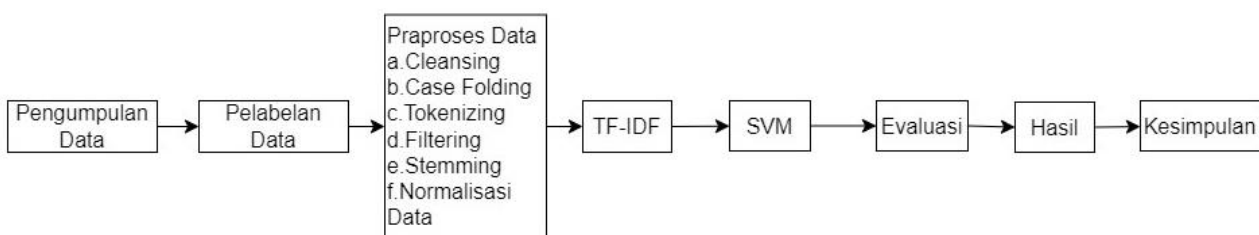
Perkembangan teknologi memfasilitasi akses informasi melalui Twitter, yang penting bagi maskapai untuk mengukur kepuasan pelanggan. Analisis sentimen digunakan untuk mengklasifikasikan opini sebagai positif, negatif, atau netral. Proses ini melibatkan preprocessing data, pembobotan kata dengan TF-IDF, dan algoritma SVM multi-class. Data opini dalam bahasa Inggris dari pengguna Twitter diuji, menunjukkan SVM kernel RBF

dengan parameter  $C = 10$  dan  $\gamma = 1$  mencapai akurasi 84,37%. Metode 10-fold cross validation meningkatkan akurasi hingga 80,41% [9].

Ulasan online restoran mempengaruhi reputasi dan membantu calon pelanggan mengevaluasi layanan. Restoran sering kesulitan menafsirkan banyak ulasan. Skripsi ini mengembangkan aplikasi untuk menganalisis ulasan restoran menggunakan text mining dan preprocessing. Data teks diubah menjadi vektor dengan TF-IDF dan dilatih dengan SVM. Ulasan berbahasa Indonesia digunakan sebagai data pelatihan. Hasil penelitian menunjukkan akurasi dan f1-score sebesar 93%. Penelitian menekankan pentingnya parameter TF-IDF yang optimal (min\_df 0,05, max\_df 0,75, norm l2, n-gram 1-2) dan penggunaan kernel SVM linear dengan C 1. Peningkatan jumlah dataset juga meningkatkan kinerja model SVM [14].

### III. METODE PENELITIAN

Tahapan penelitian yang dilakukan dimulai dari tahap pengumpulan data, pelabelan data, praproses data, proses *Term Frequency-Inverse Document Frequency* (TF-IDF), proses klasifikasi dengan *support vector machine* (SVM), evaluasi, penyajian hasil, dan penarikan kesimpulan. Tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Pengumpulan data dari platform media sosial Twitter dan YouTube dilakukan dengan menggunakan berbagai alat dan library milik Python Google Colab. Untuk pengumpulan data Twitter, digunakan *nodejs* dan *Tweet Harvest*. Sedangkan untuk pengumpulan data YouTube, digunakan library *Google API Client*. Data yang diambil berupa data komentar dan berbahasa Indonesia dari kedua platform tersebut. Proses pelabelan data dilakukan menggunakan *Microsoft Excel*. Data dilabeli secara manual: jika data tersebut memiliki kesan mendukung, maka akan dilabeli sebagai positif, sedangkan jika data tersebut mengandung kesan tidak percaya atau tidak setuju, maka akan dilabeli sebagai negatif.

Pada tahap praproses data ini, data teks yang telah dikumpulkan dari Twitter dan YouTube diproses untuk meningkatkan kualitas dan konsistensinya sebelum analisis lebih lanjut. Tahapan tersebut meliputi *cleansing*, *case folding*, *tokenizing*, *filtering*, *stemming*, dan normalisasi data. Preprocessing merupakan proses mengubah data teks mentah menjadi format yang lebih mudah difahami. Data mentah akan diproses untuk menyelesaikan masalah selama proses *preprocessing* [15].

Selanjutnya, *Term Frequency-Inverse Document Frequency* (TF-IDF) adalah metode yang digunakan untuk mengukur seberapa penting suatu istilah dalam dokumen [16]. Metode TF-IDF digunakan untuk proses pembobotan kata. Bobot ini digunakan untuk menilai tingkat kepentingan suatu kata dalam sebuah dokumen [17]. Pendekatan ini mengombinasikan dua prinsip untuk menghitung bobot yaitu frekuensi kemunculan suatu istilah dalam dokumen tertentu dan frekuensi kebalikannya (*invers*) dari dokumen yang berisi istilah tersebut [16]. TF-IDF terdiri dua komponen utama, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). *Term Frequency* (TF) berguna untuk menghitung seberapa sering istilah tertentu muncul dalam dokumen [16]. Maka, *Term Frequency* dirumuskan sebagai berikut [18]:

$$TF(t, d) = f(t, d) \quad (1)$$

Dimana,  $f$  adalah Frekuensi kemunculan,  $t$  merupakan *term* (Frasa/Kata), dan  $d$  adalah dokumen. Sedangkan, *Inverse Document Frequency* (IDF) berguna untuk Mengurangi bobot kata yang sering muncul dan meningkatkan bobot untuk kata yang jarang muncul [16]. Setelah itu, *Inverse Document Frequency* dirumuskan sebagai berikut [18]:

$$IDF = \log\left(\frac{N}{dft}\right) \quad (2)$$

Dimana, N adalah total dokumen yang berisi kata kunci, dan df(t) merupakan total dokumen yang berisi istilah (t).

SVM adalah kependekan dari *Support Vector Machine*, sebuah algoritma yang membagi dua kelompok kelas data menggunakan fungsi linear dalam ruang fitur yang memiliki dimensi yang tinggi. Proses ini mencari hyperplane terbaik untuk memisahkan kedua kelas, dengan tujuan menemukan margin maksimum antara ruang input dan ruang fitur menggunakan konsep kernel [19]. SVM juga merupakan salah satu metode yang digunakan untuk melakukan prediksi, baik dalam situasi klasifikasi ataupun regresi [20]. SVM terbukti sebagai teknik klasifikasi yang andal dalam pengenalan pola, menunjukkan tingkat keberhasilan tinggi di berbagai aplikasi [21]. Adapun rumus dalam bentuk sederhana dapat dilihat pada rumus dibawah [22]:

$$Y = w^T x_i + b \quad (3)$$

Dimana, Y merupakan nilai target dari dataset, yang dapat bernilai -1 atau 1, w merupakan vektor bobot yang berdimensi k (jumlah fitur), xi merupakan vektor baris yang berdimensi k (jumlah fitur) yang mewakili data ke-I, dan b merupakan bias atau kesalahan.

Langkah terakhir adalah evaluasi menggunakan metode *confusion matrix*, yang akan membantu menilai kinerja model dengan mengukur akurasi, presisi, *recall*, dan *F1-score* [10]. Proses evaluasi dilakukan dengan Python Google Colab.

#### IV. HASIL DAN PEMBAHASAN

Data penelitian ini dikumpulkan secara bertahap dari Twitter dan YouTube dengan kata kunci RUU Perampasan Aset menggunakan scraping di Google Colab dengan Python. Hasilnya diubah ke dalam bentuk .xlsx. Jumlah data yang terkumpul sebanyak 5000, terdiri dari 1434 data Twitter dan 3566 data YouTube. Contoh data hasil scraping dari Twitter dan Youtube seperti gambar berikut:

Total data in the DataFrame: 139

index	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url
0	1797428552310399274	Mon Jun 03 00:42:16 +0000 2024	18	FYI Prabowo Gibran dalam janji Politik nya tak sebut ada Revisi UU KPK dan UU perampasan Aset. Jadi jgn berharap pemberantasan KKN bisa berjalan dgn Baik. <a href="https://t.co/5atoICBin5">https://t.co/5atoICBin5</a>	1797428552310399274	<a href="https://pbs.twimg.com/ext_tw_video_thumb/1797428276794892288/pu/img/u5fXTwSi9PnCEM.jpg">https://pbs.twimg.com/ext_tw_video_thumb/1797428276794892288/pu/img/u5fXTwSi9PnCEM.jpg</a>
1	1801138050963304765	Thu Jun 13 06:22:29 +0000 2024	1326	Banteng Korena Menyala ayo cepat selesaikan UU perampasan aset <a href="https://t.co/IDacDvbSYb">https://t.co/IDacDvbSYb</a>	1801138050963304765	<a href="https://pbs.twimg.com/ext_tw_video_thumb/1801135984031264768/pu/img/HlztVVR-g-rov78NA.jpg">https://pbs.twimg.com/ext_tw_video_thumb/1801135984031264768/pu/img/HlztVVR-g-rov78NA.jpg</a>
2	1798703982325743794	Thu Jun 06 13:10:22 +0000 2024	20	Nih kaum 58 % jgn omon Omon undang undang Perampasan Aset..ya..knp simak deh biar kalian jgn sok teriak sok paling anti korupsi tapi dalam visi misinya kagak ada 01 ada Revisi UU KPK dan UU perampasan Aset 03 ada Revisi UU KPK dan UU perampasan Aset 02 Kagak Ada sama sekali <a href="https://t.co/4OY3wNnPig">https://t.co/4OY3wNnPig</a>	1798703982325743794	<a href="https://pbs.twimg.com/ext_tw_video_thumb/1798703921843892224/pu/img/0NcooE0olqMTfmvj.jpg">https://pbs.twimg.com/ext_tw_video_thumb/1798703921843892224/pu/img/0NcooE0olqMTfmvj.jpg</a>

Gambar 2. Pengumpulan Data Twitter

Total komentar: 91  
 Pasti takut  
 Ayo nitizen rakyat indonesia kita bubarkan dpr  
 Segerah saahkan RUU perampasan aset sebelum 17 AGUSTUS hari kemerdekaan Indonesia 🇮🇩🇮🇩🇮🇩🇮🇩🇮🇩🇮🇩🇮🇩  
 Up👍  
 Cepaaaat woiii.. sengaja di perlamaa  
 Mana mahasiswa🤔🤔🤔  
 Pajak dari uang rakyat, tidak pernah dinikmati rakyat, Pendidikan saja mahal. DPR MENYUARAKAN SUARA PARTAI, MANA KEADILAN?.  
 UANG RAKYAT DI NIKMATI OKNUM, DPR DIKENDALIKAN PARTAI. SUARAKAN KEADILAN "NKRI HARGAMATI".  
 Hh  
 UANG RAKYAT DI NIKMATI OKNUM, DPR DIKENDALIKAN PARTAI "NKRI HARGA MATI".

Gambar 3. Pengumpulan Data Youtube

Proses pelabelan data dilakukan secara manual dengan Microsoft Excel. Pelabelan data dibagi menjadi 2 jenis yaitu positif dan negatif. Seperti contoh tabel berikut:

TABEL I  
 Pelabelan Data

Kata	Nilai	Sentimen
------	-------	----------

uu rampas aset bagaimana	-1	negatif
harus saahkan uu rampas aset	1	positif

Setelah dilakukannya pengumpulan data selesai. Data tersebut masih berupa data tidak terstruktur dan perlu dibersihkan. Pada tahap awal pembersihan data dilakukan beberapa tahapan yang salah satunya bagian awal yaitu *cleansing*. *Cleansing* adalah langkah dalam pengolahan data teks yang bertujuan untuk menghapus komponen-komponen yang tidak relevan atau mengandung noise. Komponen yang sering dihapus mencakup karakter HTML, simbol emosi, hashtag (#), nama pengguna akun (@username), retweet (RT), tautan URL, dan alamat situs web. Contohnya seperti pada tabel dibawah ini:

TABEL II  
*Cleansing*

Sebelum di Cleansing	Sesudah di Cleansing
@psi_id UU PERAMPASAN ASET????	UU PERAMPASAN ASET
@Mdy_Asmara1701 Psi yg berani menyuarakan uu perampasan aset yg lain masuk angin	Psi yg berani menyuarakan uu perampasan aset yg lain masuk angin

Setelah tahap *cleansing* selesai maka dilakukanlah tahap *case folding*. *Case folding* adalah langkah standarisasi yang mengubah semua elemen dalam teks menjadi huruf kecil. Proses ini juga mencakup penghapusan karakter-karakter selain huruf yang dianggap tidak mempengaruhi proses pengolahan teks, seperti angka dan tanda baca. Contohnya seperti pada tabel dibawah ini:

TABEL III  
*Case Folding*

Sebelum di Case Folding	Sesudah di Case Folding
UU PERAMPASAN ASET	uu perampasan aset
Psi yg berani menyuarakan uu perampasan aset yg lain masuk angin	psi yg berani menyuarakan uu perampasan aset yg lain masuk angin

Setelah tahap *case folding* selesai maka dilakukanlah tahap *tokenizing*. *Tokenizing* adalah tahap di mana kalimat-kalimat dalam sebuah teks dipecah menjadi kata-kata tunggal atau token. Contohnya seperti pada tabel dibawah ini:

TABEL III  
*Tokenizing*

Sebelum di Tokenizing	Sesudah di Tokenizing
uu perampasan aset	UU,PERAMPASAN,ASET
psi yg berani menyuarakan uu perampasan aset yg lain masuk angin	Psi,yg,berani,menyuarakan,uu,perampasan,aset,yg,lain,masuk,angin

Setelah tahap *tokenizing* selesai maka dilakukanlah tahap *filtering*. *Filtering* adalah langkah dalam pengolahan data teks di mana kata-kata yang termasuk dalam kategori stopword, atau kata-kata yang dianggap tidak memberikan kontribusi signifikan dalam analisis sentimen, dihilangkan. Proses ini didukung oleh daftar stopword yang dipadankan dengan data teks yang ada. Contohnya seperti pada tabel dibawah ini:

TABEL IV  
*Filtering*

Sebelum di Filtering	Sesudah di Filtering
----------------------	----------------------

UU,PERAMPASAN,ASET	UU,PERAMPASAN,ASET
Psi,yg,berani,menyuarakan,uu,perampasan,aset,yg,lain,masuk,angin	Psi,yg,berani,menyuarakan,uu,perampasan,aset,yg,masuk,angin

Setelah tahap *filtering* selesai maka dilakukanlah tahap *stemming*. *Stemming* adalah langkah dalam pengolahan teks yang mengubah kata-kata dari bentuk variasi menjadi bentuk dasarnya sesuai struktur morfologi. Ini berarti kata-kata yang diubah tetap memiliki arti dan makna yang sama dengan kata dasarnya. Contohnya seperti pada tabel dibawah ini:

TABEL V  
*Stemming*

Sebelum di Stemming	Sesudah di Stemming
UU,PERAMPASAN,ASET	uu,ampas,aset
Psi,yg,berani,menyuarakan,uu,perampasan,aset,yg,masuk,angin	psi,yg,berani,suara,uu,ampas,aset,yg,masuk,angin

Dan yang terakhir adalah normalisasi data. Dimana proses normalisasi data dilakukan untuk menyesuaikan kata-kata dengan aturan yang ditemukan dalam Kamus Besar Bahasa Indonesia(KBBI) [15]. Normalisasi data bertujuan untuk meningkatkan konsistensi dan akurasi dalam analisis teks akurasi dalam analisis teks dengan mengubah kata-kata yang mungkin salah eja, disingkat, atau berbentuk slang menjadi bentuk baku yang diakui oleh KBBI. Hal ini membuat data lebih homogen dan representatif dari bahasa Indonesia yang benar, sehingga hasil analisis menjadi lebih valid dan dapat dipercaya. Contohnya seperti pada tabel dibawah ini:

TABEL VI  
Normalisasi Data

Sebelum di Normalisasi Data	Sesudah di Normalisasi Data
uu,ampas,aset	uu rampas aset
psi,yg,berani,suara,uu,ampas,aset,yg,masuk,angin	psi yang berani suara uu rampas aset yang masuk angin

Dalam proses perhitungan TF-IDF, digunakan dua komponen utama yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Berikut adalah langkah-langkah perhitungan menggunakan contoh tabel komentar dan sentimen:

TABEL VII  
Contoh Data

Komentar	Sentimen
harus sahan uu rampas aset	positif
kok uu rampas aset gadidukung	negatif
dorong uu rampas aset koruptor sah	positif

Dalam perhitungan TF mengukur frekuensi kemunculan kata dalam dokumen, sementara IDF mengukur kelangkaan kata di seluruh dokumen.  $IDF = \log(\text{Total jumlah dokumen} / \text{Jumlah dokumen yang mengandung term})$ . Seperti tabel berikut:

TABEL VIII  
TF dan IDF

Term	TF			IDF
	D1	D2	D3	
aset	0.2	0.2	0.166	0.0
dorong	0.000	0.000	0.166	1.098
gadidukung	0.000	0.2	0.000	1.098

harus	0.2	0.000	0.000	1.098
kok	0.000	0.2	0.000	1.098
koruptor	0.000	0.000	0.166	1.098
rampas	0.2	0.2	0.166	0.000
sah	0.000	0.000	0.166	1.098
sahkan	0.2	0.000	0.000	1.098
uu	0.2	0.2	0.166	0.000

Selanjutnya adalah hasil dari TF-IDF yang dimana diperoleh dari hasil perkalian antara TF dan IDF untuk setiap kata dalam setiap dokumen. Seperti table berikut:

TABEL IX  
Hasil TF-IDF

Term	TF-IDF		
	D1	D2	D3
aset	0.000	0.000	0.000
dorong	0.000	0.000	0.183
gadidukung	0.000	0.219	0.000
harus	0.219	0.000	0.000
kok	0.000	0.219	0.000
koruptor	0.000	0.000	0.183
rampas	0.000	0.000	0.000
sah	0.000	0.000	0.183
sahkan	0.219	0.000	0.000
uu	0.000	0.000	0.000

Proses uji menggunakan *Support Vector Machine* akan dilakukan dengan pembagian data menggunakan library sklearn, yaitu *ShuffleSplit*, yang membagi data menjadi 10 bagian, dengan 80% data train dan 20% data test pada setiap bagian. Data tarin sebanyak 4000 data dan data test sebanyak 1000. Maka diperoleh lah hasil seperti gambar berikut:

```
Banyak data train: (4000,)
Banyak data test: (1000,)
```

Gambar 4. Data Train dan Test

Untuk menyeimbangkan kelas minoritas pada data train, akan digunakan *library imblearn* dengan algoritma SMOTE yang melakukan *oversampling* dengan menambah sampel pada kelas minoritas [23]. Maka diperoleh lah hasil seperti gambar berikut:

```
Banyaknya data train yang di oversampling: (4876, 1000)
Banyaknya data test yang di oversampling: (1000, 1000)
```

Gambar 5. SMOTE

Setelah itu label diprediksi dengan library sklearn yaitu SVM. Selanjutnya menyimpan nilai-nilai metrik performa model untuk setiap *fold*. Nilai akurasi dari prediksi pada data test disimpan dalam list `accuracy`, nilai *F1 score* disimpan dalam list `f1`, dan *confusion matrix* disimpan dalam list `cms`. Berikutnya, Kita juga menampilkan laporan klasifikasi untuk *fold* saat ini menggunakan `classification_report`, yang memberikan informasi detail tentang *precision*, *recall*, *F1 score*, dan *support* untuk setiap kelas pada data test. Maka diperoleh lah hasil seperti gambar berikut:

```
precision  recall  f1-score  support
negatif    0.86    0.79    0.82     598
positif    0.72    0.82    0.76     402

accuracy                    0.80    1000
macro avg    0.79    0.80    0.79    1000
weighted avg 0.81    0.80    0.80    1000
```

Gambar 6. Tampilan Nilai Hasil

Kemudian, menghitung dan menampilkan rata-rata akurasi, F1 score, presisi dan recall setelah beberapa fold. Rata-rata akurasi dihitung dari list `accuracy` dan ditampilkan dalam persen, rata-rata F1 score yang dihitung dari list `f1`, rata-rata presisi yang dihitung dari list `precision` dan rata-rata recall yang dihitung dari list `recall`.

Maka diperoleh lah hasil seperti gambar berikut:

```
Average accuracy across fold: 79.80
Average f1 across fold: 79.96
Average precision across fold: 80.58
Average recall across fold: 79.80
```

Gambar 7. Tampilan Akurasi, F1 Score, Presisi, dan Recall

Gambaran jika ditampilkan dalam grafik untuk akurasi dan f1 score-nya:



Gambar 8. Tampilan Grafik Akurasi dan F1 Score

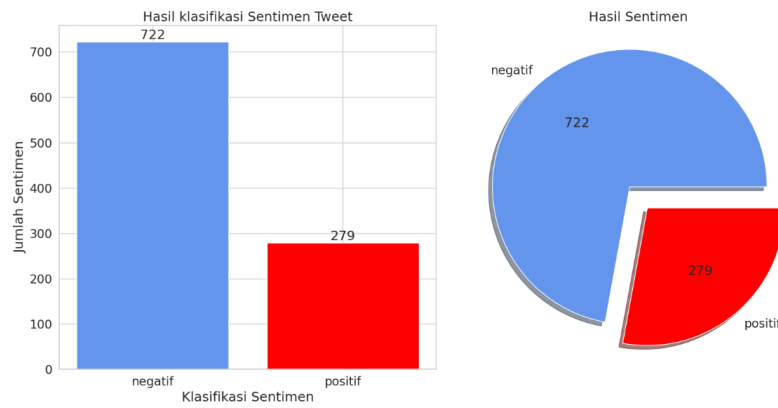
Dengan nilai rata-rata akurasi sebesar 79.8%, F1 Score sebesar 79.96%, presisi sebesar 80.58%, dan recall sebesar 79.8%. Akhirnya, juga menampilkan rata-rata confusion matrix setelah beberapa fold untuk memberikan gambaran yang lebih lengkap tentang performa model secara keseluruhan. Maka diperoleh lah hasil seperti gambar berikut:

```
Average confusion matrix across fold:
[[470. 128.]
 [ 74. 328.]]
```

Gambar 9. Tampilan Hasil Confusion Matrix

Yang berarti 470 adalah True negative, 128 False positive, 74 False Negative dan 328 True Positive. Dengan demikian, proses ini memberikan evaluasi yang komprehensif terhadap kinerja model Support Vector Machine pada data yang telah diolah. Setelah itu masukan data yang sudah di preprocessing untuk pelabelan otomatis, dengan pengaturan minimal df= 3 dokumen dan maximal df = 80% dokumen, sehingga hasilnya dapat dilihat dari gambar berikut:





Gambar 10. Hasil Pelabelan Otomatis

Dari gambar tersebut memiliki hasil sentimen negatif 722 dan positif 279 dari total 1000 sentimen, sehingga 72.2% itu mereka memiliki kesan negatif terhadap RUU Perampasan Aset dan 27.9% memiliki kesan positif.

## V. KESIMPULAN

Analisis sentimen terkait ruu perampasan aset dengan *support vector machine* telah disajikan pada artikel ini. Berdasarkan hasil yang diperoleh, metode *support vector machine* (SVM) terbukti efektif dalam mengklasifikasikan sentimen, dengan akurasi 79.8%, *F1 score* 79.96%, presisi 80.58%, dan *recall* 79.8%. Penyeimbangan kelas menggunakan SMOTE berhasil mengatasi ketidakseimbangan data, sebuah tantangan umum dalam analisis sentimen. Hasil analisis menunjukkan mayoritas sentimen publik terhadap RUU Perampasan Aset adalah negatif, mengindikasikan ketidakpuasan atau ketidakpercayaan terhadap RUU tersebut. Metode SVM dalam penelitian ini memberikan evaluasi komprehensif dan valid terhadap data yang telah dipreprocessing, menghasilkan temuan yang dapat dipercaya untuk analisis lebih lanjut.

## DAFTAR PUSTAKA

- [1] J. Ilmiah, G. Justisi, W. Yusmar, and N. S. Putri, "PERAMPASAN ASET SEBAGAI UPAYA PEMBERANTASAN TINDAK PIDANA PENCUCIAN UANG DENGAN PREDICATE CRIME TINDAK PIDANA NARKOTIKA," 2021.
- [2] M. A. Najib, U. Sunan, A. Surabaya, and K. K. Abstrak, "Sosio Yustisia : Jurnal Hukum dan Perubahan Sosial POLEMIC PENGESAHAN RANCANG UNDANG-UNDANG PERAMPASAN ASET DI INDONESIA," 2023, [Online]. Available: <https://news.detik.com/berita/d-6628926/mahfud-md-jelaskan-soal-heboh-rp->
- [3] I. Hafid, "Perampasan Aset Tanpa Pemidanaan," 2021. [Online]. Available: <https://www.hukumonline.com/berita/baca/t5d0dbe1380889/kasus-setnov--visi-asset-recovery-belum-jadi-prioritas/>,
- [4] O. Viana Agustine, "RUU PERAMPASAN ASET SEBAGAI PELUANG DAN TANTANGAN DALAM PEMBERANTASAN KORUPSI DI INDONESIA," 2019.
- [5] T. Krisdiyanto, E. Maricha, and O. Nurharyanto, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers," *Jurnal CoreIT*, vol. 7, no. 1, 2021.
- [6] Putri Wiyanie, "Penyalahgunaan Aplikasi X sebagai Media Akses Konten Pornografi," *Kalbisocio, Jurnal Bisnis dan Komunikasi*, vol. 11, Mar. 2024, Accessed: Jun. 14, 2024. [Online]. Available: <http://ojs.kalbis.ac.id/index.php/kalbisocio/article/view/3264/910#>
- [7] Parinata Dwi, "PENGARUH PENGGUNAAN APLIKASI YOUTUBEDAN FACEBOOK TERHADAP HASIL BELAJAR MATEMATIKA," *Jurnal Ilmiah Matematika Realistik (JI-MR)*, vol. 1, 2021, Accessed: Jun. 14, 2024. [Online]. Available: <https://jim.teknokrat.ac.id/index.php/pendidikanmatematika/article/view/1061/328>
- [8] M. Ardiansyah *et al.*, "ANALISIS PEMANFAATAN MEDIA PEMBELAJARAN YOUTUBE DALAM MENINGKATKAN PEMAHAMAN KONSEP MATEMATIKA PESERTA DIDIK," 2022. Accessed: Jun. 14, 2024. [Online]. Available: <https://proceeding.unindra.ac.id/index.php/semnasristek/article/view/5828/1439>
- [9] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Teknika*, vol. 10, no. 1, pp. 18–26, Feb. 2021, doi: 10.34148/teknika.v10i1.311.
- [10] D. Oktavia and Y. R. Ramadahan, "Analisis Sentimen Terhadap Penerapan Sistem E-Tilang Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Media Online*, vol. 4, no. 1, pp. 407–417, 2023, doi: 10.30865/klik.v4i1.1040.
- [11] A. Rahman Isnain, A. Indra Sakti, D. Alita, and N. Satya Marga, "SENTIMEN ANALISIS PUBLIK TERHADAP KEBIJAKAN LOCKDOWN PEMERINTAH JAKARTA MENGGUNAKAN ALGORITMA SVM," *JDMSI*, vol. 2, no. 1, pp. 31–37, 2021, [Online]. Available: <https://t.co/NfhmfMjtXw>
- [12] P. Arsi and R. Waluyo, "ANALISIS SENTIMEN WACANA PEMINDAHAN IBU KOTA INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 8, no. 1, pp. 147–156, 2021, doi: 10.25126/jtiik.202183944.
- [13] N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," *Jurnal Informatika: Jurnal pengembangan IT (JPIT)*, vol. 6, no. 3, 2021, [Online]. Available: <http://situs.com>
- [14] Y. Julianto, D. H. Setiabudi, and S. Rostianingsih, "Analisis Sentimen Ulasan Restoran Menggunakan Metode Support Vector Machine," 2022. Accessed: Jun. 14, 2024. [Online]. Available: <https://publication.petra.ac.id/index.php/teknik-informatika/article/view/11888/10482>

- [15] P. R. Alvita Wagiswari D, I. Susilawati, A. Witanti, P. Studi Informatika, F. Teknologi Informasi, and U. Mercu Buana Yogyakarta, "Analisis Sentimen pada Komentar Aplikasi MyPertamina dengan Metode Multinomial Naïve Bayes," 2023. Accessed: Jun. 14, 2024. [Online]. Available: <https://jurnal.forai.or.id/index.php/forai/article/view/4/2>
- [16] R. Al Rasyid, D. Handayani, and U. Ningsih, "Penerapan Algoritma TF-IDF dan Cosine Similarity untuk Query Pencarian Pada Dataset Destinasi Wisata," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 8, no. 1, p. 2024, 2024, doi: 10.35870/jti.
- [17] K. Tri Putra, M. Amin Hariyadi, and C. Crysdiyan, "PERBANDINGAN FEATURE EXTRACTION TF-IDF DAN BOW UNTUK ANALISIS SENTIMEN BERBASIS SVM," 2023. Accessed: Jun. 14, 2024. [Online]. Available: <https://www.ojs.cahayamandalika.com/index.php/jcm/article/view/2292>
- [18] I. S. Wibowo, A. Witanti, and I. Susilawati, "Keyword Extraction Judul Berita Online Di Indonesia Menggunakan Metode TF-IDF," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 11, Mar. 2024, [Online]. Available: <http://jurnal.mdp.ac.id>
- [19] M. Riski Qisthiano, I. Ruswita, A. Prayesy, and \* Korespondensi, "Implementasi Metode SVM dalam Analisis Sentimen Mengenai Vaksin dengan Menggunakan Python 3," *Online) Teknologi: Jurnal Ilmiah Sistem Informasi*, vol. 13, no. 1, pp. 1–7, 2023, doi: 10.26594/teknologi.v13i1.3105.
- [20] G. Radiana and A. Nugroho, "ANALISIS SENTIMEN BERBASIS ASPEK PADA ULASAN APLIKASI KAI ACCESS MENGGUNAKAN METODE SUPPORT VECTOR MACHINE," 2023. Accessed: Jun. 14, 2024. [Online]. Available: <https://ojs.cbn.ac.id/index.php/jukanti/article/view/836/329>
- [21] E. Suryati, A. Ari Aldino, N. Penulis Korespondensi, and E. Suryati Submitted, "Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)," *JURNAL TEKNOLOGI DAN SISTEM INFORMASI*, vol. 4, no. 1, pp. 96–106, Mar. 2023, doi: 10.33365/jtsi.v4i1.2445.
- [22] Suryani and Mustakim, "Estimasi Keberhasilan Siswa dalam Pemodelan Data Berbasis Learning Menggunakan Algoritma Support Vector Machine," *Bulletin of Informatics and Data Science*, vol. 1, no. 2, 2022, [Online]. Available: <https://ejournal.pdsi.or.id/index.php/bids/index>
- [23] R. A. Nurdian, Mujib Ridwan, and Ahmad Yusuf, "Komparasi Metode SMOTE dan ADASYN dalam Meningkatkan Performa Klasifikasi Herregistrasi Mahasiswa Baru," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 1, Apr. 2022, doi: 10.28932/jutisi.v8i1.4004.