



# **Analisis Sentimen Masyarakat Terhadap Rencana Kenaikan PPN 12% Di Indonesia Pada Media Sosial X Menggunakan Metode Decision Tree**

**Intan Diah Hardyatman, Firman Noor Hasan\***

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta  
Jl. Tanah Merdeka No.20, Rambutan, Kec. Ciracas, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia

Email: <sup>1</sup>intandyatan5@gmail.com, <sup>2,\*</sup>firman.noorhasan@uhamka.ac.id

Email Penulis Korespondensi: firman.noorhasan@uhamka.ac.id

Submitted: 30/12/2024; Accepted: 14/01/2025; Published: 15/01/2025

**Abstrak**—Penelitian ini menganalisis sentimen masyarakat terhadap rencana kenaikan Pajak Pertambahan Nilai (PPN) menjadi 12% di Indonesia menggunakan data dari media sosial X. Kenaikan PPN dapat memicu peningkatan belanja di luar negeri serta kenaikan harga produk dan jasa di Indonesia, yang berpotensi menurunkan penjualan dan melemahkan industri. Usulan ini juga mendapat perhatian luas di media sosial X. Rencana kenaikan PPN memiliki pro dan kontra, sehingga memicu banyak diskusi di media sosial. Metode klasifikasi Decision Tree digunakan untuk mengolah data yang diperoleh melalui proses crawling dan preprocessing teks. Penelitian ini membandingkan 80% data latih dan 20% data uji yang terdiri dari 1000 data, dengan rincian 285 sentimen negatif dan 715 sentimen positif dalam dataset. Dalam hal ini dapat digambarkan bahwa pengguna media sosial X terhadap rencana kenaikan PPN 12% di Indonesia cenderung positif. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap rencana kenaikan PPN 12% di Indonesia menggunakan Decision Tree dan mengidentifikasi faktor-faktor yang memengaruhi sentimen tersebut. Hasil analisis menunjukkan bahwa Decision Tree berhasil meningkatkan akurasi sebesar 81,34% klasifikasi sentimen dibandingkan metode sebelumnya, seperti Naïve Bayes dengan tingkat akurasi 63,1%. Hasil dari penelitian ini diharapkan dapat membantu pemerintah dalam kebijakan fiskal yang lebih responsif.

**Kata Kunci:** Analisis Sentimen; PPN 12%; Rapidminer; Decision Tree; Media Sosial

**Abstract**—This study analyzes public sentiment towards the planned increase of Value Added Tax (VAT) to 12% in Indonesia using data from X social media. The VAT hike could trigger an increase in overseas spending and higher prices for products and services in Indonesia, potentially reducing sales and weakening industries. This proposal also received widespread attention on social media X. The VAT increase plan has pros and cons, triggering many discussions on social media. The Decision Tree classification method was used to process the data obtained through crawling and text preprocessing. This research compares 80% training data and 20% test data consisting of 1000 data, with details of 285 negative sentiments and 715 positive sentiments in the dataset. In this case, it can be described that X social media users towards the plan to increase VAT by 12% in Indonesia tend to be positive. This research aims to analyze people's sentiment towards the plan to increase VAT by 12% in Indonesia using Decision Tree and identify factors that influence the sentiment. The results of the analysis show that Decision Tree succeeded in increasing the accuracy by 81.34% of sentiment classification compared to previous methods, such as Naïve Bayes with an accuracy rate of 63.1%. The results of this study are expected to help the government in a more responsive fiscal policy.

**Keywords:** Sentiment Analysis; 12% VAT; Rapidminer; Decision Tree; Social Media

## **1. PENDAHULUAN**

Kegiatan perdagangan meningkat di semua sektor seiring dengan perkembangan ekonomi global. Kegiatan perdagangan selalu terkait dengan pemenuhan kebutuhan suatu negara dalam hubungannya dengan negara lain. sehingga dapat memulai kegiatan ini dan berkontribusi pada pendapatan negara. Setiap barang komersial harus dikenakan pajak di suatu negara. Salah satu cara paling efektif bagi negara untuk mendanai program kerja dan tujuan pemerintah adalah melalui perpajakan [1].

Pajak Pertambahan Nilai (PPN) adalah pajak yang diterapkan pada penjualan barang dan jasa di Indonesia, yang bertujuan untuk meningkatkan pendapatan negara guna mendanai proyek Pembangunan nasional dan memperbaiki layanan publik [2].

Undang-Undang Harmonisasi Peraturan Perpajakan mengenai klaster PPN di Pasal 7 ayat (1) menyebutkan, tarif PPN 10% akan naik menjadi 11% pada tanggal 1 April 2022. Kemudian, pemerintah diwajibkan menaikkan tarif PPN sekali lagi menjadi 12% paling lambat pada 1 Januari 2025. Menurut pemerintah, tujuan menaikkan tarif PPN adalah untuk memaksimalkan penerimaan negara, yang diharapkan dapat mendukung pemulihan ekonomi dan pembangunan [1].

Namun, Rencana kenaikan PPN menjadi 12% di Indonesia telah memicu beragam reaksi dari masyarakat, mulai dari kekhawatiran penurunan daya beli hingga potensi dampak negatif terhadap perekonomian. Pedagang dan pengusaha khususnya, mengungkapkan keprihatinan atas penurunan daya beli jika tarif tersebut diberlakukan pada tahun 2025 [3].

Salah satu data teks tersebut diperoleh dari aplikasi X, yaitu (Twitter, sebuah platform media sosial yang memungkinkan pengguna untuk mengekspresikan pendapat mereka secara publik) [4]. Aplikasi Twitter dipilih karena (memiliki basis pengguna yang sangat besar di Indonesia, memungkinkan pengambilan data public tweets yang relevan dengan topik penelitian, dan memiliki API yang memungkinkan pengambilan data terstruktur) [5].

Data sentimen dikumpulkan dengan cara (jelaskan metode pengumpulan data, misalnya: menggunakan Twitter API untuk mengambil tweets yang mengandung kata kunci tertentu terkait rencana kenaikan PPN 12%, kemudian dilakukan preprocessing data untuk membersihkan noise dan stopwords sebelum dilakukan analisis sentimen) [6].

Peneliti sebelumnya yang sebanding yaitu menggunakan algoritma Naïve Bayes untuk analisis sentimen dilakukan dengan Eva Darwisah dan Rakhmat Kurniawan [7]. Penelitian dilakukan terhadap implementasi dari kebijakan pemerintah mengenai Tabungan Perumahan Rakyat (TAPERA). Jumlah data yang diperoleh sebanyak 1210 data tweet. Hasil accuracy sebesar 0.96%, precision sebesar 0,93%, recall sebesar 0,93% dan f1-score sebesar 0,97%.

Penelitian sebelumnya yang serupa yang dilakukan oleh Arya Adi Restu Putra Pratama, Bayu Rizky Utomo, Naufal Hanan Jati Asmara, Arig Kusuma Jati, dan Rajnaparamitha Kusumastuti. Dengan menggunakan metode Decision Tree penelitian ini tentang analisis sentimen ulasan maskapai penerbangan virgin america. Dataset diambil dan berhasil mengumpulkan 900 data melalui proses scraping pada Kaggle. Hasil pada metode Decision Tree menyatakan tingkat akurasi sebesar 52,18% [8].

Adapun penelitian sebelumnya yang serupa mengenai algoritma Decision Tree [9] yang dilakukan oleh Aslam Fatkhudin, Fenelinas Adi Artanto, Naufal Abiyu Saflil dan Dimas Wibowo menggunakan metode Decision Tree untuk analisis sentimen penggunaan Artificial Intelligence skripsi mahasiswa dengan rentang waktu 30-10-2020 sampai 30-10-2023 mendapatkan 238 data. Data yang telah dibersihkan menjadi 50 data lalu dikelompokkan menjadi 2 kelompok yaitu sentimen positif dan negatif. Setelah diklasifikasi terdapat 84.4% sentimen negatif dan 15.6% sentimen positif.

Penelitian yang serupa Jessica Kristovani Siagian dan Painem menggunakan pendekatan Naïve Bayes untuk menganalisis opini masyarakat Indonesia terkait usulan kenaikan PPN sebesar 12% di media sosial X [1]. Dari 468 dataset yang dikumpulkan antara 1 Maret hingga 15 Mei 2024, model ini mencapai akurasi 83%, presisi 68,8%, dan recall 78,6%.

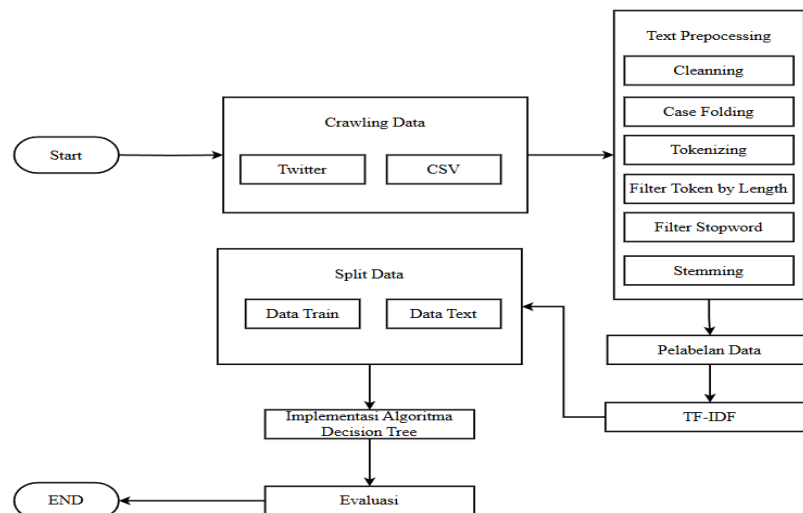
Adapun penelitian yang sebelumnya yang dilakukan oleh Ihsan Zulfahmi menggunakan metode Decision Tree menganalisis sentimen pada aplikasi PLN Mobile. Berdasarkan hasil dari evaluasi, memproses 10206 baris data dan mengklasifikasinya menggunakan metode Decision Tree menunjukkan akurasi sebesar 96%, presisi sebesar 91%, recall sebesar 96%, dan skor f1 sebesar 93% [10].

Penelitian sebelumnya menunjukkan bahwa metode seperti Naive Bayes, meskipun menghasilkan analisis sentimen yang cukup baik, memiliki keterbatasan dalam memahami secara mendalam opini publik terhadap kebijakan pemerintah, termasuk rencana kenaikan PPN menjadi 12% di Indonesia. Oleh karena itu, penelitian ini menggunakan metode Decision Tree untuk memberikan penilaian yang lebih komprehensif tentang sentimen Masyarakat. Hasil penelitian diharapkan mampu memberikan gambaran yang jelas mengenai opini pengguna platform X terhadap rencana kebijakan tersebut dan menjadi masukan berharga bagi pemerintah dalam merumuskan kebijakan.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Algoritma Decision Tree digunakan untuk melakukan analisis sentimen dalam penelitian ini dengan RapidMiner terdapat beberapa tahapan. Beberapa tahapannya sebagai berikut:



**Gambar 1.** Alur Penelitian

Berikut adalah alur pada penelitian ini yang terdapat pada Gambar 1:



- a. Crawling data: tahap pertama mengumpulkan data dari platform media sosial X yang berisi postingan dan tweet berkaitan dengan topik penelitian. Setelah itu, data yang terkumpul akan disimpan dalam format tertentu seperti file CSV [11].
- b. Text preprocessing: pada proses ini melibatkan tahapan preprocessing data yang diterima untuk membersihkan dan menghilangkan karakter yang mengganggu. Tahapan tersebut yaitu cleaning, case folding, tokenizing, filter token by length, filter stopwords removal, dan stemming [12].
- c. Pelabelan data: selanjutnya, setelah mendapatkan data mentah adalah proses pemberian label sentimen tertentu pada setiap bagian data dalam dataset agar dapat digunakan untuk pengujian, penilaian atau pelatihan model [13].
- d. Setelah membagi data ke dalam subset pelatihan dan pengujian Smote Upsampling digunakan untuk mengatasi ketidakseimbangan kelas, selanjutnya menerapkan algoritma Decision Tree [14].
- e. Tahapan akhir, yaitu melakukan evaluasi untuk mengklasifikasikan polaritas sentimen teks. Dengan menggunakan data pelatihan, teknik ini menciptakan struktur seperti pohon dengan tujuan memastikan model yang digunakan menghasilkan yang akurat dan relevan [15].

### 2.2 Text Preprocessing

Preprocessing merupakan proses yang bertujuan untuk membersihkan dan menghilangkan karakter yang mengganggu [16]. Gangguan ini dapat menghambat efektivitas pembobotan sehingga membuat kategorisasi data yang akurat menjadi tidak tepat. Terdapat bebarap tahap yang dilakukan:

- a. Cleaning : menghilangkan noise dalam bentuk simbol, termasuk karakter @#%&^&\*( ), mention"@", dan hastag "#" [16].
- b. Case Folding : mengganti kalimat pada dokumen menjadi huruf kecil.
- c. Tokenizing : menggunakan hasil pemrosesan case folding untuk membagi kalimat menjadi kata-kata berdasarkan spasi [17].
- d. Filter Token by Length : mengilangkan kalimat dengan karakter lebih sedikit dari empat karakter [18].
- e. Filter Stopwords Removal : menghilangkan kata penghubung seperti tetapi, dengan, untuk, yang, dan kata penghubung lainnya yang bukan merupakan istilah atau kata penghubung yang relevan [19].
- f. Stemming : mengubah kata menjadi bentuk yang paling mendasar [20].

### 2.3 Decision Tree

Decision Tree adalah struktur data berbasis node dan rusuk yang terdiri dari tiga jenis node: akar, cabang/internal, dan daun. Sebagai metode klasifikasi sederhana untuk sejumlah kecil kelas, node akar dan cabang diberi label sesuai atribut, rusuk diberi label dengan nilai atribut, dan node daun merepresentasikan kelas yang diklasifikasikan. [21].

$$\text{Entropy (S)} = \sum_{i=1}^n -p * \log_2 p_i \tag{1}$$

### 2.4 Confusion Matrix

Confusion Matrix merupakan wawasan tentang menjelaskan berbagai jenis kesalahan yang dilakukan oleh model. Ketika penilaian ini digabungkan, evaluasi menawarkan gambaran menyeluruh tentang seberapa baik kinerja model dalam mengidentifikasi perasaan, membantu dalam pengambilan keputusan yang tepat [22]. Pada titik ini, nilai akurasi, presisi, dan recall dipertimbangkan dalam evaluasi yang mengukur hasil kategorisasi. Confusion Matriks yang didasarkan pada empat variabel utama digunakan dalam pendekatan evaluasi ini. Nilai-nilai ini adalah true positif rate (TP rate), true negatif rate (TN rate), false positif rate (FP rate), dan false negatif rate (FN rate) [23]. Berikut ini adalah Tabel 1, yang menunjukkan Confusion Matrix:

**Tabel 1.** Confusion Matrix

	Prediksi	
Aktual	TRUE	FALSE
TRUE	TP	FP
FALSE	FN	TN

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{3}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{4}$$

Dalam evaluasi Confusion Matrix, True Positive (TP) mengacu pada data positif yang berhasil diprediksi dengan benar, sedangkan False Positive (FP) adalah data negatif namun diprediksi sebagai data positif. Sebaliknya, False Negative (FN) merupakan data positif namun diprediksi sebagai data negatif, dan True Negative (TN) adalah

data negatif yang diprediksi benar. Keempat matrix ini digunakan untuk mengevaluasi performa model prediksi, terutama dalam klasifikasi, dengan menghitung akurasi, presisi, dan recall [20].

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Pengumpulan Data

Mengambil data menggunakan teknik web scrapping yang digunakan untuk mengumpulkan data dari media sosial X dengan tujuan untuk mengestraknya. Proses ekstrak data dimulai dengan mengakses media sosial X untuk mendapatkan auth token akun twitter yang dimasukkan kedalam Google Colab dan hasilnya disimpan sebagai file CSV (Comma Separated Value). Dapat dilihat pada Gambar 2, ilustrasi dari proses ekstrak data:



Gambar 2. Proses Ekstrak Data

Bahasa pemrograman python digunakan untuk mengumpulkan data penelitian ini melalui teknik crawling, data diambil dengan limit 1000 tweet dengan keyword ppn12% dan dibatasi dengan tweet yang menggunakan bahasa Indonesia. Skrip berikut akan mengumpulkan tweet dari X berdasarkan sesuai pencarian, kerangka waktu, dan batasan jumlah yang ditentukan. Gambar 3 menunjukkan hasil data crawling dari media sosial X:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
977	Tukar guling rakyat dapat bansos pemerintah dapat kenaikan PPN 12%. Hemmmm.....tetep cuan?!	1,869E+18	2024-12-17 09:53:05+	Ascanc155885	1,72E+18	in	0	0	0	0	0	0	0	<a href="https://x.com/Ascanc155885/status/1868957580">https://x.com/Ascanc155885/status/1868957580</a>
978	Ilustrasi sederhana nya si Varel kalo suka mandi pake Aqua mau naik 12% aja dia ttp mandi pake aqua	1,869E+18	2024-12-17 09:53:08+	aditaleee	1,15E+18	in	0	0	0	0	0	0	0	<a href="https://x.com/aditaleee/status/1868957595262">https://x.com/aditaleee/status/1868957595262</a>
979	2025 meningkatkan kesejahteraan 2025 meningkatkan ppn 12%	1,869E+18	2024-12-17 09:53:10+	madefromlove	1,33E+18	in	0	0	1	1	0	0	0	<a href="https://x.com/madefromlove/status/186895761">https://x.com/madefromlove/status/186895761</a>
980	Hal paling lucu dari PPN naik 12% adalah ngeliat voter 02 yg KAGET (especially rakyat biasa) akan keb	1,869E+18	2024-12-17 09:53:30+	rudauli	1,27E+18	in	0	0	0	1	1	0	0	<a href="https://x.com/rudauli/status/18689576876249">https://x.com/rudauli/status/18689576876249</a>
981	12% Gw masih ga nyangka aja ppn segede itu di tengah gaji seminin ini.	1,869E+18	2024-12-17 09:53:35+	chonko_lolo	1,49E+18	in	0	0	0	1	1	0	0	<a href="https://x.com/chonko_lolo/status/18689577051">https://x.com/chonko_lolo/status/18689577051</a>
982	@tanvarfes Jasa pendidikan premium itu kyk gmn? Keadaan normal aja inflasi biaya pendidikan naik	1,869E+18	2024-12-17 09:53:36+	Hollyne_Retha	1,48E+18	in	0	0	0	0	0	0	0	<a href="https://x.com/Hollyne_Retha/status/1868957711">https://x.com/Hollyne_Retha/status/1868957711</a>
983	ya Allah bingung bgt mau nangsini diri sendiri krn baru dijahatin org yg paling gue sayang apa nangsini	1,869E+18	2024-12-17 09:53:42+	levicornus	1,25E+18	in	0	0	0	0	0	0	0	<a href="https://x.com/levicornus/status/1868957735620">https://x.com/levicornus/status/1868957735620</a>
984	@tempodotco Bukannya harus berbanding lurus juga dengan pendapat rata-rata warganya ya? Rasa	1,869E+18	2024-12-17 09:53:46+	naufal_2828	1,47E+18	in	0	0	0	0	0	0	0	<a href="https://x.com/naufal_2828/status/18689577510">https://x.com/naufal_2828/status/18689577510</a>
985	sedang memikirkan narasi agar konten tentang kenaikan PPN 12% ini masuk ke otak warga tiktok	1,869E+18	2024-12-17 09:54:08+	RUMROUEN	2,89E+09	in	0	0	0	0	0	0	0	<a href="https://x.com/RUMROUEN/status/18689578468">https://x.com/RUMROUEN/status/18689578468</a>
986	gmn ini ppn 12% sinting lo ya	1,869E+18	2024-12-17 09:54:16+	idkwhtodoo	1,26E+18	in	0	0	0	0	0	0	0	<a href="https://x.com/idkwhtodoo/status/18689578804">https://x.com/idkwhtodoo/status/18689578804</a>
987	ehh ini klo ppn 12% nnti bakso mas harno naik dong harganyaaaa ??? damn pdhl itu bakso terenak dan	1,869E+18	2024-12-17 09:54:19+	floufyjyuan	1,56E+18	in	0	0	0	0	0	0	0	<a href="https://x.com/floufyjyuan/status/18689578908">https://x.com/floufyjyuan/status/18689578908</a>
988	Nih ya dengan naiknya ppn 12% harga barang pasti naik penjual pasti akan menaikkan harga dagangan	1,869E+18	2024-12-17 09:54:19+	mingyoori	1,26E+18	in	0	1	0	1	0	0	0	<a href="https://x.com/mingyoori/status/186895789953">https://x.com/mingyoori/status/186895789953</a>
989	- naiknya harga layanan bikin fanatik kesukaan kalian kalau jual keychain memungkinkan untuk mer	1,869E+18	2024-12-17 09:54:20+	mingyoori	1,26E+18	in	0	0	0	1	1	0	0	<a href="https://x.com/mingyoori/status/1868957893569">https://x.com/mingyoori/status/1868957893569</a>
990	yaelah ngapain sih pake ppn 12% segala moonyttttt	1,869E+18	2024-12-17 09:54:30+	kloroform	1,53E+18	in	0	0	0	0	0	0	0	<a href="https://x.com/_kloroform/status/186895793669">https://x.com/_kloroform/status/186895793669</a>
991	ORGLI ppn 12% dikira ga gede????	1,869E+18	2024-12-17 09:54:38+	ekueuuuu	1,84E+18	in	0	0	0	0	0	0	0	<a href="https://x.com/ekueuuuu/status/1868957972615">https://x.com/ekueuuuu/status/1868957972615</a>
992	ppn 12% saatnya stop ngopi gak sih (ngetik sambil nangis perlu asupan kopi)	1,869E+18	2024-12-17 09:54:39+	lopmoons	1,31E+18	in	0	0	0	0	0	0	0	<a href="https://x.com/lopmoons/status/1868957975832">https://x.com/lopmoons/status/1868957975832</a>
993	Natal dan Tahun Baru makin seru. Telkomel hadirkan jaringan stabil untuk temani setiap momen sp	1,869E+18	2024-12-17 09:54:50+	KabarNetral	1,79E+18	in	0	0	0	0	0	0	0	<a href="https://x.com/KabarNetral/status/18689580217">https://x.com/KabarNetral/status/18689580217</a>
994	@13NDD Harga sepatu sebelum pajak 2 juta include PPN 11% jadi 2 997 000 dan klo PPN 12% ja	1,869E+18	2024-12-17 09:54:56+	im_notslim	1,24E+18	in	0	0	0	0	0	0	0	<a href="https://x.com/im_notslim/status/186895804454">https://x.com/im_notslim/status/186895804454</a>
995	bismillah ppn 12% batam	1,869E+18	2024-12-17 09:54:58+	akupiyooou	1,68E+18	in	0	1	0	0	0	0	0	<a href="https://x.com/akupiyooou/status/18689580538">https://x.com/akupiyooou/status/18689580538</a>
996	2. PPN 12% Dilakukan Secara Bertahap kebijakan ini tidak dilakukan secara mendadak melainkan ber	1,869E+18	2024-12-17 09:55:04+	sulastri65481	4,44E+09	in	0	1	0	0	0	0	0	<a href="https://x.com/sulastri65481/status/1868958079">https://x.com/sulastri65481/status/1868958079</a>
997	Terjawab sudah PPN 12 persen kapan mulai berlaku cek juga golongan warga yang dapat diskon tarif	1,869E+18	2024-12-17 09:55:13+	tribunakalim	1,04E+08	in	0	0	0	0	0	0	0	<a href="https://x.com/tribunakalim/status/18689581177">https://x.com/tribunakalim/status/18689581177</a>
998	@Heraleobis Pulangin. Masuk orderan pas PPN naik 12%	1,869E+18	2024-12-17 09:55:40+	emfahmidil	3,47E+08	in	0	0	0	0	0	0	0	<a href="https://x.com/emfahmidil/status/186895823268">https://x.com/emfahmidil/status/186895823268</a>
999	Lebih ngabisin energi liet berita PPN naik 12% dan kebrokrokan rezim sih gelud war cecinaan gak ngar	1,869E+18	2024-12-17 09:55:46+	myangzi	1,29E+18	in	0	0	0	0	0	0	0	<a href="https://x.com/myangzi/status/1868958257188">https://x.com/myangzi/status/1868958257188</a>
1000	@dear_woobin iya kan.. harga pada naik tpi daya beli menurun eh malah mau naikin ppn 12% tuh lo	1,869E+18	2024-12-17 09:55:54+	arisasahara	9,28E+17	in	0	0	0	0	0	0	0	<a href="https://x.com/arisasahara/status/186895828898">https://x.com/arisasahara/status/186895828898</a>
1001	Vietnam: 8% tetapi sembako dikenakan pajak 5% Filipina: 12% India: 18% (untuk sebagian barang dai	1,869E+18	2024-12-17 09:56:24+	sulastri65481	4,44E+09	in	0	0	0	0	0	0	0	<a href="https://x.com/sulastri65481/status/1868958415">https://x.com/sulastri65481/status/1868958415</a>

Gambar 3. Hasil Crawling Data

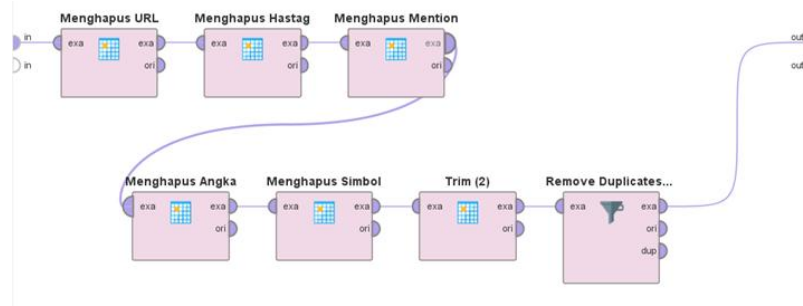
Untuk membaca data tweet dari format file CSV, memilih dan mengganti nama kolom, mengubah tipe data tanggal, mengurutkan data berdasarkan tanggal, menyimpan data yang telah diolah ke file CSV baru dan menampilkan data akhir.

#### 3.2 Preprocessing Data

Selanjutnya tahap text preprocessing setelah pengumpulan data sebanyak 1000 data. Proses ini mempersiapkan data teks untuk analisis lebih lanjut dengan melakukan pembersihan dan transformasi data. Ada beberapa langkah yang digunakan untuk melakukan text preprocessing, diantaranya yaitu:

##### a. Cleaning

Langkah pertama dalam proses cleaning yaitu menghapus elemen yang tidak relevan atau tidak penting. Elemen yang dibersihkan yaitu URL, Hastag, Mention, Angka, Simbol, Trim (menghapus spasi yang tidak dibutuhkan di awal dan akhir teks), serta Duplicates (menghapus salinan teks dalam dataset). Berikut ini Gambar 4 dari proses Cleaning Data :



**Gambar 4.** Proses Cleaning Data

Hasil dari proses cleaning data dapat dilihat pada Tabel 2.

**Tabel 2.** Hasil dari proses Cleaning Data

Sebelum	Sesudah
@kolamkopi @AnggaPutraF kamu tau tidak bahkan kenaikan pajak 12% sudah ditandatangani pada saat yg sama ketika pajak kita dinaikkan 11%. ada di pasal 7 ayat 1 UU PPN 11% per 1 april 2024 12% per 1 januari 2025. bisa di cek sendiri. anehnya kenapa DPR menyetujui wakil rakyat kok tdk menyuarakan rakyat	kamu tau tidak bahkan kenaikan pajak sudah ditandatangani pada saat yg sama ketika pajak kita dinaikkan ada di pasal ayat UU PPN per april per januari bisa di cek sendiri anehnya kenapa DPR menyetujui wakil rakyat kok tdk menyuarakan rakyat

Pada tahap cleaning processing, teks yang tidak relevan atau mengganggu analisis dihapus untuk memfokuskan data pada informasi yang lebih bermakna. Proses pembersihan ini mencakup penghapusan elemen-elemen seperti URL, hashtag, mention, angka, simbol, serta spasi berlebih (trim). Selain itu, salinan duplikat teks juga dihapus untuk mencegah bias dan memastikan keberagaman data dalam dataset.

b. Case Folding

Case Folding menjadi langkah selanjutnya, untuk menjaga konsistensi struktur teks semua huruf diubah menjadi huruf kecil. Hasil dari proses case folding dapat dilihat pada Tabel 3.

**Tabel 3.** Hasil dari Proses Case Folding

Sebelum	Sesudah
Kamu tau tidak bahkan kenaikan pajak sudah ditandatangani pada saat yg sama ketika pajak kita dinaikkan ada di pasal ayat UU PPN per april per januari bisa di cek sendiri anehnya kenapa DPR menyetujui wakil rakyat kok tdk menyuarakan rakyat	kamu tau tidak bahkan kenaikan pajak sudah ditandatangani pada saat yg sama ketika pajak kita dinaikkan ada di pasal ayat uu ppn per april per januari bisa di cek sendiri anehnya kenapa dpr menyetujui wakil rakyat kok tdk menyuarakan rakyat

Pada tahap Case Folding, perubahan yang terjadi adalah konversi semua huruf menjadi huruf kecil untuk memastikan konsistensi dan standar format teks.

c. Tokenizing

Pada tahap tokenizing, kata-kata dipecah menjadi bagian-bagian yang lebih kecil dan dipisahkan menggunakan spasi. Hasil proses tokenizing dapat dilihat pada Tabel 4.

**Tabel 4.** Hasil dari Proses Tokenizing

Sebelum	Sesudah
kamu tau tidak bahkan kenaikan pajak sudah ditandatangani pada saat yg sama ketika pajak kita dinaikkan ada di pasal ayat uu ppn per april per januari bisa di cek sendiri anehnya kenapa dpr menyetujui wakil rakyat kok tdk menyuarakan rakyat	['kamu', 'tau', 'tidak', 'bahkan', 'kenaikan', 'pajak', 'sudah', 'ditandatangani', 'pada', 'saat', 'yg', 'sama', 'ketika', 'pajak', 'kita', 'dinaikkan', 'ada', 'di', 'pasal', 'ayat', 'uu', 'ppn', 'per', 'april', 'per', 'januari', 'bisa', 'di', 'cek', 'sendiri', 'anehnya', 'kenapa', 'dpr', 'menyetujui', 'wakil', 'rakyat', 'kok', 'tdk', 'menyuarakan', 'rakyat']

Proses tokenizing mengubah teks yang semula merupakan satu kesatuan kalimat panjang menjadi kumpulan kata-kata yang lebih terstruktur, memungkinkan analisis lebih lanjut terhadap setiap kata atau token.

d. Filtering (Stopwords Removal)

Pada tahap filtering (stopwords removal) guna menghilangkan kalimat umum yang dianggap tidak berguna. Hasil proses filtering (stopwords removal) dapat dilihat pada Tabel 5.

**Tabel 5.** Hasil dari Proses Filtering (stopwords Removal)

Sebelum	Sesudah
---------	---------

['kamu', 'tau', 'tidak', 'bahkan', 'kenaikan', 'pajak', 'sudah', 'ditandatangani', 'pada', 'saat', 'yg', 'sama', 'ketika', 'pajak', 'kita', 'dinaikkan', 'ada', 'di', 'pasal', 'ayat', 'uu', 'ppn', 'per', 'april', 'per', 'januari', 'bisa', 'di', 'cek', 'sendiri', 'anehnya', 'kenapa', 'dpr', 'menyetujui', 'wakil', 'rakyat', 'kok', 'tdk', 'menyuarakan', 'rakyat']

['tau', 'kenaikan', 'pajak', 'ditandatangani', 'yg', 'pajak', 'dinaikkan', 'pasal', 'ayat', 'uu', 'ppn', 'april', 'januari', 'cek', 'sendiri', 'anehnya', 'dpr', 'menyetujui', 'wakil', 'rakyat', 'tdk', 'menyuarakan', 'rakyat']

Dengan menghapus stopwords, teks menjadi lebih fokus pada kata-kata yang mengandung makna dan esensi dari pesan yang ingin disampaikan,

e. Filter Token by Length

Pada tahap filter token by length, beberapa kata dengan kurang dari empat karakter dan lebih dari dua puluh empat karakter akan dilakukan penghapusan oleh filter token by length. Hasil dari proses filter token by length dapat dilihat pada Tabel 6.

**Tabel 6.** Hasil dari Proses Filter Token by Length

Sebelum	Sesudah
['kamu', 'tau', 'tidak', 'bahkan', 'kenaikan', 'pajak', 'sudah', 'ditandatangani', 'pada', 'saat', 'yg', 'sama', 'ketika', 'pajak', 'kita', 'dinaikkan', 'ada', 'di', 'pasal', 'ayat', 'uu', 'ppn', 'per', 'april', 'per', 'januari', 'bisa', 'di', 'cek', 'sendiri', 'anehnya', 'kenapa', 'dpr', 'menyetujui', 'wakil', 'rakyat', 'kok', 'tdk', 'menyuarakan', 'rakyat']	['tau', 'kenaikan', 'pajak', 'ditandatangani', 'yg', 'pajak', 'dinaikkan', 'pasal', 'ayat', 'uu', 'ppn', 'april', 'januari', 'cek', 'sendiri', 'anehnya', 'dpr', 'menyetujui', 'wakil', 'rakyat', 'tdk', 'menyuarakan', 'rakyat']

Pada proses Filter Token by Length, kata-kata dengan panjang kurang dari empat karakter (seperti "kamu," "tidak," dan "ada") dihapus, sementara kata-kata yang relevan dengan panjang karakter antara empat hingga dua puluh empat karakter dipertahankan.

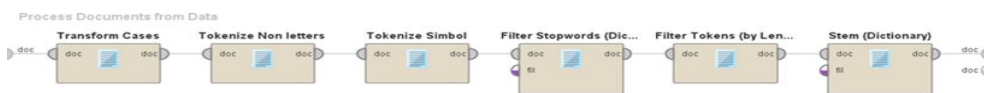
f. Stemming

Pada tahap stemming dengan menggunakan operator stem, stemming mengubah kata dengan imbuhan menjadi kata dasar. Dapat dilihat pada Tabel 7 hasil dari proses stemming.

**Tabel 7.** Hasil dari Proses Stemming

Sebelum	Sesudah
['tau', 'kenaikan', 'pajak', 'ditandatangani', 'yg', 'pajak', 'dinaikkan', 'pasal', 'ayat', 'uu', 'ppn', 'april', 'januari', 'cek', 'sendiri', 'anehnya', 'dpr', 'menyetujui', 'wakil', 'rakyat', 'tdk', 'menyuarakan', 'rakyat']	['tau', 'naik pajak', 'ditandatangani', 'yg', 'pajak', 'naik', 'pasal', 'ayat', 'uu', 'ppn', 'april', 'januari', 'cek', 'sendiri', 'aneh', 'dpr', 'tju', 'wakil', 'rakyat', 'tdk', 'suara', 'rakyat']

Pada tahap stemming, kata-kata dengan imbuhan atau bentuk variatif diubah menjadi bentuk dasar untuk menyederhanakan teks. Misalnya, "kenaikan" menjadi "naik" dan "dinaikkan" menjadi "naik". Imbuhan yang tidak diperlukan dihilangkan, seperti "menyuarakan" menjadi "suara", dan kata slang seperti "anehnya" disederhanakan menjadi "aneh". Proses ini mengurangi kompleksitas dan memastikan konsistensi kata, sehingga teks lebih mudah dianalisis. Setelah melalui tahap preprocessing, tahapan tersebut digabungkan ke dalam satu operator yang disebut 'Document from Data' di dalam aplikasi RapidMiner, seperti yang diilustrasikan pada Gambar 5.



**Gambar 5.** Proses Text Preprocessing

**3.3 Pelabelan Data**

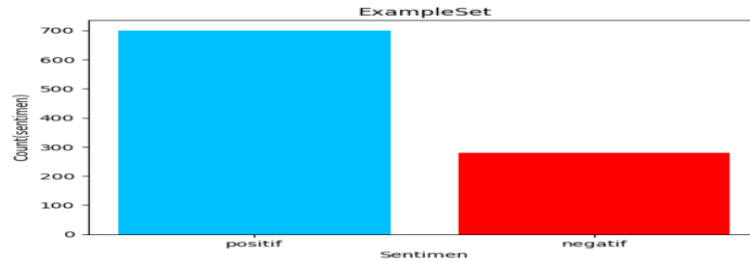
Pelabelan data bertujuan untuk memisahkan data ke dalam dua kelas, yaitu positif dan negatif, guna mengidentifikasi sentimen yang mendukung atau menentang usulan kenaikan PPN sebesar 12%. Setiap data diberi label berdasarkan sentimen yang disampaikan dalam teks, dimana sentimen mendukung diberi label "positif" dan sentimen menentang diberi label "negatif". Proses pelabelan dilakukan secara manual dengan memeriksa data, memahami konteks, dan memberikan label yang sesuai yang dapat dilihat pada Tabel 8.

**Tabel 8.** Pelabelan Data

Text	Polarity
tau naik pajak ditandatangani yg pajak naik pasal ayat uu ppn april januari cek sendiri aneh dpr tuju wakil rakyat tdk suara rakyat	positif
ppn otomatis tiket konser tiket pesawat jancokkkkk	negatif
pilih bbrp terobos yg bagus yg cetus si menteri orang ga tuju ppn	positif

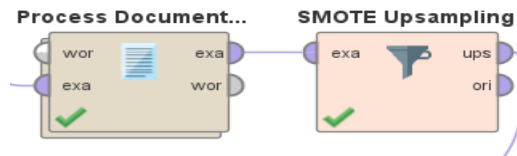
selamat siang pt pkp beli hadiah handphone komisi orang pribadi kena pph pegawai mungut ppn ia

Berikut ini merupakan perbandingan sentimen positif dan negative yang dapat dilihat pada Gambar 6.



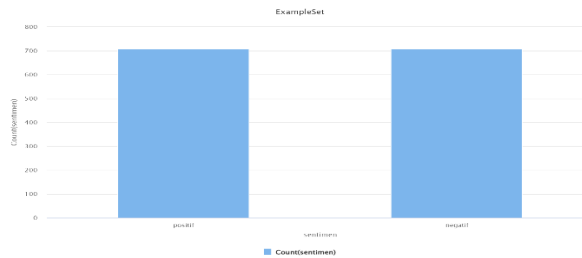
**Gambar 6.** Perbandingan Label Positif dan Negatif

Operator SMOTE Upsampling digunakan untuk mengatasi ketidakseimbangan dalam dataset dengan menyeimbangkan distribusi kelas sentimen tanpa mempengaruhi keakuratan hasil analisis. Gambar 7 berikut menunjukkan penerapan Operator SMOTE Upsampling:



**Gambar 7.** Operator SMOTE Upsampling

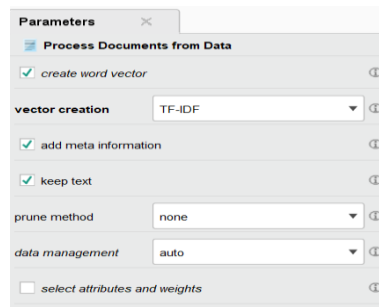
Dapat dilihat pada Gambar 8 yang merupakan hasil dari SMOTE Upsampling.



**Gambar 8.** Hasil SMOTE Upsampling

### 3.4 TF-IDF

Tahapan selanjutnya yaitu pemodelan TF-IDF untuk menghitung bobot setiap kata. Setelah pembersihan data, teks mentah dari tweet diubah menjadi fitur numerik yang dapat digunakan oleh model pembelajaran mesin dengan menggunakan teknik TF-IDF (Term Frequency-Inverse Document Frequency). `TfidfVectorizer` dari library `sklearn` digunakan untuk mengubah teks dalam kolom 'text' menjadi fitur. Setelah itu, fitur yang dihasilkan dimasukkan ke dalam model klasifikasi, yang dibagi menjadi data pelatihan dan pengujian, dengan 80% digunakan untuk pelatihan dan 20% untuk pengujian. Tujuannya adalah untuk meningkatkan akurasi model dalam prediksi model untuk sentimen tweet. Berikut ini parameter yang digunakan TF-IDF yang dapat dilihat pada Gambar 9:



**Gambar 9.** Menghitung TF-IDF pada Rapidminer

### 3.5 Klasifikasi Decision Tree

Langkah awal dalam menerapkan model dilatih dengan menghubungkan operator Decision Tree ke operator data pelatihan menggunakan algoritma Decision Tree. Selanjutnya, operator Apply Model dihubungkan ke data uji





- Terhadap Kenaikan Tarif PPN di Indonesia,” *J. Akunt. DAN Keuang.*, vol. 12, no. 2, pp. 228–238, 2024, [Online]. Available: <https://ojs.unimal.ac.id/index.php/jak/index%0APelabelan>
- [3] M. C. Kwan and B. Sarjono, “Dampak Kenaikan Tarif Pajak Pertambahan Nilai Pada Perilaku Konsumen di Indonesia,” *J. Adijaya Multidisplin*, vol. 2, no. 03, pp. 338–348, 2024, [Online]. Available: <https://e-journal.naurendigiton.com/index.php/mj>
- [4] N. Istiqomah and F. Novika, “Sentiment Analysis Penyedia layanan Asuransi dari Media Sosial Twitter,” *J. Tekno Kompak*, vol. 18, no. 1, p. 77, 2024, doi: 10.33365/jtk.v18i1.3465.
- [5] M. Zania Maharani, “Analisis Sentimen Positif Terhadap Avoskin sebagai Eco Friendly Brand di Media Sosial X dan TikTok,” *Filos. Publ. Ilmu Komunikasi, Desain, Seni Budaya*, vol. 1, no. 3, pp. 125–140, 2024, [Online]. Available: <https://doi.org/10.62383/filosofi.v1i3.169>
- [6] L. A. Pramesti and N. Pratiwi, “Analisis Sentimen Twitter Terhadap Program MBKM Menggunakan Decision Tree dan Support Vector Machine,” *J. Inf. Syst. Res.*, vol. 4, no. 4, pp. 1145–1154, 2023, doi: 10.47065/josh.v4i4.3807.
- [7] R. Harahap, Eva Darwisah Kurniawan, “Analisis Sentimen Komentar Terhadap Kebijakan Pemerintah Mengenai Tabungan Perumahan Rakyat (TAPER) Pada Aplikasi X Menggunakan Metode Naïve Bayes,” *J. Tek. Inform. Unika ST. Thomas*, vol. 9, no. 1, pp. 2657–1501, 2024, [Online]. Available: <https://ejournal.ust.ac.id/index.php/JTIUST/article/view/3911>
- [8] A. Adi et al., “Analisis Sentimen Ulasan pada Maskapai Penerbangan Virgin America Menggunakan Metode Decision Tree,” in *SEMINAR NASIONAL AMIKOM SURAKARTA (SEMNAS) 2024*, 2024, pp. 392–401.
- [9] A. Fatkhudin, F. A. Artanto, N. A. Safli, and ..., “Decision Tree Berbasis SMOTE Dalam Analisis Sentimen Penggunaan Artificial Intelligence Untuk Skripsi,” *REMIK Ris. dan E ...*, vol. 8, no. 2, pp. 494–505, 2024, [Online]. Available: <https://www.jurnal.polgan.ac.id/index.php/remik/article/view/13531%0Ahttps://www.jurnal.polgan.ac.id/index.php/remik/article/download/13531/2453>
- [10] I. Zulfahmi, J. Williem Iskandar, P. V Medan, S. Tuan, and D. Serdang, “Analisis Sentimen Aplikasi PLN Mobile Menggunakan Metode Decision Tree,” *J. Penelit. Rumpun Ilmu Tek.*, vol. 3, no. 1, pp. 11–21, 2024, [Online]. Available: <https://doi.org/10.55606/juprit.v3i1.3096>
- [11] A. R. I. Waldy Romadhona, “Analisis Sentimen Pengguna Media Sosial Terhadap Identitas Kependudukan Digital Menggunakan Metode Support Vector Machine (SVM),” *J. Inf. Syst. ...*, vol. 9, no. 4, pp. 2185–2195, 2024, [Online]. Available: <http://ejournal.seminar-id.com/index.php/josh/article/view/5614%0Ahttps://ejournal.seminar-id.com/index.php/josh/article/download/5614/2906>
- [12] J. S. Gea and H. Budiati, “Analisis Sentimen Masyarakat Terhadap Direktorat Jenderal Pajak,” *J. Sains Dan Komput.*, vol. 8, no. 01, pp. 30–36, 2024, doi: 10.61179/jurnalinfact.v8i01.466.
- [13] F. N. Hasan and M. Dwijayanti, “Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naïve Bayes Classifier,” *J. Linguist. Komputasional*, vol. 4, no. 2, pp. 52–58, 2021, doi: <https://doi.org/10.26418/jlk.v4i2.61>.
- [14] T. Setiawan, S. Liem, and D. M. R. Pribadi, “Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Komentar Tiktok pada Produk Skincare,” *Appl. Inf. Technol. Comput. Sci.*, vol. 3, no. 2, pp. 28–32, 2024, [Online]. Available: <https://jurnal.politap.ac.id/index.php/aicoms>
- [15] Maharani and Fathoni, “Analisis Sentimen Pengguna Terhadap Faktor Penggunaan PayPal Menggunakan Metode Decision Tree,” *J. Ilm. Teknol. Inf. Asia*, vol. 18, no. 1, pp. 71–83, 2024, doi: 10.32815/jitika.v18i1.1002.
- [16] R. Cantika Larasati, C. Dewi, and C. H. Juli, “Analisis sentimen produk kecantikan jenis moisturizer di twitter menggunakan algoritma super vector machine,” *Teknikom*, vol. 7, no. 1, pp. 124–134, 2024, doi: 10.37600/teknikom.v7i1.1243.
- [17] A. Sentimen, “Perbandingan Pelabelan Data dalam Analisis Sentimen Kurikulum Proyek di platform TikTok: Pendekatan Naïve Bayes,” *J. EKSPLORA Inform.*, vol. 14, no. 1, pp. 96–107, 2024, doi: 10.30864/eksplora.v14i1.1093.
- [18] A. M. Hilda et al., “Metode Klasifikasi Support Vector Machine ( SVM ) Untuk Analisis Sentimen Aplikasi Bing : Chat with AI & GPT-4 Di Google Play Store,” *J. Pengkaj. dan Penerapan Tek. Inform.*, vol. 17, no. 1, pp. 68–76, 2024, [Online]. Available: <https://jurnal.itpln.ac.id/petir/article/download/2283/1226/10752>
- [19] H. Z. Muflih, A. R. Abdillah, and F. N. Hasan, “Analisis Sentimen Ulasan Pengguna Aplikasi Ajaib Menggunakan Metode Naïve Bayes,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 3, pp. 1613–1621, 2023, doi: 10.30865/klik.v4i3.1303.
- [20] R. A. Pranata and N. A. Verdikha, “Metode Pembobotan TF-IDF untuk Klasifikasi Teks Quick Count Pemilihan Wakil Presiden Indonesia 2024 pada X Twitter dengan Metode SVM,” *J. Teknol. Inf.*, vol. 18, no. 2, pp. 126–138, 2024, [Online]. Available: <https://e-journal.upr.ac.id/index.php/JTI%0A126>
- [21] A. H. Nasrullah, “Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik,” *J. Pilar Nusa Mandiri*, vol. 7, no. 2, p. 217, 2021, [Online]. Available: <http://ejournal.fikom-unasman.ac.id>
- [22] A. Kartika Sari, Akhmad Irsyad, Dinda Nur Aini, Islamiyah, and Stephanie Elfriede Ginting, “Analisis Sentimen Twitter Menggunakan Machine Learning untuk Identifikasi Konten Negatif,” *Adopsi Teknol. dan Sist. Inf.*, vol. 3, no. 1, pp. 64–73, 2024, doi: 10.30872/atasi.v3i1.1373.
- [23] K. Nurfebria, “Sentiment Analysis of Skincare Products Using the Naive Bayes Method,” vol. 6, no. 3, pp. 1663–1676, 2024, doi: 10.51519/journalisi.v6i3.817.