

# Layanan Perpustakaan UHAMKA

## Intan Diah Hardyatman - Analisis Sentimen Masyarakat Terhadap Rencana Kenaikan PPN 12% Di Indonesia Pada Med...

 07012025

 Fakultas Teknologi Industri dan Informatika

 Universitas Muhammadiyah Prof. Dr. Hamka

---

### Document Details

Submission ID

trn:oid:::1:3125852453

9 Pages

Submission Date

Jan 8, 2025, 2:45 PM GMT+7

3,345 Words

Download Date

Jan 8, 2025, 3:24 PM GMT+7

21,387 Characters

File Name

Intan\_Diah\_Hardyatman\_-\_Analisis\_Sentimen\_Masyarakat\_Terhadap\_Rencana\_Kenaikan\_PPN\_1....docx

File Size

490.0 KB

# 17% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

## Filtered from the Report

- ▶ Bibliography
- 

## Top Sources

14%	 Internet sources
7%	 Publications
5%	 Submitted works (Student Papers)

## Integrity Flags

### 0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

## Top Sources

- 14% Internet sources  
7% Publications  
5% Submitted works (Student Papers)
- 

## Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

Rank	Type	Source	Percentage
1	Internet	ejurnal.seminar-id.com	3%
2	Internet	senafti.budiluhur.ac.id	1%
3	Internet	jim.unindra.ac.id	1%
4	Publication	Rijal, Eka Prasetyaningrum, Agung Purwanto, Abdul Aziz. "Analisis Sentimen Ke...	1%
5	Publication	Risa Sulistiawati, Mia Kamayani. "Analisis Sentimen Aplikasi Maskapai Penerbanga...	1%
6	Internet	polgan.ac.id	1%
7	Internet	www.researchgate.net	0%
8	Internet	www.scribd.com	0%
9	Publication	David J. Putnarubun, C. F. Palembang. "PENERAPAN METODE MACHINE LEARNING...	0%
10	Internet	core.ac.uk	0%
11	Publication	Atik Nurmasani, Yoga Pristyanto. "Algoritme Stacking Untuk Klasifikasi Penyakit J...	0%

12	Student papers	Sriwijaya University	0%
13	Student papers	Universitas Brawijaya	0%
14	Student papers	Universitas Dian Nuswantoro	0%
15	Internet	djournals.com	0%
16	Internet	ejournal.ikmi.ac.id	0%
17	Internet	ojs.amikomsolo.ac.id	0%
18	Publication	Anissya Agsani Pratiwi, Mia Kamayani. "Perbandingan Pelabelan Data dalam Anal...	0%
19	Publication	Muhammad Al Aziiz, Fariz Herlando, Muhammad Jeral Palepa, Reva Qintara Roh...	0%
20	Student papers	Universitas Muhammadiyah Purwokerto	0%
21	Internet	zbook.org	0%
22	Internet	scienceon.kisti.re.kr	0%
23	Internet	id.123dok.com	0%
24	Internet	digilib.uin-suka.ac.id	0%
25	Internet	ejurnal.stmik-budidarma.ac.id	0%

26	Internet	
repository.unair.ac.id		0%
27	Internet	
vdocuments.site		0%
28	Internet	
www.philadelphia.edu.jo		0%
29	Publication	
Tundo, Dea Noer Rachmawati. "Implementasi Algoritma Naive Bayes untuk Analis...		0%
30	Internet	
developer.android.com		0%
31	Internet	
dimaslefi.weebly.com		0%
32	Internet	
johschool.com		0%
33	Internet	
repository.uin-suska.ac.id		0%
34	Internet	
repository.unpkediri.ac.id		0%
35	Publication	
Henny Leidiyana, Titik Misriati, Riska Aryanti. "Klasifikasi Sentimen Terhadap Keb...		0%

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

1

**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: 99/99/999; Accepted: 99/99/999; Published: 99/99/999

**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 [7] dilakukan dengan Eva Darwisah dan Rakhmat Kurniawan. 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 [8] 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%.

Adapun penelitian sebelumnya yang serupa mengenai algoritma Decision Tree [9] yang dilakukan oleh Aslam Fathhudin, Fenelinus Adi Artanto, Naufal Abiyu Safli 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 [1] Jesica Kristovani Siagian dan Painem menggunakan pendekatan Naïve Bayes untuk menganalisis opini masyarakat Indonesia terkait usulan kenaikan PPN sebesar 12% di media sosial X. 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 [10] 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%.

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.

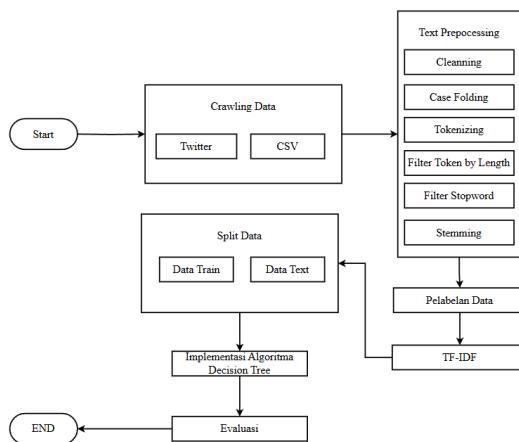
## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

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

- 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].

Berikut ini merupakan alur pada penelitian ini :

**Gambar 1.** Alur Penelitian

## 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 beberapa tahap yang dilakukan:

- Cleaning* : menghilangkan noise dalam bentuk simbol, termasuk karakter @#\$%^&\*\*(), mention”@”, dan hashtag “#” [16].
- Case Folding* : mengganti kalimat pada dokumen menjadi huruf kecil.
- Tokenizing* : menggunakan hasil pemrosesan case folding untuk membagi kalimat menjadi kata-kata berdasarkan spasi [17].
- Filter Token by Length* : menghilangkan kalimat dengan karakter lebih sedikit dari empat karakter [18].
- 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].
- 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].

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

## 2.2 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 merupakan tabel Confusion Matrix :

**Tabel 1.** Confusion Matrix

		Prediksi	
		TRUE	FALSE
Aktual	TRUE	TP	FP
	FALSE	FN	TN

14

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

Keterangan :

TP (True Positive) : Data ini merupakan data positif yang diprediksi benar

FP (False Positive) : Data ini merupakan data negatif namun diprediksi sebagai data positif

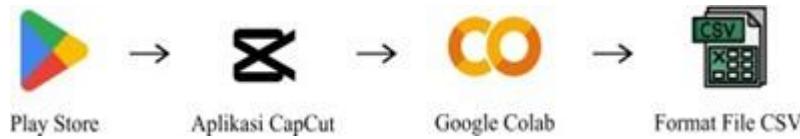
FN (False Negative) : Data ini merupakan data positif namun diprediksi sebagai data negatif

TN (True Negative) : Data ini merupakan data negatif yang diprediksi benar

### 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*). Berikut adalah 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. Berikut ini adalah 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 bantuan pemerintah dapat kenaikan PPN 12%. Hemmm....teteep cuan?!	1,869E+18	2024-12-17 09:53:05+C	Ascan155885	1,72E+18	in	0	0	0	0	https://x.com/Ascan155885/status/18689575955262			
978 Ilustrasi sederhananya si Varel kalo suka mandi pake Aqua mau naik 12% aja dia tpp mandi pake aqua	1,869E+18	2024-12-17 09:53:08+4C	aditaleee	1,15E+18	in	0	0	0	0	https://x.com/aditaleee/status/18689575955262			
979 2025 meningkatkan kesejahteraan 2025 meningkatkan ppon 12%	1,869E+18	2024-12-17 09:53:10+4C	madefromlove	1,33E+18	in	0	0	1	1	https://x.com/madefromlove/status/18689575955262			
980 Hal paling lucu dari PPN naik 12% adalah neglat voter 02 yg KAGET (especially rakyat biasa) akan keb	1,869E+18	2024-12-17 09:53:30+4C	rudauhl	1,27E+18	in	0	0	1	1	https://x.com/rudauhl/status/186895768766249			
981 12% Gv masih ga nyangka aja ppn segede itu di tengah gaji semimin ini.	1,869E+18	2024-12-17 09:53:34+4C	chonlo_lolo	1,49E+18	in	0	0	1	1	https://x.com/chonlo_lolo/status/186895770512			
982 @tanyaries Jasa pendidikan premium itu kyk mnjin? Keadaan normal aja inflasi biaya pendidikan naik	1,869E+18	2024-12-17 09:53:36+4C	Hollyne_Retha	1,48E+18	in	0	0	0	0	https://x.com/Hollyne_Retha/status/186895771716			
983 ya Allah bingung bgt mau nangisin diri sendiri krn baru dijatihin org yg paling gue sayang apa nangisng	1,869E+18	2024-12-17 09:53:34+4C	levisorius	1,25E+18	in	0	0	0	0	https://x.com/levisorius/status/186895773620			
984 @tempodotco Biukannya harus berbanding lurus juga dengan pendapat rata-rata warganya ya? Rasa	1,869E+18	2024-12-17 09:53:34+4C	naufal_2828	1,47E+18	in	0	0	0	0	https://x.com/naufal_2828/status/18689577510			
985 sedarn memikiran narasi agar konten tentang kenaikan PPN 12% ini masuk ke otak warga tiktok	1,869E+18	2024-12-17 09:53:40+4C	RUMROJEN	2,89E+09	in	0	0	0	0	https://x.com/RUMROJEN/status/18689578468			
986 gmn ini ppon 12% sinting lo ya	1,869E+18	2024-12-17 09:53:41+4C	idkwhthodo	1,26E+18	in	0	0	0	0	https://x.com/idkwhthodo/status/18689578804			
987 ehhi ini klo ppon 12% ntll bakso mas harno naik dong harganya ?? dam pdhl itu bakso terenak	1,869E+18	2024-12-17 09:53:41+4C	loufyijunan	1,56E+18	in	0	0	0	0	https://x.com/loufyijunan/status/18689578908			
988 Nih ya dengan naiknya ppon 12% harga barang pasti naik jadi pasti akan menaiknya harga dagangan	1,869E+18	2024-12-17 09:53:41+4C	mingyoori	1,26E+18	in	0	1	0	1	https://x.com/mingyoori/status/186895789953			
989 -naiknya harga layanan bikin fanartis ketusukan kalian kalaun jalur keychain memungkinkan untuk mre	1,869E+18	2024-12-17 09:54:20+4C	mingyoori	1,26E+18	in	0	0	0	1	https://x.com/mingyoori/status/1868957893569			
990 yaelah ngapain sih pak ppon 12% segala monyetttttt	1,869E+18	2024-12-17 09:54:30+4C	kloroform	1,53E+18	in	0	0	0	0	https://x.com/kloroform/status/186895793669			
991 ORGIL ppon 12% dikira ga gede?????	1,869E+18	2024-12-17 09:54:38+4C	ekeuuuuuu	1,84E+18	in	0	0	0	0	https://x.com/ekeuuuuuu/status/186895792615			
992 ppon 12% saatinya stop ngeop gak sih (ngetik sampil nangis perlu asupan kopli)	1,869E+18	2024-12-17 09:54:39+4C	lopnoons	1,31E+18	in	0	0	0	0	https://x.com/lopnoons/status/186895795832			
993 Natal dan Tahun Baru makin seru. Telkomsel hadirkan jaringan stabil untuk temani setiap momen sp	1,869E+18	2024-12-17 09:54:50+4C	KabarNetral	1,79E+18	in	0	0	0	0	https://x.com/KabarNetral/status/186895802176			
994 @_13NDDH Harga sepatu sebelum pajak 2 7ή lalu include PPN 11% jadi 2 997 000 dan klo PPN 12% ja	1,869E+18	2024-12-17 09:54:54+4C	akupiyooou	1,24E+18	in	0	0	0	0	https://x.com/akupiyooou/status/186895804454			
995 bismillah ppon 12% batam	1,869E+18	2024-12-17 09:54:58+4C	akupiyooou	1,68E+18	in	0	1	0	0	https://x.com/akupiyooou/status/18689580538C			
996 2. PPN 12% Dilakukan Secara Bertahap kebijakan ini tidak dilakukan secara mendadak melainkan ber	1,869E+18	2024-12-17 09:55:04+4C	sulastri65481	4,44E+09	in	0	1	0	0	https://x.com/sulastri65481/status/18689580797			
997 Terjawab sudah PPN 12 persen kapau mulai berlaku cek juga golongan warga yang dapat diskon tarif	1,869E+18	2024-12-17 09:55:13+4C	tribunkalimantan	1,04E+08	in	0	0	0	0	https://x.com/tribunkalimantan/status/186895811776			
998 @_Heraloebs Pulangin. Masuk orderan pas PPN naik 12%	1,869E+18	2024-12-17 09:55:40+4C	emfahmifitl	3,47E+08	in	0	0	0	0	https://x.com/emfahmifitl/status/186895823268			
999 Lebih ngabisin energi liat berita PPN naik 12% dan keborongan rezim sih gelud war cecinaan gak ngar	1,869E+18	2024-12-17 09:55:46+4C	myyngazi	1,29E+18	in	0	0	0	0	https://x.com/myyngazi/status/1868958257180			
1000 @dear_woobin iya kan.. harga pada naik tiap daya beli menuron eh malah mau naikin ppon 12% tuh loq	1,869E+18	2024-12-17 09:55:54+4C	arisashara	9,28E+17	in	0	0	0	0	https://x.com/arisashara/status/186895828898			
1001 Vietnam: 8% tetapi sembako dikenakan pajak 5% Filipina: 12% India: 18% (untuk sebagian barang da	1,869E+18	2024-12-17 09:56:24+4C	sulastri65481	4,44E+09	in	0	0	0	0	https://x.com/sulastri65481/status/1868958451			

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.

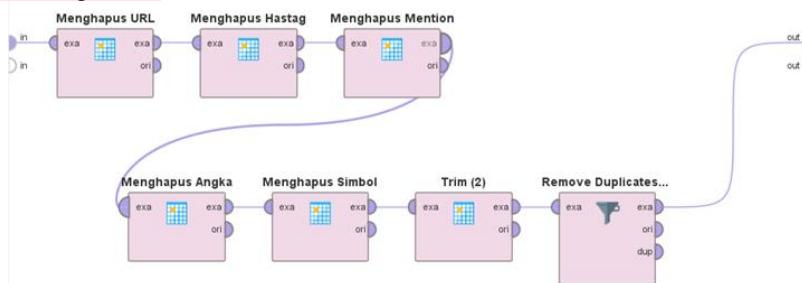
#### 3.2 Preprocessing Data

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*, *Hashtag*, *Mention*, Angka, Simbol, *Trim* (menghapus spasi yang tidak

dibutuhkan di awal dan akhir teks), serta *Duplicates* (menghapus salinan teks dalam dataset). Berikut ini gambar 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 sndiri. 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 sndiri 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.

**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 sndiri 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 sndiri 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.

**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 sndiri 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', 'sndiri', 'anehnya', 'kenapa', 'dpr', 'menyetujui', 'wakil', 'rakyat', 'kok', 'tdk', 'menyuarkan', '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.

**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’, ‘sndiri’, ‘anehnya’, ‘kenapa’, ‘dpr’, ‘menyetujui’, ‘wakil’, ‘rakyat’, ‘kok’, ‘tdk’, ‘menyuarkan’, ‘rakyat’]	[‘tau’, ‘kenaikan’, ‘pajak’, ‘ditandatangani’, ‘yg’, ‘pajak’, ‘dinaikkan’, ‘pasal’, ‘ayat’, ‘uu’, ‘ppn’, ‘april’, ‘januari’, ‘cek’, ‘sndiri’, ‘anehnya’, ‘dpr’, ‘menyetujui’, ‘wakil’, ‘rakyat’, ‘tdk’, ‘menyuarkan’, ‘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*.

**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’, ‘sndiri’, ‘anehnya’, ‘kenapa’, ‘dpr’, ‘menyetujui’, ‘wakil’, ‘rakyat’, ‘kok’, ‘tdk’, ‘menyuarkan’, ‘rakyat’]	[‘tau’, ‘kenaikan’, ‘pajak’, ‘ditandatangani’, ‘yg’, ‘pajak’, ‘dinaikkan’, ‘pasal’, ‘ayat’, ‘uu’, ‘ppn’, ‘april’, ‘januari’, ‘cek’, ‘sndiri’, ‘anehnya’, ‘dpr’, ‘menyetujui’, ‘wakil’, ‘rakyat’, ‘tdk’, ‘menyuarkan’, ‘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

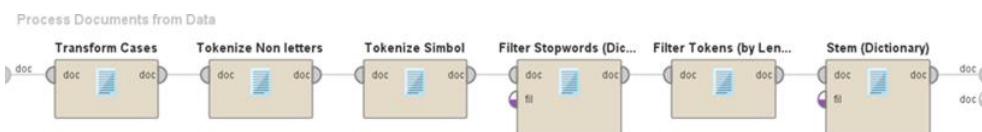
Pada tahap stemming dengan menggunakan operator stem, stemming mengubah kata dengan imbuhan menjadi kata dasar.

**Tabel 7.** Hasil dari Proses Stemming

Sebelum	Sesudah
[‘tau’, ‘kenaikan’, ‘pajak’, ‘ditandatangani’, ‘yg’, ‘pajak’, ‘dinaikkan’, ‘pasal’, ‘ayat’, ‘uu’, ‘ppn’, ‘april’, ‘januari’, ‘cek’, ‘sndiri’, ‘anehnya’, ‘dpr’, ‘menyetujui’, ‘wakil’, ‘rakyat’, ‘tdk’, ‘menyuarkan’, ‘rakyat’]	[‘tau’, ‘naik pajak’, ‘ditandatangani’, ‘yg’, ‘pajak’, ‘naik’, ‘pasal’, ‘ayat’, ‘uu’, ‘ppn’, ‘april’, ‘januari’, ‘cek’, ‘sndiri’, ‘aneh’, ‘dpr’, ‘tuju’, ‘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 "menyuarkan" 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

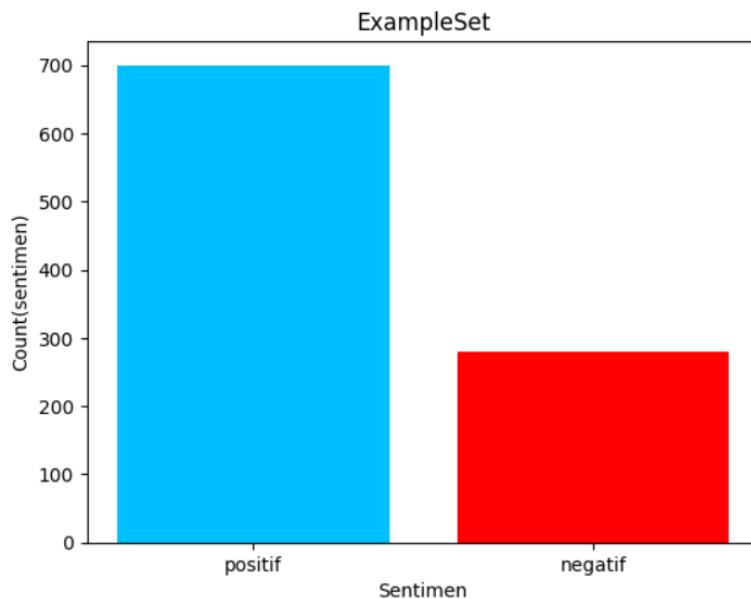
### 3.3 Pelabelan Data

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.

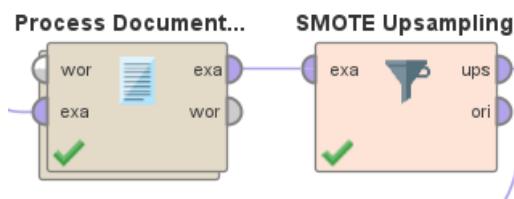
**Tabel 8.** Pelabelan Data

Text	Polarity
tau naik pajak ditandatangani yg pajak naik pasal ayat uu ppn april januari cek sndiri 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
<u>selamat siang pt pkp beli hadiah handphone komisi orang pribadi kena pph pegawai mungut ppn ia</u>	positif

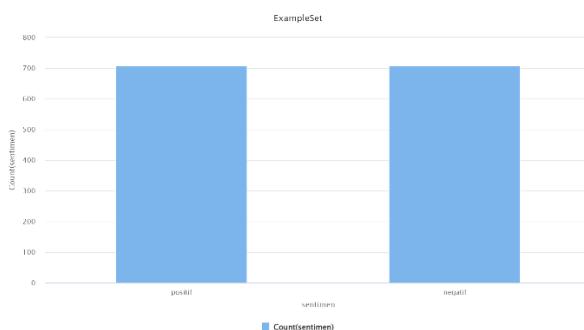


**Gambar 6.** Perbandingan Label Positif dan Negatif

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



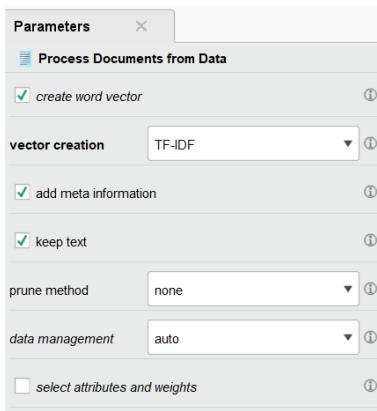
**Gambar 7.** Operator SMOTE Upsampling



**Gambar 8.** Hasil SMOTE Upsampling

### 3.4 TF-IDF

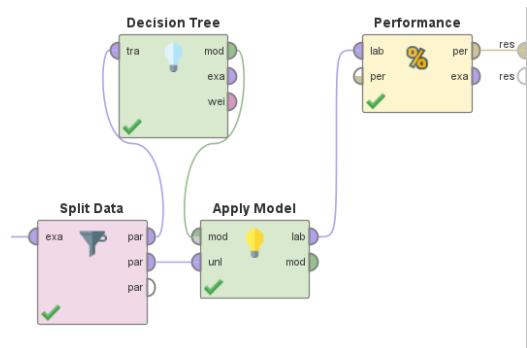
Tahapan selanjutnya yaitu pemodelan TF-IDF untuk menghitung bobot setiap kata. Setelah pembersihan data, teks mentah dari tweet diubah menjadi fitur numerik yang dapat digunakan oleh model pembelajaran mesin dengan menggunakan teknik TF-IDF (*Term Frequency-Inverse Document Frequency*). `TfidfVectorizer` 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 :



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 untuk menguji kinerja model. Prosedur ini menguji model yang telah dilatih menggunakan data latih, seperti yang ditunjukkan pada gambar 10.



Gambar 10. Implementasi Algoritma Decision Tree

Setelah langkah preprocessing, pelabelan, dan klasifikasi, Decision Tree digunakan untuk menghasilkan prediksi. Untuk kata-kata yang sering muncul dalam 1000 tweet yang dianalisis, hasil kategorisasi Decision Tree divisualisasikan dalam bentuk wordcloud. Wordcloud ini menampilkan 30 kata paling sering muncul, diurutkan secara descending, seperti yang ditunjukkan pada gambar 11.



Gambar 11. Visualisasi WordCloud

### 3.6 Evaluasi

Hasil evaluasi menunjukkan bahwa algoritma Decision Tree mencapai tingkat akurasi sebesar 81,34%, *precision* sentimen positif 75,72% sementara untuk sentimen negatif 90,09%, *recall* sentimen positif 92,25% dan untuk sentimen negatif 70,42%. Ilustrasi hasil Confusion Matrix ditampilkan di bawah ini.

accuracy: 81.34%			
	true positif	true negatif	class precision
pred. positif	131	42	75.72%
pred. negatif	11	100	90.09%
class recall	92.25%	70.42%	

Gambar 12. Hasil Confusion Matrix dari Algoritma Decision Tree

## 4. KESIMPULAN

Hasil analisis sentimen terhadap rencana kenaikan PPN 12% di Indonesia memberikan hasil berjumlah 1000 dari pemrosesan *dataset*. Selanjutnya setelah dilakukan proses *preprocessing* dataset tersebut terdapat sentimen postif berjumlah 715, lalu sentimen negatif berjumlah 285. Hasil ini menunjukkan bahwa rencana kenaikan PPN12% di Indonesia mendapatkan respon yang positif menunjukkan bahwa mayoritas di *platform* media sosial X cenderung mendukung, sedangkan sentimen negatif cenderung lebih ke kritik atau kekhawatiran terhadap rencana kenaikan PPN12% di Indonesia. Pengujian algoritma Decision Tree mendapatkan akurasi sebesar 81,34%, *precision* negatif mencapai 90,09%, sedangkan sentimen positif sebesar 75,72%, *recall* untuk sentimen negatif 70,42% dan untuk sentimen positif 92,25%.