

Abdul Syakir - Analisis Sentimen Masyarakat Terhadap Perilaku Korupsi Pejabat Pemerintah Berdasarkan Tweet Menggunakan Naive Bayes Classifier

by Abdul Syakir By Arien B

Submission date: 12-Aug-2023 11:09AM (UTC+0700)

Submission ID: 2144695009

File name: ABDUL_SYAKIR_TI_TURNITIN_KE-2_-_Abdul_Syakir.docx (1.87M)

Word count: 3501

Character count: 22787

Analisis Sentimen Masyarakat Terhadap Perilaku Korupsi Pejabat Pemerintah Berdasarkan Tweet Menggunakan Naive Bayes Classifier

Abdul Syakir¹, Firman Noor Hasan^{2,*}

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, DKI Jakarta, Indonesia

Email: ¹syakirkiki17@gmail.com, ²firman.noorhasan@uhamka.ac.id
Email Penulis Korespondensi: firman.noorhasan@uhamka.ac.id

Abstrak—Perilaku korupsi pejabat pemerintah menjadi masalah yang meresahkan masyarakat dan mengancam integritas sistem pemerintahan. Dalam era digital saat ini, media sosial menjadi sarana penting bagi masyarakat untuk menyampaikan pendapat dan sentimen terhadap isu-isu sosial, termasuk korupsi dari pejabat pemerintah. Penelitian ini bertujuan untuk melakukan analisis sentimen masyarakat terhadap perilaku korupsi dari pejabat pemerintah berdasarkan data tweet di media sosial menggunakan metode Naive Bayes Classifier. Data tweet diambil dari media sosial twitter yang berhubungan dengan kasus-kasus korupsi yang melibatkan pejabat pemerintah dalam periode waktu tertentu. Data tersebut kemudian diproses untuk menghilangkan unsur yang tidak relevan dan mengekstraksi sentimen yang terkandung dalam tweet-tweet tersebut. Metode Naive Bayes Classifier diterapkan untuk mengklasifikasikan tweet-tweet tersebut ke dalam kategori sentimen positif, negatif, atau netral terhadap perilaku korupsi dari pejabat pemerintah. Dalam penerapan algoritma Naive Bayes dan pengujian menggunakan metode 10-fold cross-validation, didapatkan hasil akurasi sebesar 89.73%. Hasil analisis menunjukkan bahwa metode Naive Bayes Classifier efektif dalam mengklasifikasikan sentimen masyarakat terhadap perilaku korupsi dari pejabat pemerintah berdasarkan data tweet di media sosial. Mayoritas tweet yang diamati mengandung sentimen negatif terhadap perilaku korupsi, mencerminkan kekhawatiran dan ketidakpuasan masyarakat terhadap tindakan korupsi yang melibatkan pejabat pemerintah. Penelitian ini memberikan wawasan yang berharga tentang pandangan masyarakat bagi upaya pencegahan dan penanggulangan korupsi dalam sistem pemerintahan.

Kata Kunci: Analisis Sentimen; Korupsi Pejabat Pemerintah; Media Sosial; Tweet; Naive Bayes Classifier.

Abstract—The corrupt behavior of government officials is a problem that worries public and threatens the integrity of the government system. In today's digital era, social media is an important means for the public to voice their opinions and sentiments on social issues, including the corruption of government officials. This study aims to analyze public sentiment toward the corrupt behavior of government officials based on tweet data on social media using the Naive Bayes Classifier method. Tweet data is taken from social media Twitter related to corruption cases involving government officials within a certain period. The data is then processed to remove irrelevant elements and extract the sentiments contained in the tweets. The Naive Bayes Classifier method is applied to classify these tweets of positive, negative, or neutral sentiment toward corrupt behavior from government officials. In applying the Naive Bayes algorithm and testing using the 10-fold cross-validation method, the results obtained an accuracy of 89.73%. The results of the analysis show that the Naive Bayes Classifier method is effective in classifying sentiments towards corrupt behavior of government officials based on tweet data on social media. The majority of tweets observed contained negative sentiments towards corrupt behavior, reflecting public concern and dissatisfaction with acts of corruption involving government officials. This research provides valuable insights into the public's view of efforts to prevent and combat corruption in the government system.

Keywords: Sentiment Analysis; Corruption of Government Officials; Social Media; Tweet; Naive Bayes Classifier.

1. PENDAHULUAN

Pemberitaan mengenai korupsi sepertinya tak pernah berhenti mengisi tayangan televisi. Banyak dari pelaku korupsi adalah pejabat pemerintahan yang menduduki posisi penting. Korupsi merupakan tindakan yang melanggar hukum dengan maksud untuk memperkaya diri sendiri atau orang lain, sehingga menyebabkan kerugian finansial bagi negara atau perekonomian negara (berdasarkan UU No. 20/2001) [1]. Korupsi telah merugikan masyarakat dengan menghambat pembangunan, merusak kepercayaan publik terhadap lembaga pemerintahan, serta menciptakan ketidakadilan dalam distribusi sumber daya [2]. Salah satu bentuk korupsi yang menjadi perhatian serius adalah perilaku korupsi yang dilakukan oleh pejabat pemerintah.

Perilaku korupsi dari pejabat pemerintah memiliki dampak yang buruk dan merugikan [3]. Dalam ranah sektor publik, [4] korupsi dapat menyebabkan pemborosan sumber daya dan mengakibatkan penurunan kualitas pelayanan publik [5]. Selain itu, korupsi juga memiliki dampak di luar aspek keuangan, seperti merusak kepercayaan pada sistem akuntabilitas [6] pemerintah, merendahkan citra pemerintahan di mata publik, dan mengganggu kualitas pelayanan publik secara keseluruhan [7]. Hal ini berakibat negatif terhadap kepercayaan masyarakat terhadap pemerintahan dan mempengaruhi tingkat partisipasi publik dalam proses politik.

Penggunaan media sosial telah menyebar luas dan sangat pesat dalam berbagai macam masyarakat [8]. Media sosial tidak hanya untuk bersosialisasi, dan komunikasi namun juga untuk menyampaikan aspirasi dan mengutarakan pendapat masyarakat secara mudah dan luas [9], salah satu media sosial yang digunakan untuk mempersentasikan pemikiran masyarakat melalui twitter [10] dalam menanggapi pejabat pemerintah yang korupsi. Hal ini dapat dijadikan sebagai sebuah acuan untuk mengetahui sentimen [11] masyarakat terhadap perilaku korupsi pejabat pemerintah.

4 Data mining merupakan suatu proses teknik dalam pengolahan data yang bertujuan untuk 4 menemukan hubungan atau pola dari data yang tidak diketahui oleh pengguna, kemudian menyajikannya dalam bentuk yang mudah dipahami [12]. Dari hubungan data yang ditemukan, informasi 7 tersebut dapat digunakan untuk pengambilan keputusan [13]. Data mining terdiri dari beberapa kelompok yaitu: Deskripsi, Estimasi, Prediksi, Klasifikasi, Clustering, dan Asosiasi [12]. Klasifikasi 16 adalah proses pengelompokan objek data ke dalam kelas-kelas tertentu berdasarkan karakteristiknya [14]. Salah satu metode yang sering digunakan dalam klasifikasi adalah metode naive Bayes [15]. Analisis sentimen adalah suatu teknik untuk mengumpulkan komentar dan ulasan dari orang lain mengenai suatu hal, [16] misalnya informasi yang terdapat pada media sosial berbasis web, dan kemudian mengklasifikasikan tingkat emosi yang diungkapkan oleh orang tersebut, apakah 5 bersifat netral, positif, atau negatif [17]. Beberapa metode Analisis Sentimen yang umum digunakan meliputi K-Nearest Neighbor (K-NN), Naive Bayes Classifier (NBC), Random Forest, dan Decision Tree [18].

Analisis Sentimen berdasarkan tweet di media 13 sosial sudah dilakukan oleh beberapa peneliti diantaranya yaitu I. R. Afendi, F. Noor Hasan, dkk [19] telah melakukan penelitian analisis sentimen berdasarkan tweet di media sosial terkait pelayanan jasa ekspedisi Anteraja. Penelitian ini menggunakan metode Naive Bayes dan mengumpulkan data tweet berbahasa Indonesia dari tanggal 1 Januari 2021 hingga 16 Agustus 2022. Hasil penelitian menyimpulkan bahwa metode Naive Bayes dapat memprediksi sentimen opini masyarakat terhadap pelayanan Anteraja dengan tingkat akurasi sebesar 85.06%. Tujuan dari penelitian untuk membantu dalam evaluasi, analisis, penilaian, dan pemahaman sikap masyarakat terhadap pelayanan Anteraja berdasarkan data 14 dari media sosial. Peneliti lain yaitu F. Noor Hasan dan M. Dwijayanti [11] juga melakukan penelitian terkait analisis sentimen ulasan pelanggan terhadap layanan Grab di Indonesia. Penelitian ini menggunakan metode klasifikasi Multinomial Naive Bayes Classifier 15 (NBC) dengan dataset berisi 1000 tweet dari Twitter. Hasil kesimpulan dari penelitian ini menunjukkan bahwa metode klasifikasi Naive Bayes Classifier (NBC) memberikan akurasi 17 tertinggi, mencapai 92,5%. Untuk sentimen negatif, nilai presisi yang didapat adalah 57%, nilai recall sebesar 67%, dan f1-score sebesar 62%. Sedangkan untuk sentimen positif, nilai presisi mencapai 97%, nilai recall sebesar 95%, dan f1-score sebesar 96%. Hal ini menunjukkan bahwa metode NBC berhasil dalam mengidentifikasi sentimen negatif dan positif dari ulasan pelanggan tentang layanan Grab Indonesia berdasarkan data tweet.

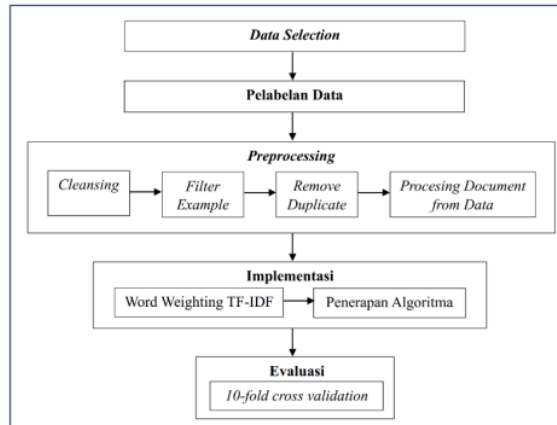
Penelitian terdahulu lainnya dilakukan 5 oleh P. Arsi, B. A. Kusuma, dkk [16] yang meneliti sentimen terkait pemindahan ibu kota menggunakan metode Naive Bayes Classifier. bertujuan untuk mengklasifikasikan sentimen yang bersumber di media sosial Twitter. Dataset berisi 1007 tweet dari twitter. Hasil penelitian menunjukkan tingkat akurasi sebesar 94,33%, nilai presisi yang didapat adalah 0,87% dan recall sebesar 0,99%. Artinya, metode Naive Bayes Classifier berhasil dengan baik dalam mengklasifikasikan sentimen dari tweet-tweet yang berkaitan dengan pemindahan ibu kota, dengan tingkat keberhasilan yang tinggi dan nilai presisi serta recall yang baik. Penelitian lainnya dilakukan oleh Syifa Khairunnisa, Adiwijaya, dkk [20] melakukan penelitian terkait Komentar Masyarakat pada Kasus Pandemi COVID-19. Penelitian menggunakan Support Vector Machine (SVM) untuk mengklasifikasikan komentar-komentar tersebut, menghasilkan tingkat akurasi sebesar 77.77%. Kemudian penelitian lain dilakukan oleh Hakam Febadianrano Putro, Retno Tri Vlandari, dkk [12] melakukan penelitian terkait penerapan klasifikasi pelanggan. Peneliti menggunakan metode naive bayes untuk mengklasifikasi pelanggan yang 2 menghasilkan akurasi sebesar 92%, presisi 100%, dan recall sebesar 91%.

Dalam penelitian ini, peneliti memilih menggunakan Naive Bayes Classifier (NBC) sebagai metode pengelompokan 12 data berdasarkan pada teorema Bayes [21]. Alasan pemilihan metode ini adalah karena NBC memiliki tingkat akurasi yang lebih baik dibandingkan dengan metode K-Nearest Neighbor (K-NN), Random Forest, dan Decision Tree [16].

Tujuan Penelitian ini, Peneliti akan melakukan penentuan klasifikasi sentimen positif atau negative berdasarkan Tweet di Media Sosial dalam mengenai Perilaku Korupsi dari Pejabat Pemerintah. Hasil klasifikasi yang didapat akan di hitung akurasinya pada Metode Naive Bayes Classifier dengan menggunakan confusion matriks [19]. Diharapkan penelitian ini akan memberikan manfaat dalam melakukan penelitian tentang opini masyarakat berdasarkan data tweet yang berkaitan dengan Korupsi Pejabat Pemerintah dan mengidentifikasi sentimen positif atau negatif yang terkandung dalam tweet tersebut [22].

12 2. METODOLOGI PENELITIAN

22 Penelitian yang dilakukan oleh Peneliti melibatkan beberapa tahapan, yang ditunjukkan pada gambar di bawah ini.



Gambar 1. Metode Penelitian

2.1 Data Selection

Tahap awal dimulai dengan proses pengumpulan data (crawling data) [23] dengan menggunakan tools RapidMiner untuk mengambil data yang bersumber dari media sosial Twitter. Seperti ditunjukkan dalam ilustrasi berikut.



Gambar 2. Proses Ekstraksi Data Tweet

Dalam penelitian ini, peneliti melakukan proses ekstraksi data dengan menggunakan kata kunci "Korupsi Pejabat Pemerintah." Data tersebut diambil dari tanggal 29 Mei 2023 hingga 7 Juli 2023, dengan menggunakan teks berbahasa Indonesia. Sebanyak 599 data tweet berhasil dikumpulkan dan disimpan dalam format file dokumen CSV.

Text
@Stev3n_Peg3l Selama ini bangun jalan berbanjar pake hutang, lantas uang pajak yg seharusnya di pergunakan...
@BI00dandsteel @kumparan Mending kita minta bikin search engine buat korupsi bisa bantu KPK dan setidaknya ...
@rezarezarezare @aldyfiqri @bhulukhuduktv Disini uang pajak rakyat dikorupsi, yg korupsi gatau malu?;"166037...
Ada korupsi besar lagi ya ..? Halo Johny G Plate ? diduga beberapa parpol terima aliran dana. Halooo..? Parpol m...
@Relawananies Jgn jd pejabat bermental penjajah, uang hasil ngutang di korupsi kroni-kroninya, begitu negara A...
RT @CutSarina5: Dari segala sudut di korupsi dan tidak henti henti di rezim ini dari satu ke satu kasus korupsi belu...
Terus hasil pajak rakyat kemana? Hasil kekayaan alam Indonesia untuk apa? Korupsi ya? https://t.co/YKDBa20sUk...

Gambar 3. Output Ekstraksi Data Tweet

Ilustrasi di atas menampilkan beberapa contoh output proses ekstraksi data tweet yang terkait dengan sentimen masyarakat terhadap perilaku korupsi dari pejabat pemerintah.

2.2 Pelabelan Data

Setelah data didapatkan dari tahap Data Selection [24], tahap selanjutnya adalah menentukan pelabelan data secara sistem manual dengan 2 peneliti. Proses pelabelan ini dilakukan dengan memberikan sentimen positif dan negatif pada data menggunakan aplikasi *Microsoft Excel 2019*, seperti yang terlihat pada gambar berikut.

1	Text	Sentimen
2	@Stev3n_Peg3l Sementara para pejabat kita kaya raya dengan uang pajak rak	Negatif
3	@BI00dandsteel @kumparan Mending kita minta bikin search engine buat ko	Positif
4	@rezarezarezare @aldyfiqri @bhulukhuduktv Disini uang pajak rakyat dikoru	Negatif
5	Ada korupsi besar lagi ya ..? Halo Johny G Plate ? diduga beberapa parpol teri	Negatif
6	@Relawananies Jgn jd pejabat bermental penjajah, uang hasil ngutang di koru	Positif
7	RT @CutSarina5: Dari segala sudut di korupsi dan tidak henti henti di rezim in	Negatif

Gambar 4. Output Proses pelabelan data tweet

2.3 Preprocessing

Tahap selanjutnya Preprocessing merupakan langkah mengubah teks yang tidak teratur menjadi teks yang terstruktur dalam mengolah data menggunakan perangkat lunak rapidminer studio [25]. Pada tahapan ini dilakukan untuk membersihkan data yang diinginkan sebelum dilakukannya analisis lebih lanjut.

- a. Cleansing yaitu tahapan untuk membersihkan karakter-karakter yang tidak dibutuhkan [25]. Pada tahapan cleansing terdapat beberapa sub-proses, yaitu :
1. Replace RT
Merupakan tahapan yang berfungsi untuk menghapus kata – kata RT dalam data komentar Twitter.
 2. Replace URL 1
Merupakan tahapan yang berfungsi untuk menghapus kata URL yang terletak diawal dan ditengah tweet dalam data komentar Twitter.
 3. Replace URL 2
Merupakan tahapan yang berfungsi untuk menghapus kata URL yang terletak diakhir tweet dalam data komentar Twitter.
 4. Replace Hashtag
Merupakan tahapan yang berfungsi untuk menghapus simbol Hashtag pada tweet dalam data komentar Twitter.
 5. Replace Mention
Merupakan tahapan yang berfungsi untuk menghapus simbol mention pada tweet dalam data komentar Twitter.
 6. Replace simbol
Merupakan tahapan yang berfungsi untuk menghapus simbol – simbol pada tweet dalam data komentar Twitter.
 7. Trim
Merupakan tahapan yang berfungsi untuk menghapus wide space atau spasi – spasi yang tidak perlu pada data komentar Twitter.



Gambar 5. Operator cleansing yang digunakan

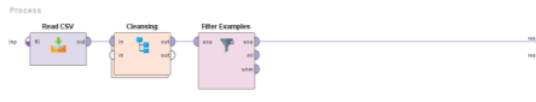
Berikut tabel proses cleansing tweet.

Tabel 1. Hasil Cleansing

Before	After	Sentimen
@Stev3n_Peg3l Sementara para pejabat kita kaya raya dengan uang pajak rakyat nya;"1659885144667484160"	Sementara para pejabat kita kaya raya dengan uang pajak rakyat nya1659885144667484160	Negatif
@Relawananies Jgn jd pejabat bermental penjajah, uang hasil ngutang di korupsi kroni-kroninya, begitu negara APBNnya minus, tinggal naikan pajak rakyatnya,,;"1660306175722938370"	Jgn jd pejabat bermental penjajah uang hasil ngutang di korupsi kroni kroninya begitu negara APBNnya minus tinggal naikan pajak rakyatnya1660306175722938370	Positif
@rezarezarezare @aldyfiqri @bhulukhuduktv Disini uang pajak rakyat dikorupsi, yg korupsi gatau malu ?;"1660370110631247872"	Disini uang pajak rakyat dikorupsi yg korupsi gatau malu 1660370110631247872	Negatif

b. Filter Example

Merupakan tahapan penyaringan data yang berupa tweet komentar kosong atau tweet yang tidak memiliki karakter satupun [26]. Pada tahap ini data awal yang berjumlah 599 data komentar, setelah dilakukan proses filter data komentar berubah jumlahnya menjadi 579 data.



Gambar 6. Operator Filter Example yang digunakan

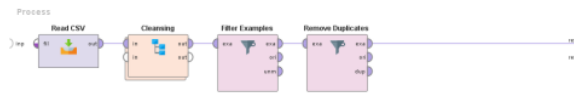
Berikut tabel proses Filter Example tweet.

Tabel 2. Hasil Filter Example

Before	After	Sentimen
?	Sementara para pejabat kita kaya raya dengan uang pajak rakyat nya1659885144667484160	Negatif
Sementara para pejabat kita kaya raya dengan uang pajak rakyat nya1659885144667484160	Jgn jd pejabat bermental penjajah uang hasil ngutang di korupsi kroni kroninya begitu negara APBNnya minus tinggal naikan pajak rakyatnya1660306175722938370	Positif
Jgn jd pejabat bermental penjajah uang hasil ngutang di korupsi kroni kroninya begitu negara APBNnya minus tinggal naikan pajak rakyatnya1660306175722938370	Disini uang pajak rakyat dikorupsi yg korupsi gatau malu 1660370110631247872	Negatif

c. Remove Duplicate

Merupakan tahapan untuk menghilangkan data komentar yang memiliki value yang sama [25]. Pada tahap ini data sebelumnya berjumlah 579 data, setelah dilakukan proses remove duplicate jumlah nya menjadi 319 data. Dan jumlah kategori yang paling banyak setelah proses remove duplicate yaitu kategori sentimen negatif sebesar 224 dan kategori yang paling sedikit yaitu sentimen Positif sebanyak 95.



Gambar 7. Operator Remove Duplicate yang digunakan

Berikut tabel proses Remove Duplicate tweet.

Tabel 3. Hasil Remove Duplicate

Before	After	Sentimen
iya enak aja uang pajak kita t alias milyar di korupsi jamaah belum yg lain	Sementara para pejabat kita kaya raya dengan uang pajak rakyat nya1659885144667484160	Negatif
iya enak aja uang pajak kita t alias milyar di	Jgn jd pejabat bermental penjajah uang hasil	Positif

korupsi jamaah belum yg lain	ngutang di korupsi kroni kroninya begitu negara APBNnya minus tinggal naikan pajak rakyatnya16603061757 22938370	
Sementara para pejabat kita kaya raya dengan uang pajak rakyat nya16598851446674841 60	Disini uang pajak rakyat dikorupsi yg korupsi gatau malu 1660370110631247872	Negatif

d. Processing Document from Data

Setelah print data, selanjutnya melakukan Tahapan Processing Document from Data yang terdapat beberapa sub-proses [19], yaitu : Tokenize, Transform Cases, Filter Stopwords, dan Filter Tokens.



Gambar 8. Operator Processing Document from Data yang digunakan

1. Tokenize

Merupakan sub-proses yang berfungsi untuk mempermudah proses stopwords, kalimat perlu dipecah menjadi kata-kata yang terpisah [20].

Berikut tabel proses tokenize tweet.

Tabel 4. Hasil Tokenize

Before	After	Sentimen
Sementara para pejabat kita kaya raya dengan uang pajak rakyat nya16598851446674841 60	Sementara para pejabat kita kaya raya dengan uang pajak rakyat nya	Negatif
Jgn jd pejabat bermental penjahang uang hasil ngutang di korupsi kroni kroninya begitu negara APBNnya minus tinggal naikan pajak rakyatnya166030617572 2938370	Jgn jd pejabat bermental penjahuang hasil ngutang di korupsi kronikroninya begitu negara APBNnya minustinggal naikan pajak rakyatnya	Positif
Disini uang pajak rakyat dikorupsi yg korupsi gatau malu 1660370110631247872	Disini uang pajak rakyat dikorupsi yg korupsi gatau malu	Negatif

2. Transform Cases

Merupakan sub-proses yang berfungsi untuk merubah kata-kata yang berkapital dalam dataset menjadi huruf kecil [27].

Berikut tabel proses transform case tweet.

Tabel 5. Hasil Transform Cases

Before	After	Sentimen
Sementara para pejabat kita kaya raya dengan uang pajak rakyat nya	sementara para pejabat kita kaya raya dengan uang pajak rakyat nya	Negatif
Jgn jd pejabat bermental penjahauang hasil ngutang di korupsi kronikroninya begitu negara APBNnya minustinggal naikan pajak rakyatnya	jgn jd pejabat bermental penjahauang hasil ngutang di korupsi kronikroninya begitu negara apbnya minustinggal naikan pajak rakyatnya	Positif
Disini uang pajak rakyat dikorupsi yg korupsi gatau malu	disini uang pajak rakyat dikorupsi yg korupsi gatau malu	Negatif

3. Filter Stopwords

Merupakan sub-proses yang berfungsi untuk menghapus ¹ semua kata penghubung dan kata-kata yang tidak relevan dari dataset [19], peneliti menggunakan kamus stopwords yang diperoleh dari situs web www.keagle.com.

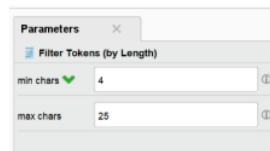
Berikut tabel proses filter stopwords tweet.

Tabel 6. Hasil Filter Stopwords

Before	After	Sentimen
sementara para pejabat kita kaya raya dengan uang pajak rakyat nya	pejabat kaya raya uang pajak rakyat nya	Negatif
jgn jd pejabat bermental penjahauang hasil ngutang di korupsi kronikroninya begitu negara apbnya minustinggal naikan pajak rakyatnya	jgn jd pejabat bermental penjahauang hasil ngutang korupsi kronikroninya negara apbnya minustinggal naikan pajak rakyatnya	Positif
disini uang pajak rakyat dikorupsi yg korupsi gatau malu	uang pajak rakyat dikorupsi yg korupsi gatau malu	Negatif

4. Filter Tokens

Merupakan sub-proses yang berfungsi untuk penghapusan kata-kata dengan jumlah huruf yang telah ditentukan [19]. Data yang digunakan hanya terdiri dari kalimat dengan panjang minimal 4 kata dan maksimal 25 kata.



Gambar 9. Parameters Filter Tokens yang dipergunakan

Berikut tabel proses filter tokens by length tweet.

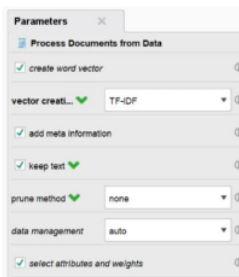
Tabel 7. Hasil Filter Tokens

Before	After	Sentimen
pejabat kaya raya uang pajak rakyat nya	pejabat kaya raya uang pajak rakyat	Negatif
jgn jd pejabat bermental penjahauang hasil ngutang korupsi kronikroninya negara apbnya minustinggal naikan pajak rakyatnya	pejabat bermental penjahauang hasil ngutang korupsi kronikroninya negara apbnya minustinggal naikan pajak rakyatnya	Positif
uang pajak rakyat dikorupsi yg korupsi gatau malu	uang pajak rakyat dikorupsi korupsi gatau malu	Negatif

3. HASIL DAN PEMBAHASAN.

3.1 Word Weighting TF-IDF

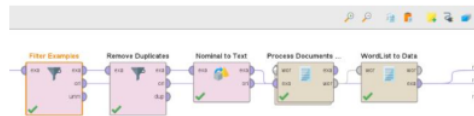
Proses ini bertujuan untuk menentukan nilai bobot pada beberapa tweet yang telah melalui proses preprocessing. Metode yang digunakan yaitu vector creation TF-IDF di operator proses dokumen from data yang telah menggabungkan langkah-langkah Tokenize, Transform Cases, Filter Stopwords, dan Filter Tokens [19]. Seperti dalam ilustrasi gambar berikut.



2

Gambar 10. Parameters Process Documents from Data yang digunakan

Tahap selanjutnya Operator Process Document from Data dihubungkan dengan Operator WordList to Data untuk dilakukan perhitungan bobot nilai kata dan frekuensi yang bermunculan [19] pada kumpulan data, seperti yang di ilustrasikan pada gambar berikut.



2

Gambar 11. Mengubungkan Operator Process Document from Data ke Operator WordList to Data

Row No.	word	in docu...	total	in class (Po...	in class (Ne...
1168	pajak	240	269	61	208
779	korupsi	191	217	57	160
1199	pejabat	89	102	32	70
1353	rakyat	69	87	16	71
1618	uang	50	55	12	43
121	bayar	47	51	1	50
1092	negara	46	52	19	33
1157	orang	24	27	6	21
1218	pemerintah	19	20	0	20
547	hukum	17	18	4	14

Gambar 12. Setelah hasil Proses Word Weighting TF-IDF

Gambar di atas menunjukkan bahwa kata "Pajak" adalah kata yang paling sering muncul dalam dataset dengan frekuensi sebanyak 269 kata. Setelah itu, diikuti oleh kata "Korupsi" dengan 217 kemunculan, dan kata "Pejabat" dengan 102 kemunculan. Selain itu, peneliti telah melakukan visualisasi data dengan membuat wordcloud untuk menampilkan 40 kata dengan frekuensi kemunculan yang sering digunakan, seperti yang di ilustrasikan pada gambar berikut.



Gambar 13. Wordcloud dari 40 kata yang paling sering muncul

Dari wordcloud yang dihasilkan, dapat dilihat bahwa ukuran kata dalam wordcloud mencerminkan jumlah kemunculan kata tersebut dalam dataset. Semakin besar ukuran kata, semakin banyak kata tersebut muncul dalam dataset. Oleh karena itu, dapat disimpulkan bahwa kata-kata "pajak", "korupsi", dan "pejabat" adalah kata-kata yang paling sering muncul dalam dataset.

3.2 Penerapan Algoritma

Penelitian ini memanfaatkan algoritma Naive Bayes untuk melakukan analisis pada dataset yang telah melalui tahap pembobotan dengan metode TF-IDF. Dataset yang digunakan terdiri dari dua kategori sentimen, yaitu sentimen positif dan sentimen negatif [24]. Untuk membagi data menjadi data test dan data Training, peneliti menggunakan perangkat lunak RapidMiner Studio yang melakukan proses pembagian secara otomatis.

Berikut ini adalah contoh langkah manual dalam menghitung algoritma Naive Bayes untuk tweet yang dipilih secara acak dari kumpulan data (dataset) yang telah melalui tahap preprocessing dan pembobotan TF-IDF [19]. Tweet yang digunakan sebagai contoh adalah "Pejabat Korupsi Uang Negara," yang telah diberi label manual sebagai kategori sentimen negatif. Di bawah ini adalah tabel yang menunjukkan frekuensi kata-kata tersebut.

Tabel 8. Frekuensi Kata Pada Tweet

Kata	Class Positif	Class Negatif	Total
Pejabat	32	70	102
Korupsi	57	160	217
Uang	12	43	55
Negara	19	33	52

Keterangan :

$P1$ = Pejabat Class Positif $P2$ = Pejabat Class Negatif Np = Total Pejabat

$K1$ = Korupsi Class Positif $K2$ = Korupsi Class Negatif Nk = Total Korupsi

$U1$ = Uang Class Positif $U2$ = Uang Class Negatif Nu = Total Uang

$N1$ = Negara Class Positif $N2$ = Negara Class Negatif Nn = Total Negara

$Npositif$ = Total Sentimen Positif $Nnegatif$ = Total Sentimen Negatif

Nilai Kelas Positif

$$= \left(\frac{P1}{Np} \times \frac{K1}{Nk} \times \frac{U1}{Nu} \times \frac{N1}{Nn} \right) \times \frac{Npositif}{Npositif + Nnegatif}$$

$$= \left(\frac{70}{102} \times \frac{160}{217} \times \frac{43}{55} \times \frac{33}{52} \right) \times \frac{224}{319}$$

$$= 0$$

Nilai Kelas Negatif

$$= \left(\frac{P2}{Np} \times \frac{K2}{Nk} \times \frac{U2}{Nu} \times \frac{N2}{Nn} \right) \times \frac{Nnegatif}{Npositif + Nnegatif}$$

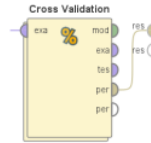
$$= \left(\frac{32}{102} \times \frac{57}{217} \times \frac{12}{55} \times \frac{19}{52} \right) \times \frac{95}{319}$$

$$= 0.176$$

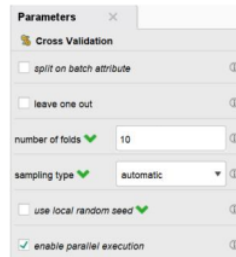
3.3 Pengujian Evaluasi

Peneliti melakukan pengujian *10-fold cross validation* menggunakan operator *cross validation*, untuk mengevaluasi hasil klasifikasi dan akurasi data dalam penerapan algoritma *Naive Bayes* [24].

Berikut operator dan parameter yang digunakan oleh peneliti.



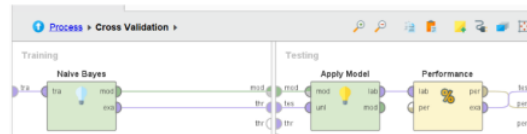
Gambar 14. Operator *cross validation*



Gambar 15. Parameter *cross validation*

Setelah mengatur operator dan parameter yang dibutuhkan, langkah berikutnya adalah menghubungkan data training dan data test ke algoritma *Naive Bayes* untuk mengevaluasi penerapannya [12]. Dengan kata lain, data test yang telah diolah sebelumnya akan digunakan untuk melatih model *Naive Bayes*, sedangkan data training akan digunakan untuk menguji performa model yang telah dilatih.

Evaluasi dilakukan dengan menggunakan operator *performance*, yang ditunjukkan seperti gambar berikut.



Gambar 16. Proses *Cross Validation*

Setelah menjalankan semua operator yang ditentukan menggunakan aplikasi RapidMiner Studio.

Berikut hasil dari nilai akurasi implementasi algoritma *Naive Bayes* dengan *10-Fold Cross Validation*.

Table View Plot View

accuracy: 89.73% +/- 5.33% (micro average: 89.73%)

	true Positif	true Negatif	class precision
pred. Positif	212	34	86.18%
pred. Negatif	12	190	94.06%
class recall	94.64%	84.82%	

Gambar 17. Hasil Accuracy dari Performance Vector

		Pred. Positif	Pred. Negatif
Predicted Values	Pred. Positif	[True Positif] 210	[False Positif] 36
	Pred. Negatif	[False Negatif] 14	[True Negatif] 188

Gambar 18. Visualisasi Confusion matrix

Berikut cara perhitungan manual mencari nilai akurasi dengan menggunakan rumus berikut:

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \\
 &= \frac{212+190}{212+190+34+12} \times 100\% \\
 &= \frac{402}{448} \times 100\% \\
 &= 89.73
 \end{aligned}$$

Pada gambar di atas, terlihat hasil pengujian dataset menggunakan metode *10-Fold cross validation* dan algoritma *naive bayes* dalam aplikasi *RapidMiner Studio* menghasilkan akurasi nilai yang diperoleh adalah 89.73%. Dari pengolahan dataset tersebut, masyarakat sangat marah dan kecewa atas perilaku korupsi yang dilakukan oleh pejabat pemerintah. Oleh karena itu, pemerintah harus meningkatkan transparansi dalam pengelolaan dana, menegakkan hukum secara adil dan tegas terhadap pelaku korupsi untuk memberikan efek jera dan memberantas praktik korupsi.

4. KESIMPULAN

Berdasarkan analisis penelitian pada pemrosesan dataset yang dikumpulkan dari tanggal 29 Mei 2023 hingga 7 Juli 2023 berjumlah 599 data tweet, ditemukan bahwa masyarakat memiliki sentimen negatif terhadap Perilaku Korupsi Pejabat Pemerintah. Dibuktikan dengan kategori sentimen yang paling dominan adalah sentimen negatif dengan jumlah data sebanyak 224 dan terdapat 95 data yang masuk ke dalam kategori sentimen positif. Pada Proses Pembobotan TF-IDF terdapat 3 kata yang sangat mendominasi dan sering muncul pada sentimen-sentimen masyarakat yaitu kata "Pajak", "Korupsi" dan kata "Pejabat". Hal ini menunjuk bahwa masyarakat sangat marah, kecewa, serta tingkat kepercayaan rendah terhadap perilaku korupsi yang dilakukan oleh pejabat pemerintah. Khususnya yang dilakukan oleh pejabat pajak. Selain itu, tingkat kepercayaan masyarakat terhadap pejabat pemerintah semakin berkurang. Dalam penerapan algoritma *Naive Bayes* dan pengujian menggunakan metode *10-fold cross validation*, didapatkan hasil akurasi sebesar 89.73%. True Positif 210, True Negatif 188, False Positif 36, dan False Negatif 14.

Abdul Syakir - Analisis Sentimen Masyarakat Terhadap Perilaku Korupsi Pejabat Pemerintah Berdasarkan Tweet Menggunakan Naive Bayes Classifier

ORIGINALITY REPORT

14%

SIMILARITY INDEX

12%

INTERNET SOURCES

7%

PUBLICATIONS

2%

STUDENT PAPERS

PRIMARY SOURCES

1	djournals.com Internet Source	1%
2	ejurnal.seminar-id.com Internet Source	1%
3	ecampus.pelitabangsa.ac.id Internet Source	1%
4	id.123dok.com Internet Source	1%
5	journal.upgris.ac.id Internet Source	1%
6	Muhammad Aqil Emeraldi, Inna Ekawati, Malikus Sumadyo. "Analisis Sentimen Masyarakat Terhadap Kebijakan Pemerintah Selama Pandemi Covid-19 Menggunakan Algoritma Naïve Bayes", Journal of Students' Research in Computer Science, 2022 Publication	1%

repository.bsi.ac.id

7	Internet Source	1 %
8	core.ac.uk Internet Source	1 %
9	Rani Nooraeni, Heny Dwi Sariyanti, Aulia Fikri Fadhilah Iskandar, Siti Fatimatul Munawwaroh et al. "Analisis Sentimen Data Twitter Mengenai Isu RUU KPK Dengan Metode Support Vector Machine (SVM)", Paradigma - Jurnal Komputer dan Informatika, 2020 Publication	1 %
10	www.coursehero.com Internet Source	<1 %
11	Rangga Gelar Guntara. "Aplikasi Deteksi Phising Berbasis Android Menggunakan Metode Pengembangan Perangkat Lunak DSRM", Jurnal Minfo Polgan, 2023 Publication	<1 %
12	doku.pub Internet Source	<1 %
13	Rina Noviana, Isram Rasal. "PENERAPAN ALGORITMA NAIVE BAYES DAN SVM UNTUK ANALISIS SENTIMEN BOY BAND BTS PADA MEDIA SOSIAL TWITTER", Jurnal Teknik dan Science, 2023 Publication	<1 %

14 Fajar Sidik, Ibnu Suhada, Azhar Haikal Anwar, Firman Noor Hasan. "Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier", Jurnal Linguistik Komputasional (JLK), 2022
Publication <1 %

15 Jimmy, Eni Heni Hermaliani, Laela Kurniawati. "ANALISIS KLASIFIKASI SENTIMEN PENGGUNA MEDIA SOSIAL TWITTER TERHADAP PENUNDAAN PEMILU PRESIDEN TAHUN 2024", Jurnal Indonesia : Manajemen Informatika dan Komunikasi, 2023
Publication <1 %

16 jti.respati.ac.id
Internet Source <1 %

17 Submitted to Academic Library Consortium
Student Paper <1 %

18 Submitted to Universitas Amikom
Student Paper <1 %

19 www.scribd.com
Internet Source <1 %

20 dergipark.org.tr
Internet Source <1 %

21 ejurnal.ars.ac.id
Internet Source <1 %

es.scribd.com

22

Internet Source

<1 %

23

fantowiworld.blogspot.com

Internet Source

<1 %

24

Gunawan Gunawan, Abd. Charis Fauzan, Harliana Harliana. "Implementasi Algoritma Decision Tree Iterative Dichotomiser 3 (ID3) untuk Prediksi Keberhasilan Pengobatan Penyakit Kutil Menggunakan Cryotherapy", Jurnal Bumigora Information Technology (BITe), 2022

Publication

<1 %

25

publishing-widyagama.ac.id

Internet Source

<1 %

26

qdoc.tips

Internet Source

<1 %

27

repository.widyatama.ac.id

Internet Source

<1 %

28

e-journal.hamzanwadi.ac.id

Internet Source

<1 %

29

journals.ju.edu.et

Internet Source

<1 %

30

marsiajarbe.blogspot.com

Internet Source

<1 %

31

text-id.123dok.com

Internet Source

<1 %

32

toffeede.com

Internet Source

<1 %

33

www.riss.kr

Internet Source

<1 %

34

Sepyan Purnama Kristanto, Junaedi Adi Prasetyo, Edwin Pramana. "Naive Bayes Classifier on Twitter Sentiment Analysis BPJS of HEALTH", 2019 2nd International Conference of Computer and Informatics Engineering (IC2IE), 2019

Publication

<1 %

35

Jarwandi Jarwandi, Hedy Ramadhan Putra Pembangunan. "PENANAMAN NILAI-NILAI KARAKTER ANTI KORUPSI PADA PEMBELAJARAN KIMIA", Academy of Education Journal, 2022

Publication

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off