

ANALISIS KEPUASAN MASYARAKAT TERHADAP PERILAKU KORUPSI PEMERINTAH BERDASARKAN KOMENTAR PADA SOSIAL MEDIA MENGUNAKAN NAIVE BAYES CLASSIFIER

Faldy Irwiensyah^{1*}, Firman Noor Hasan²

¹Program Studi Sistem dan Teknologi Informasi

²Program Studi Teknik Informatika

^{1,2}Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

Correspondence email: faldy@uhamka.ac.id

Article history: Revised: 3 November, 2025

Accepted: November 28, 2025

Published: November 30, 2025

ABSTRACT

Corruption in Indonesia is dominated by government officials, which frequently occurs and is disturbing to the public. Social media has now become a very important platform for the public. Through social media, the public can voice their opinions on social issues and also express their satisfaction with government performance, including corrupt behavior carried out by officials. This study aims to analyze public satisfaction with corrupt behavior carried out by government officials based on public tweets on social media using the Naïve Bayes Classifier method. Data was obtained from platform X with keywords closely related to corruption cases involving government officials. The results of this study concluded that the word weighting process using TF-IDF resulted in three very dominant words that frequently appeared in public tweets, namely the words "Corruption", "Official", and the word "Tax". This shows that the public is very dissatisfied with government performance, resulting in a very low level of trust.

Keywords: *Public Satisfaction Analysis, Corrupt Behavior of Government Officials, Naïve Bayes Classifier, Social Media, 10-Fold Cross Validation.*

ABSTRAK

Perilaku korupsi yang terjadi di Indonesia didominasi oleh oknum pemerintah yang sering terjadi dan meresahkan dimasyarakat. Media sosial saat ini telah menjadi platform yang sangat penting dimasyarakat. Dengan media sosial, masyarakat dapat menyuarakan pendapat terhadap isu-isu sosial dan juga menyatakan kepuasan terhadap kinerja pemerintah, termasuk perilaku korupsi yang dilakukan oleh oknum ditingkat pejabat pemerintah. Penelitian ini bertujuan untuk menganalisis kepuasan masyarakat terhadap perilaku korupsi yang dilakukan oleh pejabat pemerintah berdasarkan cuitan masyarakat di media sosial dengan menggunakan metode *Naïve Bayes Classifier*. Data diperoleh dari platform X dengan kata kunci yang berkaitan erat dengan kasus korupsi yang melibatkan pejabat pemerintah. Hasil penelitian ini menyimpulkan bahwa dengan proses pembobotan kata dengan menggunakan TF-IDF, maka menghasilkan tiga kata yang sangat dominan dan sering muncul dalam cuitan masyarakat, yaitu kata "*Corruption*", "*Official*", dan kata "*Tax*". Hal ini memperlihatkan bahwa publik sangat tidak puas terhadap kinerja pemerintah, sehingga mengakibatkan tingkat kepercayaan yang sangat rendah.

Kata Kunci: *Analisis Kepuasan Masyarakat, Perilaku Korupsi Oknum Pemerintah, Naïve Bayes Classifier, Media Sosial, 10-Fold Cross Validation.*

PENDAHULUAN

Kabar tentang korupsi yang dilakukan oleh pemerintah seakan tak pernah berhenti memenuhi berita (Kartiko, 2024).

Korupsi yang terjadi di Indonesia lebih banyak dilakukan oleh oknum pemerintah, terutama yang memegang jabatan strategis (Paraswansa & Utomo, 2024). Korupsi merupakan suatu perbuatan melawan hukum yang bertujuan agar memperkaya orang lain atau diri sendiri sehingga menimbulkan kerugian bagi



keuangan negara (UU No. 20/2001). Akibat dari korupsi yang dilakukan oleh oknum pemerintah, dapat menghambat pembangunan, merugikan masyarakat, menurunkan dan bahkan merusak tingkat kepercayaan masyarakat terhadap lembaga pemerintah (Chiu, Hung, Yang, Chen, & Chiu, 2025). Salah satu yang sangat memprihatinkan adalah adanya perilaku korup yang selama ini dilakukan oleh oknum pemerintah, khususnya yang memegang jabatan strategis (Sadik-sZadaa, Gatto, & Niftiyev, 2024). Perilaku korup pejabat pemerintah mempunyai dampak buruk dan merugikan (Wanga, Guan, Zeng, Liang, & Donga, 2024). Pada sektor publik mengakibatkan rusaknya rendahnya kepercayaan masyarakat, sehingga menyebabkan turunnya kualitas pelayanan publik oleh pemerintah (Puspita, L, & Gultom, 2024). Selain itu, korupsi juga mempunyai dampak yang lebih luas dari sekedar aspek keuangan, sehingga dapat menyebabkan hilangnya kepercayaan terhadap akuntabilitas pemerintah, menurunnya citra pemerintah dimata masyarakat, dan terganggunya kualitas pelayanan publik (Bani-Mustafa, Nimer, Uyar, & Schneider, 2024). Hal ini berdampak negatif pada kepercayaan masyarakat terhadap pemerintah dan mempengaruhi, sehingga nantinya dapat menyebabkan penurunan tingkat partisipasi masyarakat didalam berpartisipasi pada kegiatan proses politik (Castro & Lopes, 2023).

Penggunaan media sosial oleh masyarakat sebagai media penyebaran informasi dan menyampaikan aspirasi, telah menjadi umum dan hampir digunakan disemua level tingkat masyarakat (Hasan, Aziz, & Nofendri, 2023). Karena masyarakat sudah menganggap bahwa penggunaan media sosial saat ini tidak hanya berkomunikasi saja melainkan untuk menyampaikan pendapat dan aspirasi mereka, karena dianggap sebagai salah satu cara yang efektif dan mudah (Hasan et al., 2024). Salah satu platform berbasis media sosial yang saat ini digunakan oleh hampir semua level masyarakat untuk menyampaikan pemikiran, adalah melalui media sosial X atau yang dulunya bernama Twitter (Alam, Aftab, Abbas, Ugli, & Bokhari, 2023). Sehingga hal tersebut dijadikan referensi untuk mengetahui tingkat kepuasan masyarakat terhadap perilaku korup yang dilakukan oleh pejabat pemerintah (Syakir & Hasan, 2023).

Data mining merupakan salah satu proses teknis dalam mengolah dan menggali data dengan tujuan menemukan pola dan keterkaitan pada data yang selama ini belum diketahui, kemudian disajikan kedalam format yang lebih mudah untuk dipahami oleh pengguna (Andriyani et al., 2024). Hal ini dapat dilihat dari keterkaitan antar data yang ditemukan dari hasil pengolahan data mining, sehingga informasi tersebut nantinya dapat digunakan untuk pengambilan keputusan (Cantini, Peron, Carlo, & Sgarbossa, 2024). Data mining

terdiri dari beberapa teknik yaitu Description, Estimation, Prediction, Classification, Clustering, dan Association (Sani et al., 2023). Salah satu teknik dalam data mining adalah klasifikasi, yaitu proses pengelompokan objek data ke dalam kelas-kelas tertentu berdasarkan karakteristiknya (Hasan & Ariyansah, 2024). Salah satu teknik atau metode yang paling umum digunakan dalam teknik klasifikasi adalah metode Naive Bayes (Peretz, Koren, & Koren, 2024). Analisis kepuasan masyarakat merupakan teknik didalam mengumpulkan ulasan dan komentar dari media sosial yang didasarkan pada suatu fenomena tertentu, dan kemudian mengklasifikasikan tingkat kepuasan masyarakat menjadi beberapa kategori apakah itu negatif, netral, atau positif (Romano, Contu, Mola, & Conversano, 2024). Beberapa metode yang umum digunakan antara lain Naive Bayes Classifier (NBC), Decision Tree, K-Nearest Neighbor (K-NN), dan Random Forest (Suhud, Febriandirza, Permatasari, & Ramadhan, 2023)(Putra et al., 2023).

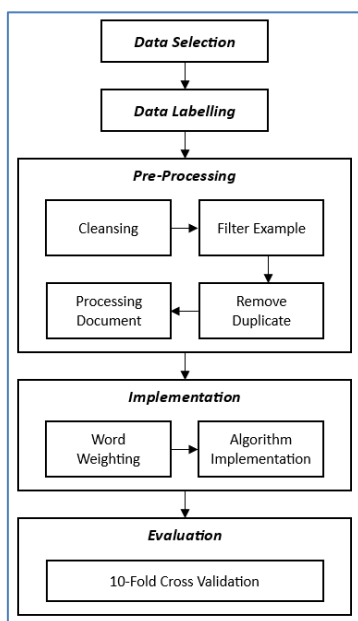
Analisis kepuasan masyarakat yang didapatkan melalui komentar dan ulasan pada media sosial telah dilakukan oleh beberapa peneliti, termasuk penelitian tentang analisis sentimen berdasarkan tweet di media sosial terkait layanan ekspedisi Anteraja. Penelitian ini bertujuan untuk menganalisis, mengevaluasi berdasarkan data dari media sosial terkait penilaian, sikap publik terhadap layanan Anteraja (Afandi, Hasan, Rizki, Pratiwi, & Halim, 2022). Peneliti lain meneliti topik analisis sentimen terhadap ulasan pelanggan pada layanan Grab diIndonesia pada twitter, dengan menggunakan metode *Multinomial Naive Bayes Classifier* (MNBC). Penelitian ini menyimpulkan bahwa metode (MNBC) menunjukkan bahwa untuk sentimen negatif nilai presisi yang diperoleh sebesar 57%, dengan tingkat akurasi sebesar 92,5%, *F1-score* adalah 62%, dan *recall* 67%. Sementara itu, nilai recall 95%, nilai presisi 97%, serta *f1-score* 96% untuk sentimen positif. Sehingga mendapatkan kesimpulan bahwa metode *Multinomial Naive Bayes Classifier* (MNBC) telah berhasil mengidentifikasi sentimen pelanggan layanan Grab Indonesia berdasarkan ulasan data pada tweet (Hasan & Dwijayanti, 2021). Penelitian sebelumnya lainnya yang mengkaji topik adanya pemindahan ibu kota negara Indonesia dengan menggunakan metode Naive Bayes. Tujuan dari penelitian tersebut adalah untuk mengklasifikasikan sentimen pada Twitter. Penelitian menghasilkan bahwa nilai presisi yang diperoleh 0,87%, dengan tingkat akurasi 94,33%, serta recall 0,99%. Hal ini berarti bahwa metode Naive Bayes Classifier berhasil melakukan klasifikasi sentimen dari tweet terkait pemindahan ibu kota negara Indonesia ke IKN dengan keberhasilan tinggi serta nilai precision yang baik (Bu, Liu, & Ju, 2024). Penelitian lain yang dilakukan terkait Ulasan Publik yang terkait Pandemi

COVID-19, peneliti menggunakan *Support Vector Machine* untuk dapat mengklasifikasikan ulasan publik. Hasil penelitian menunjukkan bahwa tingkat akurasi 77,77% (Das & Singh, 2023). Penelitian selanjutnya terkait dengan implementasi klasifikasi komentar pelanggan, yang menggunakan metode klasifikasi dan teknik naive bayes untuk mengklasifikasikan pelanggan dengan akurasi 92%, recall 91%, dan presisi 100% (Rodríguez-Ibáñez, Casáñez-Ventura, Castejón-Mateos, & Cuenca-Jiménez, 2023).

Penelitian ini menggunakan *Naive Bayes Classifier* didalam metode pengelompokan data (Syakir & Hasan, 2023). Salah satu alasannya adalah karena metode tersebut mempunyai tingkat akurasi yang tinggi apabila dibandingkan dengan metode lain seperti Decision Tree, Random Forest, dan K-Nearest Neighbor (Muflih, Abdillah, & Hasan, 2023). Perbedaan penelitian ini dengan penelitian sebelumnya adalah didalam menentukan klasifikasi sentimen, kemudian hasil yang diperoleh akan dihitung tingkat akurasinya menggunakan metode NBC dan *confusion matrix*. Penelitian yang dilakukan ini diharapkan dapat memberikan kontribusi dalam menganalisis kepuasan publik terkait korupsi yang dilakukan oleh oknum pemerintah.

METODE PENELITIAN

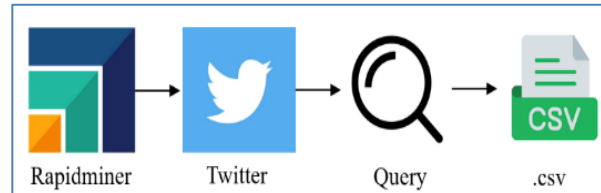
Penelitian ini dilakukan dengan melalui beberapa tahapan, seperti yang diperlihatkan oleh Gambar.1 berikut.



Gambar 1. Tahapan Penelitian

A. Data Selection

Tahap *data selection* merupakan dimulai dengan proses pengumpulan data (*data crawling*) memanfaatkan RapidMiner yang mengekstraksi data dimedia sosial X. Gambar.2 menunjukkan ilustrasi proses ekstraksi data yang dilakukan oleh peneliti.



Gambar 2. Proses Data Crawling

Peneliti mengekstraksi data di media sosial X dengan bantuan kata kunci yaitu "*Government Corruption*". Data dikumpulkan atau diekstraksi mulai sampai dengan Juli 2024, menggunakan teks Bahasa Inggris. Hasilnya mendapatkan sejumlah 599 twit yang berhasil diekstraksi pada berkas dokumen CSV. Gambar.2 memperlihatkan ilustrasi yang menunjukkan contoh keluaran proses dari hasil ekstraksi data twit terkait sentimen publik terhadap korupsi yang dilakukan oleh oknum pemerintah (Steven, Hariyanto, Arijanto, & Wijaya, 2021).

B. Data Labelling

Tahap selanjutnya adalah bagaimana peneliti dapat memberikan label pada data secara manual, tapan ini dinamakan *data labelling* (Fathurrohman, Afandi, & Hasan, 2024). Pelabelan secara manual dilakukan oleh pakar dengan memberi label sentimen negatif, netral, dan positif pada data dengan bantuan tools pengolah data Ms.Excel (Rosida, Firmansyah, & Suherman, 2024).

C. Pre-Processing

Tahap selanjutnya adalah *pre-processing* yaitu dengan mengubah teks yang formatnya tidak beraturan sehingga menjadi teks yang terstruktur pada pengolahan data dengan menggunakan operator pada alat Rapidminer Studio. (Septiana & Pastika, 2024). Tahap ini dilakukan dengan tujuan agar data yang diinginkan dibersihkan sebelum nantinya dapat dilakukan analisis lanjutan.

D. Implementation

Tahap selanjutnya yaitu implementasi, membahas pembobotan kata menggunakan metode TF-IDF dan implementasi algoritmanya. Berikut rumus yang digunakan dalam implementasi algoritma tersebut (Amini & Setiawan, 2024).

$$Positive\ Class\ Value = \frac{\left(\frac{OF1}{Nof} \times \frac{CR1}{Ncr} \times \frac{TA1}{Nta} \times \frac{CU1}{Ncu} \right) \times \frac{Npositive}{Npositive + Nnegative}}{\dots\dots\dots(1)}$$



$OF1$ = Positive Class of Officials
 Nof = Total Officials
 $CR1$ = Positive Class of Corruption
 Ncr = Total Corruption
 $TA1$ = Positive Class of Tax
 Nta = Total Tax
 $CU1$ = Positive Class of Country
 Ncu = Total Country
 $Npositive$ = Total Positive Sentiment
 $Nnegatif$ = Total Negative Sentiment

$$Negative\ Ckass\ Value = \left(\frac{OF2}{Nof} \times \frac{CR2}{Ncr} \times \frac{TA2}{Nta} \times \frac{CU2}{Ncu} \right) \times \frac{Nnegative}{Npositive + Nnegative} \dots\dots\dots(2)$$

$OF2$ = Negative Class of Official
 Nof = Total Officials
 $CR2$ = Negative Class of Corruption
 Ncr = Total Corruption
 $TA2$ = Negative Class of Tax
 Nta = Total Tax
 $CU2$ = Negative Class of Country
 Ncu = Total Country
 $Npositifve$ = Total Positive Sentiment
 $Nnegatif$ = Total Negative Sentiment

E. Evaluation

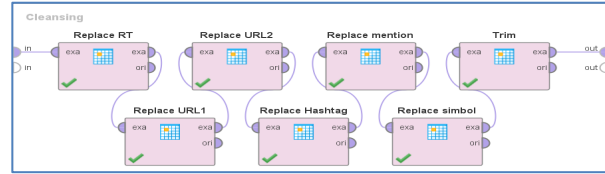
Tahap terakhir adalah evaluasi menggunakan confusion matrix untuk membahas evaluasi model yang telah dibuat dan menganalisis nilai akurasi, dll.

HASIL DAN PEMBAHASAN

Pada bagian ini, akan dibahas secara rinci tahapan *pre-processing*, implementasi, dan evaluasi penelitian yang telah dilakukan. *Pre-processing* merupakan tahapan dalam pemrosesan data menggunakan perangkat lunak Rapid Miner Studio untuk mengubah teks tidak beraturan menjadi teks terstruktur (Hermawan & Thoyyibah, 2023). Tahap *pre-processing* dilakukan untuk membersihkan data yang diinginkan sebelum dilakukan analisis lebih lanjut, tahap ini meliputi pembersihan, penyaringan contoh, penghilangan duplikat, dan pemrosesan dokumen.

A. Cleansing

Tahap *cleansing* adalah langkah-langkah untuk membersihkan karakter yang tidak diperlukan. Langkah-langkah pada tahapan *cleansing* seperti ditunjukkan pada Gambar.3 berikut.



Gambar 3. *Celaning Stages in RapidMiner*

Tahap *cleansing* melibatkan beberapa sub-proses operator pada RapidMiner, yaitu: pertama, operator *Replace RT*, yaitu operator yang berfungsi untuk menghapus kata retweet (RT) pada data komentar dimedia sosial X. Selanjutnya, operator *Replace URL 1*, yaitu operator yang tujuannya agar menghapus teks URL di awal dan di tengah tweet pada komentar dimedia sosial X. Sedangkan *Replace URL 2*, yaitu operator yang berfungsi untuk menghapus kata URL yang terletak di akhir tweet pada data komentar dimedia sosial X. Selanjutnya adalah operator *Replace Hashtag*, yaitu operator yang tujuannya agar menghapus simbol-simbol seperti *Hashtag* pada tweet pada komentar dimedia sosial X. Operator *Replace Mention*, yaitu operator yang tujuannya agar menghapus simbol yang disebutkan pada tweet pada komentar dimedia sosial X. Selanjutnya, operator *Replace Simbol*, yaitu tahap untuk menghapus simbol pada tweet pada data komentar dimedia sosial X. Terakhir, operator *Trim*, yaitu tahap untuk menghapus spasi yang lebar atau spasi yang tidak diperlukan pada data komentar dimedia sosial X. Tabel.1 menunjukkan perubahan pada tahapan *cleansing*.

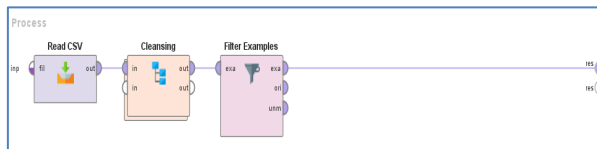
Tabel 1. Hasil Sampel tweet pada Tahap *Cleansing*

Sebelum	Sesudah	Sentimen
@Relawananies Don't be an official with a colonial mentality, the money from debt is used to corrupt his cronies, when the state budget is in the red, all he has to do is increase the people's taxes,,,"160330017522938305"	@Relawananies Don't be an official with a colonial mentality, the money from debt is used to corrupt his cronies, when the state budget is in the red, all he has to do is increase the people's taxes,,,"160330017522938305"	Positive
@Stev3n_Peg3l While our officials are rich with the people's tax money1631887444017408183 money;"163188	While our officials are rich with the people's tax money1631887444017408183	Negative

Sebelum	Sesudah	Sentimen
7444017408183		
"		
@rezarezarezare	@rezarezarezare	Negative
@aldyfiqri	@aldyfiqri	
@bhulukhudukt	@bhulukhudukt	
v Here people's	Here people's tax	
tax money is	money is being	
being corrupted,	corrupted, those who	
those who are	are corrupt have no	
corrupt have no	shame?;"1604375710	
shame?;"160437	001260872"	
5710001260872		
"		

B. Filter Example

Gambar.4 menunjukkan tahapan dalam operator *Filter Example* yang digunakan dalam RapidMiner pada tahapan tersebut.

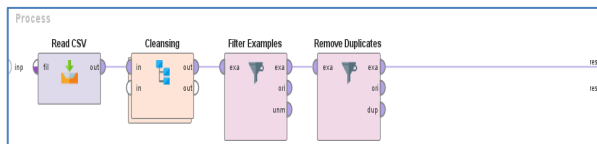


Gambar 4. Tahapan *Filter Example* pada RapidMiner

Tahap *Filtering* data berupa tweet yang berisi ulasan kosong atau bisa dikatakan sebagai tweet yang tidak mempunyai karakter. Tahap ini dilakukan setelah proses penyaringan data komentar, dari data awal yang semula berjumlah 601 data komentar berubah menjadi hanya 581 komentar. Tabel 2 menunjukkan hasil setelah tahap *Filter Example*.

C. Remove Duplicate

Gambar 5 menunjukkan langkah-langkah pada tahapan *Remove Duplicate* dan operator yang digunakan pada RapidMiner.



Gambar 5. *Remove Duplicate Operator* in RapidMiner

Remove Duplicate merupakan tahap untuk menghapus data komentar dengan nilai yang sama. Pada tahap ini, data yang sebelumnya berjumlah 581 data, dan setelah melalui tahapan ini maka jumlahnya menjadi hanya 442 data. Jumlah kategori terbesar setelah proses *remove duplicate* adalah kategori sentimen negatif, yaitu sebanyak 314, dan kategori terkecil yaitu sentimen positif yaitu hanya sejumlah 128. Tabel.2 menunjukkan hasil *Remove Duplicate*.

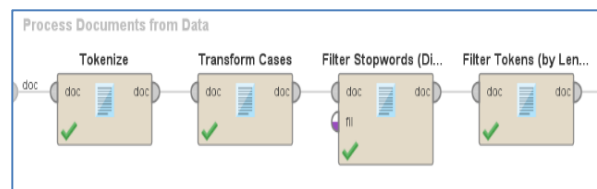


Tabel 2. Contoh Perubahan pada tahapan *Remove Duplicate*

Sebelum	Sesudah	Sentimen
Don't be an official with a colonial mentality, borrow money from debts and corrupt your cronies, when the state budget is in the red, you just have to increase the people's taxes1603300175722938305	Don't be an official with a colonial mentality, borrow money from debts and corrupt your cronies, when the state budget is in the red, you just have to increase the people's taxes1603300175722938305	Positive
While our officials are rich with the people's tax money163188744017408183	While our officials are rich with the people's tax money163188744017408183	Negative
Here people's tax money is being corrupted, those who are corrupt are shameless1604375710001260872	Here people's tax money is being corrupted, those who are corrupt have no shame1604375710001260872	Negative

D. Processing Document from Data

Tahap terakhir dari proses *pre-processing* adalah *Processing Document from Data*, yang terdiri dari beberapa sub-proses, termasuk Tokenisasi, Transformasi Kasus, Filter Stopword, dan Filter Token. Gambar.6 dibawah ini menunjukkan operator yang digunakan dalam tahap *Processing Document from Data* pada RapidMiner.



Gambar 6. Operator *Processing Document from Data* pada RapidMiner

D.1. Tokenize

Tokenize adalah sub-proses yang tujuannya agar memfasilitasi proses *stopword*, yaitu memecah kalimat sehingga menjadi kata-kata yang terpisah-pisah. Tabel.3 dibawah ini menunjukkan hasil dari proses *Tokenize*.

Tabel 2. Perubahan pada tahap *Tokenize*

Sebelum	Sesudah	Sentimen
Don't be an official with a colonial mentality, borrowing money from his cronies to corrupt him, so when the state budget is in the red, he just raises the people's taxes. 1603300175722938305	Don't be an official with a colonial mentality, borrowing money from his cronies to corrupt him, so when the state budget is in the red, he just raises the people's taxes.	Positive
While our officials are rich with the people's tax money 1631887444017408183	While our officials are rich with the people's tax money	Negative
Here people's tax money is being corrupted, those who are corrupt have no shame 1604375710001260872	Here, people's tax money is being corrupted, those who are corrupt have no shame	Negative

D.2. *Transform Cases*

Tahapan ini adalah sub-proses yang tujuannya agar mengubah kata-kata berhuruf kapital pada dataset sehingga menjadi kata-kata dengan huruf kecil semua. Tabel.4 di bawah ini menunjukkan hasil dari tahapan *transform cases*.

Tabel 3. Perubahan pada tahapan *Transform Cases*

Sebelum	Sesudah	Sentimen
Don't be an official with a colonial mentality, borrowing money from his cronies to corrupt him, so when the state budget is in the red, he just raises the people's taxes.	don't be an official with a colonial mentality, the money from debts is corrupted by his cronies, once the state budget is in the red, all that's left is to increase the people's taxes.	Positive
While our officials are rich with the people's tax money	while our officials are rich with the people's tax money	Negative
Here, people's tax money is being	here people's tax money is	Negative

Sebelum	Sesudah	Sentimen
corrupted, those who are corrupt have no shame	being corrupted, those who are corrupt don't know shame	

D.3. *Filter Stopwords*

Tahapan ini adalah sub-proses yang tujuannya agar menghapus kata yang kurang relevan dan kata sambung pada dataset. Pada tahapan ini peneliti dibantu oleh kamus *stopwords* yang didapatkan dari situs kaggle.com. Tabel.5 di bawah ini menunjukkan hasil *Stopwords*.

Tabel 4. Perubahan pada *Stopwords*

Sebelum	Sesudah	Sentimen
don't be an official with a colonial mentality, the money from debts is corrupted by his cronies, once the state budget is in the red, all that's left is to increase the people's taxes.	don't be an official with a colonial mentality, the money is the result of debts, corruption, and cronies, the state budget is in the red, all that's left is to increase the people's taxes.	Positive
while our officials are rich with the people's tax money	officials are rich from people's tax money	Negative
here people's tax money is being corrupted, those who are corrupt don't know shame	people's tax money is being corrupted, those who are corrupt have no shame	Negative

D.4. *Filter Tokens*

Tahapan ini adalah *sub-proses* yang tujuannya agar menghapus kata yang terhimpun dari jumlah huruf tertentu. Tabel.6 dibawah ini menunjukkan hasil perubahan pada *Filter Token*.

Tabel 5. Perubahan pada *Filter Tokens*

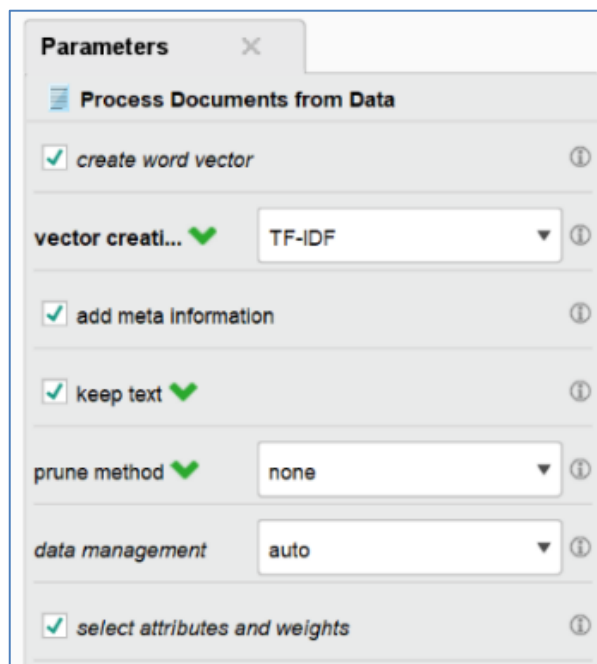
Sebelum	Sesudah	Sentimen
don't be an official with a colonial mentality, the money is the result of debts, corruption, and cronies, the state budget is in the red, all that's left	officials with a colonial mentality, money from debts, corruption, and cronies, the state budget is in the red, all that's left is to increase the people's taxes	Positive



Sebelum	Sesudah	Sentimen
is to increase the people's taxes.	officials are rich from people's tax money	Negative
officials are rich from people's tax money	people's tax money is being corrupted, those who are corrupt have no shame	Negative
people's tax money is being corrupted, those who are corrupt have no shame	people's tax money is being corrupted, corruption is shameless	

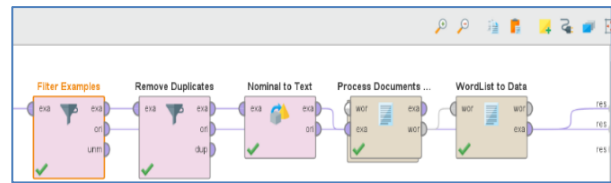
E. Word Weighting TF-IDF

Tahapan in tujuannya agar peneliti dapat menentukan nilai dan bobot dari beberapa tweet yang telah diproses sebelumnya, yaitu dengan pembuatan vektor TF-IDF pada operator *Process Documents from Data* yang sebelumnya telah berhasil didalam mengkombinasikan *Tokenization*, *Filter Stopword*, *Transform Cases*, dan *Filter Token*, seperti yang ditunjukkan pada Gambar.7 berikut.



Gambar 7. Parameter-parameter yang digunakan didalam tahapan *Process Documents form Data*

Tahap selanjutnya adalah menghubungkan operator *Process Document from Data* yang nantinya akan dihubungkan ke operator *WordList to Data*, dengan tujuan agar menghitung bobot dari frekuensi dan nilai kata- yang muncul pada dataset, seperti yang diilustrasikan pada Gambar.8 di bawah ini.



Gambar 8. Menghubungkan operator *Process Documents from Data* kepada operator *WordList to Data*

Hasil hubungan antara kedua operator tersebut menghasilkan Pembobotan Kata TF-IDF, seperti yang ditunjukkan pada Gambar.9 di bawah ini..

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 9. Hasil dari tahapan *Word Wegihting Process* dengan TF-IDF

Tabel.7 di bawah ini memperlihatkan bahwa kata "Corruption" merupakan kata yang sangat sering ada pada dataset, yaitu dengan kemunculan 221 kata. Selanjutnya kata "Official" yaitu dengan 106 kemunculan, serta kata "Tax" dengan 59 kemunculan. Oleh karena itu, kita dapat menyimpulkan bahwa kata "Corruption", "Official", dan "Tax" merupakan kata-kata frekuensi kemunculan terbanyak pada dataset.

Tabel 7. *Word Frequency on Twitter*

Word	Positive Class	Negative Class	Total
Official	34	72	106
Corruption	59	162	221
Tax	14	45	59
Country	21	35	56

F. Algorithm Implementation

Penelitian menggunakan dan mengimplementasikan algoritma *Naive Bayes Classifier* didalam menganalisis komentar yang telah melalui tahap TF-IDF. Perlu diketahui bahwa dataset yang digunakan yaitu dari 2 kategori sentimen, yaitu kategori negatif dan kategori positif dan didalam melakukan pembagian data sehingga menjadi data uji dan data latih.



Tabel.8 di bawah ini menunjukkan hasil frekuensi kata dalam tweet. Berikut tabel dibawah ini merupakan hasil dari pelabelan secara manual serta perhitungannya didalam impelementasi algoritma NBC pada twit yang terpilih acak. Twit yang dipilih dan digunakan untuk contoh yaitu "Official, Corruption, Tax, Country", yang telah dilabeli secara manual sebagai kategori negatif.

Tabel 8. Frekuensi Kemunculan kata pada Twitter

Word	Positive Class	Negative Class	Total
Official	34	72	106
Corruption	59	162	221
Tax	14	45	59
Country	21	35	56

1. Calculate Positive Class Value

$$= \left(\frac{OF1}{Nof} \times \frac{CR1}{Ncr} \times \frac{TA1}{Nta} \times \frac{CU1}{Ncu} \right) \times \frac{Npositive}{Npositive + Nnegative}$$

$$= \left(\frac{34}{106} \times \frac{59}{221} \times \frac{14}{59} \times \frac{21}{56} \right) \times \frac{128}{442} = 0.002.....(3)$$

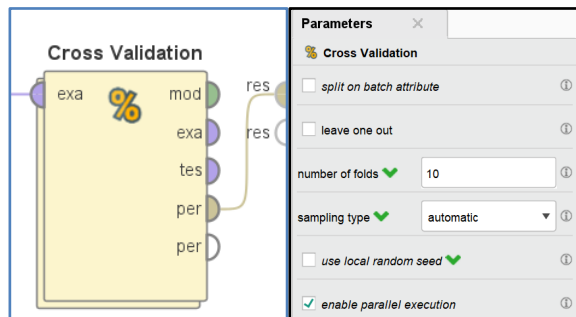
2. Calculate Negative Class Value

$$= \left(\frac{OF2}{Nof} \times \frac{CR2}{Ncr} \times \frac{TA2}{Nta} \times \frac{CU2}{Ncu} \right) \times \frac{Nnegative}{Npositive + Nnegative}$$

$$= \left(\frac{72}{106} \times \frac{162}{221} \times \frac{45}{59} \times \frac{35}{56} \right) \times \frac{314}{442} = 0.169.....(4)$$

G. Evaluation

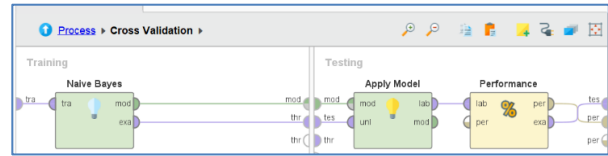
Peneliti melakukan uji dengan menggunakan operator 10-Fold Cross Validation. Tujuannya adalah untuk menghasilkan akurasi data dan klasifikasi serta mengevaluasi penerapan dari algoritma NBC. Gambar.10 menunjukkan parameter yang dimasukkan melalui operator tersebut yang digunakan dalam pengujian.



Gambar 10. Parameter pada operator Cross Validation

Setelah menetapkan parameter pada operator yang diperkukan, maka langkah selanjutnya yaitu menghubungkan data latih dan data uji ke algoritma

NBC agar menguji model tersebut yang dengan kata lain data latih yang diproses sebelumnya nantinya dilatih pada model NBC, sedangkan data uji digunakan untuk mengevaluasi model NBC tersebut. Gambar.11 menunjukkan evaluasi yang dilakukan pada RapidMiner dengan operator *Performance*.



Gambar 11. Tahapan Cross Validation pada RapidMiner

Setelah *running* pada RapidMiner untuk semua operator yang digunakan, maka didapatkan hasil pengukurannya. Gambar.12 berikut adalah tampilan dari hasil nilai akurasi terhadap implementasi algoritma NBC dengan pengujian menggunakan 10-Fold Cross Validation, seperti yang ditunjukkan pada Gambar.12 dibawah ini.

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 12. Hasil dari Accuracy dan Performance Vector pada RapidMiner

Berikut sampel untuk cara menghitung nilai akurasi secara manual, yaitu menggunakan rumus berikut:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \times 100\%$$

$$= \frac{190 + 212}{34 + 12 + 212 + 190} \times 100\%$$

$$= \frac{402}{448} \times 100\% = 89.73.....(5)$$

Terlihat bahwa hasil pengujian dengan metode 10-fold cross-validation pada RapidMiner, menghasilkan nilai akurasi sebesar 89,73%.

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

Gambar 13. Hasil dari Confusion Matrix

Gambar.13 menunjukkan hasil dari visualisasi *confusion matrix*. Berdasarkan hasil dari pengolahan, publik sangat kecewa dengan perilaku korup yang dilakukan oknum pemerintah khususnya yang memegang jabatan strategis. Sebagai masukan, pemerintah wajib mengembalikan kepercayaan publik salah satunya dengan menegakkan hukum secara tegas dan adil.

KESIMPULAN

Hasil penelitian terhadap pengolahan data yang berhasil dikumpulkan yaitu sebanyak 601 data, maka ditemukan bahwa analisis kepuasan masyarakat terhadap mayoritas menyatakan negatif kepada perilaku korupsi yang dilakukan oleh pejabat pemerintah. Hal ini dinyatakan dengan adanya temuan bahwa tweet dengan label sentimen positif hanya sejumlah 128, sedangkan mayoritas tweet dengan label sentimen negatif yaitu sebanyak 314. Dalam melakukan pembobotan TF-IDF, muncul tiga kata yang dominan dalam kepuasan masyarakat di media sosial yaitu kata “Corruption”, “Official” dan “Tax”. Hal ini membuktikan bahwa masyarakat menyatakan tidak puas dan kecewa. Sehingga hal ini mengakibatkan tingkat kepercayaan yang sangat rendah terhadap kinerja pemerintah khususnya pejabat pemerintah di sektor pajak. Implementasi algoritma *naive bayes classifier* setelah dilakukan evaluasi dengan menggunakan *10-fold cross-validation*, mendapatkan nilai akurasi 89,74% dengan komposisi *confusion matrix* untuk nilai *True Negative* sebesar 188, kemudian untuk nilai *False Negative* sebesar 14, kemudian untuk nilai *False Positive* sebesar 36, dan untuk nilai *True Positive* sebesar 210.

KONTRIBUSI PENULIS

Faldy Irwiensyah, membuat konsep penelitian, pengumpulan dan pengolahan data, penulisan dan revisi draf artikel. Firman Noor Hasan membantu dalam pengumpulan dan pre-processing data, penulisan pendahuluan, tinjauan pustaka, dan kesimpulan penelitian.

DAFTAR PUSTAKA

Afandi, I. R., Hasan, F. N., Rizki, A. A., Pratiwi, N., & Halim, Z. (2022). Analisis Sentimen Opini Masyarakat Terkait Pelayanan Jasa Ekspedisi Anteraja Dengan Metode Naive Bayes. *Jurnal Linguistik Komputasional*, 5(2), 63–70. <https://doi.org/https://doi.org/10.26418/jlk.v5i2.107>

- Alam, T., Aftab, M., Abbas, Z., Ugli, K. M. M., & Bokhari, S. A. A. (2023). Impact of E-Government Initiatives to Combat Corruption Mediating by Behavioral Intention: A Quantitative Analysis from Emerging Economies. *Sustainability*, 15(3). <https://doi.org/https://doi.org/10.3390/su15032694>
- Amini, T. A., & Setiawan, K. (2024). Application of the Naive Bayes Algorithm in Twitter Sentiment Analysis of 2024 Vice Presidential Candidate Gibran Rakabuming Raka using Rapidminer. *IJSECS: International Journal Software Engineering and Computer Science*, 4(1), 234–246. <https://doi.org/https://doi.org/10.35870/ijsecs.v4i1.2236>
- Andriyani, W., Natsir, F., Lubis, H., Tyas, S. H. Y., Meidelfi, D., Faizah, S., ... Hikmawati, E. (2024). *Perangkat Lunak Data Mining* (A. A. Soebroto & Y. Ramdhani, eds.). Retrieved from https://www.google.co.id/books/edition/PERAN_GKAT_LUNAK_DATA_MINING/bdouEQAAQBAJ?hl=id&gbpv=0
- Bani-Mustafa, A., Nimer, K., Uyar, A., & Schneider, F. (2024). Effect of Government Efficiency on Tax Evasion: The Mediating Role of Ethics and Control of Corruption. *International Journal of Public Administration*, 47(2).
- Bu, K., Liu, Y., & Ju, X. (2024). Efficient Utilization of Pre-trained Models: A Review of Sentiment Analysis via Prompt Learning. *Knowledge-Based System*, 283. <https://doi.org/https://doi.org/10.1016/j.knosys.2023.111148>
- Cantini, A., Peron, M., Carlo, F. De, & Sgarbossa, F. (2024). A Decision Support System for Configuring Spare Parts Supply Chains Considering Different Manufacturing Technologies. *International Journal of Production Research*, 62(8), 3023–3043. <https://doi.org/https://doi.org/10.1080/00207543.2022.2041757>
- Castro, C., & Lopes, I. C. (2023). E-Government as a Tool in Controlling Corruption. *International Journal of a Public Administration*, 46(16), 1137–1150. <https://doi.org/https://doi.org/10.1080/01900692.2022.2076695>
- Chiu, S.-Y., Hung, H.-W., Yang, C.-Y., Chen, C.-M., & Chiu, Y. (2025). Strategies to Control Corruption in Economic Development: The Role of Government Spending and Public Satisfaction. *Socio-Economic Planning Sciences*, 98(April). <https://doi.org/https://doi.org/10.1016/j.seps.2024>



- .102144
Das, R., & Singh, T. D. (2023). Multimodal Sentiment Analysis: A Survey of Methods, Trends, and Challenges. *ACM Computing Surveys*, 55(13), 01–38.
<https://doi.org/https://doi.org/10.1145/3586075>
- Fathurrohman, S., Afandi, I. R., & Hasan, F. N. (2024). Sentiment Analysis of TIMNAS Indonesia's Participation in the Asian Cup U23 2024 on X Using Naive Bayes and SVM. *IJID: International Journal on Informatics for Development*, 13(1), 434–447.
<https://doi.org/https://doi.org/10.14421/ijid.2024.4504>
- Hasan, F. N., & Ariyansah, R. (2024). Utilization of the FP-Growth Algorithm on MSME Transaction Data : Recommendations for Small Gifts from The Padang Region. *JTI: Jurnal Teknik Informatika*, 17(1), 70–78.
<https://doi.org/https://doi.org/10.15408/jti.v17i1.37966>
- Hasan, F. N., Aziz, A. S., & Nofendri, Y. (2023). Utilization of Data Mining on MSMEs using FP-Growth Algorithm for Menu Recommendations. *MATRIK: Jurnal Manajemen, Teknik Informatika, Dan Rekayasa Komputer*, 22(2), 261–270.
<https://doi.org/10.30812/matrik.v22i2.2166>
- Hasan, F. N., & Dwijayanti, M. (2021). Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naïve Bayes Classifier. *Jurnal Linguistik Komputasional*, 4(2), 52–58.
<https://doi.org/https://doi.org/10.26418/jlk.v4i2.61>
- Hasan, F. N., Munawar, Z., Wiranata, A. D., Indra, D., Kamdan, Sukmana, R. N., ... Sipahutar, L. (2024). *Teknologi Informasi: Konsep Dasar dan Aplikasi* (H. H. Solihin, B. Hardiyana, & C. K. Sastradipraja, eds.). Retrieved from <https://books.google.co.id/books/about?id=qVlgEQAAQBAJ>
- Hermawan, Y., & Thoyyibah, T. (2023). Comparison of K-Means Clustering Algorithm with Naive Bayes Classifier Algorithm to Know Students Reading Interest Using Library Data Mining. *JCOSBIDA: Journal of Computer Science and Big Data*, 1(2), 20–25. Retrieved from <https://jurnal.jcosbida.com/index.php/jcosbida/article/view/210>
- Kartiko, N. D. (2024). Does Government Effectiveness and Corruption Control Support Political Stability? *Integritas: Jurnal Anti Korupsi*, 10(1), 81–94.
<https://doi.org/https://doi.org/10.32697/integritas.v10i1.1209>
- Muflih, H. Z., Abdillah, A. R., & Hasan, F. N. (2023). Analisis Sentimen Ulasan Pengguna Aplikasi Ajaib Menggunakan Metode Naïve Bayes. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 4(3), 1613–1621.
<https://doi.org/https://doi.org/10.30865/klik.v4i3.1303>
- Paraswansa, A. D., & Utomo, D. C. (2024). Whistleblowing dan Korupsi Pada Sektor Publik: A Systematic Review. *JAA: Jurnal Akademi Akuntansi*, 7(1), 94–113.
<https://doi.org/https://doi.org/10.22219/jaa.v7i1.31336>
- Peretz, O., Koren, M., & Koren, O. (2024). Naive Bayes classifier – An ensemble procedure for recall and precision enrichment. *Engineering Applications of Artificial Intelligence*, 136.
<https://doi.org/https://doi.org/10.1016/j.engappai.2024.108972>
- Puspita, A. C., L, Y. M., & Gultom. (2024). The Effect of E-Procurement Policy on Corruption in Government Procurement: Evidence from Indonesia. *International Journal of Public Administration*, 47(2), 117–129.
<https://doi.org/https://doi.org/10.1080/01900692.2022.2093900>
- Putra, Y. W. S., Dawis, A. M., Novi, Natsir, F., Fitria, Windhiyanti, A. A. S., ... Maniah. (2023). *Pengantar Aplikasi Mobile* (Pertama; W. Andriyani & Erlangga, eds.). Retrieved from <https://penerbithaura.com/product/pengantar-aplikasi-mobile/>
- Rodríguez-Ibáñez, M., Casáñez-Ventura, A., Castejón-Mateos, F., & Cuenca-Jiménez, P.-M. (2023). A Review on Sentiment Analysis from Social Media Platforms. *Expert Systems with Applications*, 223.
<https://doi.org/https://doi.org/10.1016/j.eswa.2023.119862>
- Romano, M., Contu, G., Mola, F., & Conversano, C. (2024). Threshold-based Naïve Bayes Classifier. *Advances in Data Analytics and Classification*, 18, 325–361.
<https://doi.org/https://doi.org/10.1007/s11634-023-00536-8>
- Rosida, E., Firmansyah, A., & Suherman. (2024). Comparative Analysis of Classification of K-Nearest Neighbor (KNN) Algorithm and Decision Tree in Breast Cancer Using Rapidminer. *International Journal of Applied Research and Sustainable Sciences*, 2(12).
<https://doi.org/https://doi.org/10.59890/ijarss.v2i12.48>
- Sadik-Zadaa, E. R., Gatto, A., & Niftiyev, I. (2024). E-



- Government and Petty Corruption in Public Sector Service Delivery. *Technology Analysis & Strategic Management*, 36(12), 3987–4003. <https://doi.org/https://doi.org/10.1080/09537325.2022.2067037>
- Sani, A., Samuel, Suryadi, D., Hasan, F. N., Wiranata, A. D., & Aisyah, S. (2023). Predicting the Success of Garment Sales on Transaction Data using the Classification Method with the Naïve Bayes Algorithm. *ICCoSITE: 2023 International Conference on Computer Science, Information Technology and Engineering*, 234–239. <https://doi.org/10.1109/ICCoSITE57641.2023.10127693>
- Septiana, D. W., & Pastika, P. B. (2024). Classification of Travel Class with K-Nearest Neighbors Algorithm Using Rapidminer. *JOSRE: Journal of Student Research Exploration*, 2(2), 88–100. <https://doi.org/https://doi.org/10.52465/josre.v2i2.357>
- Steven, K., Hariyanto, S., Arijanto, R., & Wijaya, A. H. (2021). Penerapan Business Intelligence Untuk Menganalisis Data Pada PT.Suryaplas Intitama Menggunakan Microsoft Power BI. *Jurnal ALGOR*, 2(2), 41–50. <https://doi.org/https://doi.org/10.31253/algor.v2i2.550>
- Suhud, R., Febriandirza, A., Permatasari, I., & Ramadhan, F. (2023). Recognizing Public Satisfaction Toward Kampus Mengajar Program with Naive Bayes. *IC3INA: 2023 International Conference on Computer, Control, Informatics and Its Applications*. <https://doi.org/10.1109/IC3INA60834.2023.10285769>
- Syakir, A., & Hasan, F. N. (2023). Analisis Sentimen Masyarakat Terhadap Perilaku Korupsi Pejabat Pemerintah Berdasarkan Tweet Menggunakan Naive Bayes Classifier. *MIB: Media Informatika Budidarma*, 7(4), 1796–1805. <https://doi.org/http://dx.doi.org/10.30865/mib.v7i4.6648>
- Wanga, Z., Guan, X., Zeng, Y., Liang, X., & Donga, S. (2024). Utilizing Data Platform Management to Implement “5W” Analysis Framework for Preventing and Controlling Corruption in Grassroots Government. *Heliyon*, 10(7). <https://doi.org/10.1016/j.heliyon.2024.e28601>

