Published online on the journal's webpage: **http://journal.ittelkom-pwt.ac.id/index.php/dinda**

Journal of Dinda

Data Science, Information Technology, and Data Analytics

Vol. 5 No. 2 (2025) 241 - 249 E-ISSN: 2809-8064

Analysis of Public Sentiment Toward the Increase in VAT Rates Using the SVM Algorithm

Elsa Azila Rahman[1*], Aidil Halim Lubis[2]

[1*,2] Department of Computer Science, Faculty of Science and Technology, State Islamic University of North Sumatra

[1*]elsaaziahman@gmail.com, [2]aidilhalimlubis@uinsu.ac.id

Abstract

The Policy Of Increasing the Value Added Tax (VAT), particularly on luxury goods as stipulated in Minister of Finance Regulation (PMK) Number 131 of 2024, has sparked various public responses. In the digital era, social media serves as a primary platform for the public to express opinions about government policies. This study aims to analyze public sentiment toward the VAT policy to provide data-driven insights that can support more responsive policymaking. A total of 4,000 comments were collected from the X platform using web crawling techniques. After preprocessing, 3,553 clean comments were obtained. Sentiment labeling was carried out automatically using a lexicon-based approach. Sentiment classification was then performed using the Support Vector Machine (SVM) algorithm with a polynomial kernel and an 80:20 train-test data split. Lexicon-based labeling revealed that 73.3% of comments expressed positive sentiment, while 26.7% were negative. The SVM model achieved an accuracy of 76.65%. For positive sentiment, the model obtained precision of 76.18%, recall of 100%, and F1-score of 86.51%. For negative sentiment, precision was 100%, recall 7.78%, and F1-score 14.44%. The SVM model performed well in detecting positive sentiment but was less effective for negative sentiment, indicating the need for optimization. These findings can help policymakers better understand public perceptions, anticipate resistance, and improve policy communication strategies. The proposed method can also be applied to other public policy analyses using social media.

Keywords: *Preprocessing, Sentiment Analysis,* SVM, *Value Added Tax* (VAT), X

1. Introduction

Taxes are a right of the state. Therefore, the collection of taxes by the state is mandatory for citizens [1]. Taxes have a basic system of collecting contributions in the form of money (funds) obtained from the community, and the final output will be used for the community itself [2]. This can be seen from several examples, such as road construction, public facilities, and examples that can be directly observed, such as subsidies for the public ranging from education, daily necessities (water, electricity), to health subsidies, as well as other examples that are fundamentally used for the welfare of the people [3].

According to Prof. Dr. H. Rochmat Soemitro, SH, "Taxes are contributions made by the people to the state treasury based on the law without receiving any direct reciprocal services (consideration) that can be directed and used to pay for public expenditures." This definition was later revised to read: "Taxes are a transfer of wealth from the people to the state treasury to finance routine expenditures, with the surplus used for public savings, which is the main source of funding for public investment"[4].

In the digital age, social media has become the main platform for people to openly express their opinions, one of which is the X app. This platform generates a large amount of data that reflects public sentiment on various issues, including the VAT rate increase policy. Analysis of this data can provide important insights for policymakers to understand public response and design more effective measures to manage the impact of the policy [5].

The tax increase policy issued by the Minister of Finance has become a controversial issue, as it greatly affects the economy of the community. Under Minister of Finance

Received: 07-08-2025 | Accepted: 18-08-2024 | Published: 19-08-2024

Regulation (PMK) No. 131 of 2024, which governs the implementation of Value-Added Tax (VAT) and serves as the basis for applying a 12% VAT rate, the regulation will take effect on January 1, 2025, as stated in Article 6 of PMK No. 131 of 2024.

In a previous study by Jesica Kristovani Siagian and Painem entitled "Analysis of Indonesian Public Sentiment Towards the Plan to Increase VAT to 12% on Social Media X Using the Naïve Bayes Method," [6] the results showed that the Naïve Bayes method was able to provide an overview of public sentiment with a certain degree of accuracy. However, this study uses the Support Vector Machine (SVM) algorithm, which is expected to enhance the accuracy of sentiment analysis more effectively, particularly in handling more complex and non-linear data patterns.

To gain a deeper understanding of how the public views this policy, a sentiment analysis is needed to reveal the positive, negative, or neutral sentiments of various affected parties. Using the support vector machine (SVM) algorithm approach, this study can group and analyze public sentiment data through the social media application X [7].

This study aims to analyze public sentiment toward the VAT increase using the SVM algorithm. Data collected from social media will be processed to identify sentiment patterns and understand public perception of this policy. The results of this study are expected to contribute to the government and policymakers in understanding public opinion and assist in formulating policies that are more responsive to the needs of the community.

2. Research Methods

This study uses a quantitative method, which is a research method that tests theories by examining the relationship between variables. This study goes through stages that will be applied when conducting research so that the research stages are more structured.

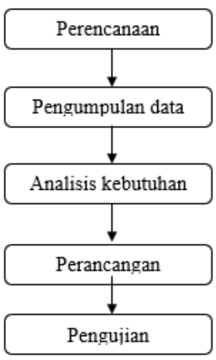


Figure 1. Research Framework

2.1. System Design

The system design in this study aims to define the workflow in analyzing public sentiment towards the increase in Value Added Tax (VAT) using the Support Vector Machine (SVM) algorithm. This system is designed to collect, process, and analyze data systematically to produce accurate sentiment classification.

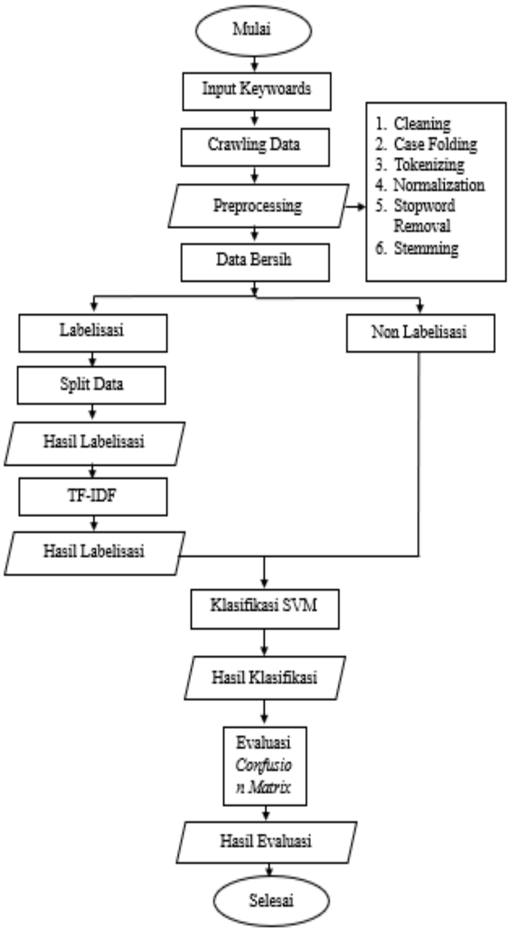


Figure 2. Research Flow

2.2. Data Collection

This study applies three main methods in data collection, namely literature study, bibliographic study, and data collection from application X. Literature and bibliographic studies are used to understand theories related to sentiment analysis and SVM algorithms, while data from application X is collected through web scraping or utilization of application features. After the data was collected, a pre-processing stage was conducted, including text cleaning, normalization, and feature extraction, before being analyzed using the SVM algorithm.

The data in this study was obtained through social media X using the official API from Twitter (Twitter API v2). Data crawling successfully collected 4,000 tweets. There were three keywords used as search references, namely "PPNBM", "Pajak Barang Mewah" (Luxury Goods Tax), and "PPN Barang Mewah" (Luxury Goods VAT).

2.3. Preprocessing

Data preprocessing is the initial data processing used to convert raw data obtained from various sources into cleaner information that can be used for further analysis [8].

a. Cleaning

During the cleaning stage, punctuation marks and unnecessary characters such as periods, commas, question marks, exclamation marks, HTML and URLs, hashtags and mentions, emojis, and irrelevant characters will be removed.

Table 1. Cleaning

Before Cleaning	After Cleaning
https://t.co/EFZztn8zwz Alhamdulillah Presiden Prabowo mengumumkan kenaikan PPN 12% hanya dikenakan untuk barang dan jasa mewah seperti pesawat jet pribadi kapal pesiar rumah mewah kendaraan tertentu yang diatur sebagai objek PPnBM bukan barang kebutuhan pokok.Jenis	Alhamdulillah Presiden Prabowo mengumumkan kenaikan PPN hanya dikenakan untuk barang dan jasa mewah seperti pesawat jet pribadi kapal pesiar rumah mewah kendaraan tertentu yang diatur sebagai objek PPnBM bukan barang kebutuhan pokok Jenis

b. Case Folding

In the case folding stage, the collection of tweets will be converted to lowercase letters

Table 2. Case Folding

Before Tokenization	After Tokenization
Alhamdulillah Presiden Prabowo mengumumkan kenaikan PPN hanya dikenakan untuk barang dan jasa mewah seperti pesawat jet pribadi kapal pesiar rumah mewah kendaraan tertentu yang diatur sebagai objek PPnBM bukan barang kebutuhan pokok Jenis	alhamdulillah presiden prabowo mengumumkan kenaikan ppn hanya dikenakan untuk barang dan jasa mewah seperti pesawat jet pribadi kapal pesiar rumah mewah kendaraan tertentu yang diatur sebagai objek ppnbm bukan barang kebutuhan pokok jenis

c. Tokenization

At this tokenization stage, a sentence from tweets will be separated into word segments before further analysis.

Table 3. Tokenization

Before Tokenization	After Tokenization
alhamdulillah presiden prabowo mengumumkan kenaikan ppn hanya dikenakan untuk barang dan jasa mewah seperti pesawat jet pribadi kapal pesiar rumah mewah kendaraan tertentu yang diatur sebagai objek ppnbm bukan barang kebutuhan pokok jenis	['alhamdulillah', 'presiden', 'prabowo', 'mengumumkan', 'kenaikan', 'ppn', 'hanya', 'dikenakan', 'untuk', 'barang', 'dan', 'jasa', 'mewah', 'seperti', 'pesawat', 'jet', 'pribadi', 'kapal', 'pesiar', 'rumah', 'mewah', 'kendaraan', 'tertentu', 'yang', 'diatur', 'sebagai', 'objek', 'ppnbm', 'bukan', 'barang', 'kebutuhan', 'pokok', 'jenis']

d. Normalization

Normalization is the process of converting non-standard words or word variations into a standard or consistent form. The purpose is to: Standardize words with the same meaning but different forms, reduce unnecessary word variation, thereby improving the accuracy of analysis results, and facilitate stopword removal, stemming, and classification processes. This process uses a dataset file containing a standard language dictionary accessible from https://www.kaggle.com/datasets/fornigulo/kamus-slag.

Table 4. Normalization

Before Normalization	After Normalization
['alhamdulillah', 'presiden', 'prabowo',	['alhamdulillah', 'presiden', 'prabowo',

'mengumumkan', 'kenaikan', 'ppn', 'hanya', 'dikenakan', 'untuk', 'barang', 'dan', 'jasa', 'mewah', 'seperti', 'pesawat', 'jet', 'pribadi', 'kapal', 'pesiar', 'rumah', 'mewah', 'kendaraan', 'tertentu', 'yang', 'diatur', 'sebagai', 'objek', 'ppnbm', 'bukan', 'barang', 'kebutuhan', 'pokok', 'jenis']	'mengumumkan', 'kenaikan', 'ppn', 'hanya', 'dikenakan', 'untuk', 'barang', 'dan', 'jasa', 'mewah', 'seperti', 'pesawat', 'jet', 'pribadi', 'kapal', 'pesiar', 'rumah', 'mewah', 'kendaraan', 'tertentu', 'yang', 'diatur', 'sebagai', 'objek', 'ppnbm', 'bukan', 'barang', 'kebutuhan', 'pokok', 'jenis']

e. Stopword Removal

Stopword removal is the process of removing common words that have no significant meaning in analysis, such as: “which”, “and”, “in”, “to”, “is”, ‘that’, “I”, etc. The purpose: To reduce noise (unimportant words) in the text, make the model more focused on meaningful words (e.g., “increase,” “tax,” “PPNBM”), reduce the number of features, and speed up the analysis process.

Table 5. Stopword Removal

Before Stopword Removal	After Stopword Removal
['alhamdulillah', 'presiden', 'prabowo', 'mengumumkan', 'kenaikan', 'ppn', 'hanya', 'dikenakan', 'untuk', 'barang', 'dan', 'jasa', 'mewah', 'seperti', 'pesawat', 'jet', 'pribadi', 'kapal', 'pesiar', 'rumah', 'mewah', 'kendaraan', 'tertentu', 'yang', 'diatur', 'sebagai', 'objek', 'ppnbm', 'bukan', 'barang', 'kebutuhan', 'pokok', 'jenis']	['alhamdulillah', 'presiden', 'prabowo', 'mengumumkan', 'kenaikan', 'ppn', 'dikenakan', 'barang', 'jasa', 'mewah', 'pesawat', 'jet', 'pribadi', 'kapal', 'pesiar', 'rumah', 'mewah', 'kendaraan', 'diatur', 'objek', 'ppnbm', 'barang', 'kebutuhan', 'pokok', 'jenis']

f. Stemming

Stemming is a process in text preprocessing that aims to convert words to their basic form or root word. This program uses a literary library to perform stemming on Indonesian text. In other words, stemming removes affixes such as prefixes, suffixes, infixes, or combinations of affixed words [9].

Table 6. Stemming

Before Stemming	After Stemming
['alhamdulillah', 'presiden', 'prabowo', 'mengumumkan', 'kenaikan', 'ppn', 'dikenakan', 'barang', 'jasa', 'mewah', 'pesawat', 'jet', 'pribadi', 'kapal', 'pesiar', 'rumah', 'mewah', 'kendaraan', 'diatur', 'objek', 'ppnbm', 'barang', 'kebutuhan', 'pokok', 'jenis']	alhamdulillah presiden prabowo umum naik ppn kena barang jasa mewah pesawat jet pribadi kapal pesiar rumah mewah kendara atur objek ppnbm barang butuh pokok jenis

2.4. Lexicon-Based

Labeling in this study uses a lexicon-based approach, which determines sentiment (positive, negative, or neutral) based on a list of words (lexicon) that have been previously assigned polarity values. The dictionary used is the Indonesia Sentiment Lexicon (InSet), which can be obtained at https://github.com/fajri91/InSet and consists of 3,069 positive words and 6,609 negative words.

The lexicon-based labeling method uses a predetermined list of words to identify sentiment. However, this approach has limitations, such as the inability to capture slang, abbreviations, or the use of sarcasm that is common on social media. For example, a sentence that lexically contains positive words can have a negative tone if used ironically. Additionally, the vocabulary in the lexicon may not fully represent the language and communication style of social media users in Indonesia, potentially leading to bias in the labels assigned.

Table 7. Sample Labeling Results

Hasil Preprocessing	Bobot Kata	Total Bobot	Label
telat kocak mending pagi besok langsung suruh staff asih tau distributor agen ppn barang mewah	{'telat': -5, 'mending': -3, 'pagi': 1, 'suruh': -2, 'asih': -1}	-10	Negatif

alhamdulillah presiden prabowo umum naik ppn kena barang jasa mewah pesawat jet pribadi kapal pesiar rumah mewah kendara atur objek ppnbm barang butuh pokok jenis	{'alhamdulillah': 5, 'rumah': 3}	8	Positif

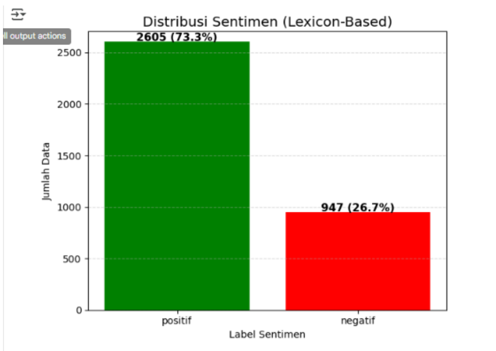


Figure 3. Percentage and Frequency of Sentiment Labels

2.5. TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) is a method for measuring how important a word is in a document compared to the entire document collection. In the feature extraction stage, the TF-IDF method is used to convert preprocessed text into numerical form.

Table 8. Calculation of TF and DF Values

Term	TF					DF
	D1	D2	D3	D4	D5	
akurat	0	0	0	0	1	1
barang	0	1	1	1	1	4
alah	1	0	0	0	0	1
bela	0	0	0	0	1	1
…	…	…	…	…	…	…
mewah	0	1	1	1	1	4

$$IDF = In\left(\frac{D+1}{df+1}\right) + 1$$

IDF : Inversed Document Frequency
D : Data frequency on D
df : Many documents containing the search term

Table 9. IDF value of training data

Term	DF	IDF
akurat	1	2.0986
barang	4	1.182
alah	1	2.0986
bela	1	2.0986
…	…	…
mewah	4	1.182

$$w = TF \;\; X \; IDF$$

w : Weight of document d relative to word t
TF : Number of words in the searched document
IDF : *Iinversed Document Frequency*

Table 10. TF-IDF Values from Training Data

Term	TF-IDF				
	D1	D2	D3	D4	D5
alah	2.098612	0	0	0	0
pusing	2.098612	0	0	0	0
ppn	1.182322	1.182322	0	1.182322	1.182322
kepala	2.098612	0	0	0	0
…	0	2.098612	0	0	0
pajak	0	0	2.098612	0	0

$$TF_{\text{norm}}\,(t,d) = \frac{TF\,(t,d)}{\sqrt{\sum_i (t,d))^2}}$$

d : Document d
t : The first word of the keyword
TF : Number of words in the document being searched

Table 11. Data Normalization

Term	TF-IDF

	D1	D2	D3	D4	D5
alah	0.549036	0	0	0	0
pusing	0.549036	0	0	0	0
ppn	0.309317	0.204318	0	0.219245	0.167938
kepala	0.549036	0	0	0	0
…	0	0	0	0	0
pajak	0	0	0.28944	0	0

2.6. Splitting Data

Data splitting is the process of dividing a dataset into several parts for the purposes of training and testing machine learning models. At this stage, researchers divide the data into 80% training data and 20% testing data. Out of the total 3,553 comments obtained, 2,842 data points (80%) were used as training data and 711 data points (20%) as testing data.

3. Results and Discussion

In the classification stage, the algorithm used is a support vector machine (SVM) with a polynomial kernel. The polynomial kernel was chosen because it is capable of capturing non-linear patterns in the data, particularly in text data that has been transformed into numerical form using the TF-IDF method. The polynomial kernel projects the data into a higher-dimensional space, enabling the separation of positive and negative classes even though they cannot be linearly separated in the original space. By using this kernel, the SVM can form a more flexible and complex hyperplane, enhancing its classification capability for diverse sentiment data. The classification results are then evaluated to measure the model's performance in recognizing sentiment from text data.

a. Menghitung Kernel

Pada klasifikasi yang dilakukan, jenis kernel yang digunakan adalah kernel polynomial. Kernel polynomial menghitung kemiripan antara dua vektor fitur dengan rumus:

$K = (x, y) = (\gamma . x . y + r)^2$

Berikut contoh representasi data pada 5 buah data:

Table 12. *Sample Representasi*

	x1	x2	x3	x4	x5
y1	K(x1, y1)	K(x2, y1)	K(x3, y1)	K(x4, y1)	K(x5, y1)
y2	K(x1, y2)	K(x2, y2)	K(x3, y2)	K(x4, y2)	K(x5, y2)
y3	K(x1, y3)	K(x2, y3)	K(x3, y3)	K(x4, y3)	K(x5, y3)
Y4	K(x1, y4)	K(x2, y4)	K(x3, y4)	K(x4, y4)	K(x5, y4)
y2	K(x1, y5)	K(x2, y5)	K(x3, y5)	K(x4, y5)	K(x5, y5)

Berikut adalah sample hasil perhitungan normalisasi data yang dilakukan

Table 13. Perhitungan Kernel Polynomial

	d1	d2	d3	d4	d5
d1	1	0.063199	0	0.067816	0.051946
d2	0.063199	1	0.134961	0.134387	0.102939
d3	0	0.134961	1	0.071503	0.11093
d4	0.067816	0.134387	0.071503	1	0.185967
d5	0.051946	0.102939	0.11093	0.185967	1

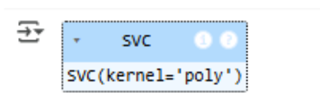


Figure 4. Kernel Polynomial

3.1. Model Evaluation

After the SVM model training process with polynomial kernel is complete, the next step is to evaluate the model to measure how well it performs sentiment classification. The evaluation is carried out using test data that has been separated from the training data.

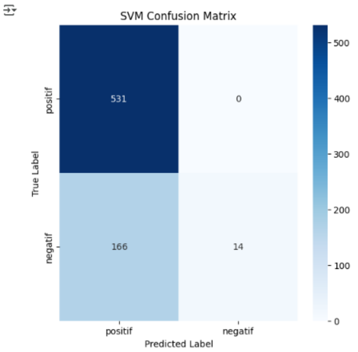


Figure 5. Results of the SVM Model Confusion Matrix

Based on Figure 5, the True Positive (TP) value is 531, False Positive (FP) is 166, False Negative (FN) is 0, and True Negative (TN) is 248.

Table 13. Matrix confusion results

Predicted Value	Actual Value	
	Positive	Negative
Positive	531	0
Negative	166	14

Dari diatas diketahui hasil *accuracy, precision*, *recall*, dan *F1-Score* kernel polynomial dapat dihitung sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{531+14}{531+14+166+0} = \frac{545}{711} = 0.7665 \approx 76.65\%$$

$$Precission\ positive\ class = \frac{TP}{TP+FP} = \frac{531}{531+166} = \frac{531}{697} = 0.7618 \approx 76.18\%$$

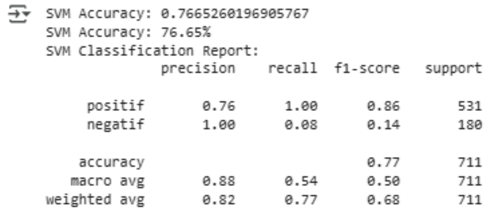
$$Precission\ \text{negative class} = \frac{TN}{TN+FN} = \frac{14}{14+0} = \frac{14}{14} = 1.0 \approx 100\%$$

$$Recall\ \text{positive class} = \frac{TP}{TP+FN} = \frac{531}{531+0} = \frac{531}{531} = 1.0 \approx 100\%$$

$$Recall\ \text{negative class} = \frac{TN}{TN+FP} = \frac{14}{14+166} = \frac{14}{180} = 0.0778 \approx 7.78\%$$

$$F1\text{-}Score\ positive\ class = 2x\,\frac{\text{presisi x recall}}{\text{presisi+ recall}} = 2x\,\frac{0.7618\ x\ 1.0}{0.7618+ 1.0} = \frac{1.5236}{1.7618} = 0.8651 \approx 86.51\ \%$$

$$F1\text{-}Score\ \text{negative class} = 2x\,\frac{\text{presisi x recall}}{\text{presisi+ recall}} = 2x\,\frac{1.0X0.0778}{1.0+0.778} = \frac{0.1556}{1.0778} = 0.1444 \approx 14.44\%$$



```
SVM Accuracy: 0.7665260196905767
SVM Accuracy: 76.65%
SVM Classification Report:
              precision    recall  f1-score   support

     positif       0.76      1.00      0.86       531
     negatif       1.00      0.08      0.14       180

    accuracy                           0.77       711
   macro avg       0.88      0.54      0.50       711
weighted avg       0.82      0.77      0.68       711
```

Figure 6. SVM Classification Report Results

From the image above, the results for macro precision, macro recall, macro F1-score, weighted precision, weighted recall, and weighted F1-score can be calculated as follows:

$$Macro\ Precision = \frac{Precision\_positif+precision\_negatif}{2} = \frac{0.76+1.0}{2} = 0.88$$

$$Macro\ Recall = \frac{Recall\_positif+Recall\_negatif}{2} = \frac{1.0+0.08}{2} = 0.54$$

$$Macro\ f1\text{-}score = \frac{f1-score\ positif+f1-score\ negatif}{2} = \frac{0.86+0.14}{2} = 0.5$$

$$Weighted\ precision = \frac{Precision\_Pos * Support\_Pos * Precision\_Neg * Support\_Neg}{Total\ Support} = \frac{0.76*531+1.0*180}{711} = \frac{583.56}{711} = 0.8203$$

$$Weighted\ recall = \frac{Recall\_Pos * Support\_Pos * Recall\_Neg * Support\_Neg}{Total\ Support} = \frac{1.0*531+0.08*180}{711} = \frac{545.4}{711} = 0.767$$

$$Weighted\ f1\text{-}score = \frac{F1-score_{Pos} * Support_{Pos} * F1-score\_Neg * Support\_Neg}{Total\ Support} = \frac{0.86*531+0.14*180}{711} = \frac{481.86}{711} = 0.6775$$

3.2. World Cloud

The following is a word cloud showing positive, negative, and all words in the data from the public sentiment analysis regarding the VAT increase.

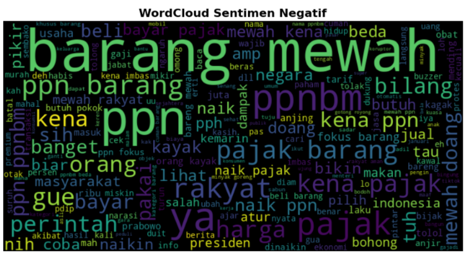


Figure 7. Negative Sentiment Word Cloud

The negative sentiment WordCloud image shows a collection of words that frequently appear in statements with negative undertones regarding the policy being analyzed. Words such as "hit," "pay," "price hike," "regulate," "luxury goods," and "VAT" dominate the visualization. These words indicate public dissatisfaction with the impact of policies that are perceived as burdensome, particularly regarding taxes and expenditures. This reflects the negative perceptions of the public who feel financially and socially affected.

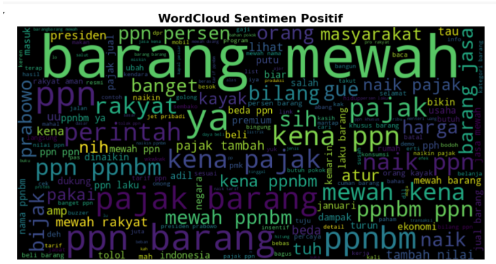


Figure 8. Positive Sentiment Word Cloud

In the positive sentiment WordCloud, words such as "luxury goods," "really," "goods tax," "people," and "economy" appear dominantly. This shows that some people see this policy as a form of consumption control and increased state revenue that can have a positive effect on the economy and equity. This visualization shows that there is a view that a tax on luxury goods is a wise and fair step toward fiscal balance.

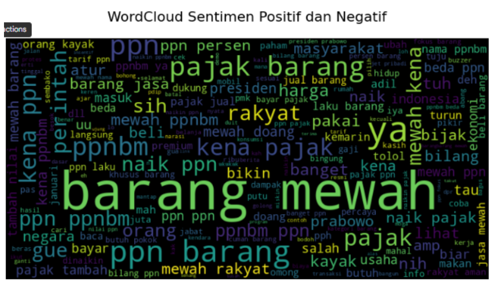


Figure 9. Word Cloud All Words in Data

The overall WordCloud of the data illustrates the frequency of words from the entire text without considering sentiment. Combinations of words from both sentiment camps appear here, such as "ppn," "barang," "kena," "pajak," and "ekonomi." This visualization provides an overview of the main topics and words that frequently appear in public discussions, and serves as a starting point for further sentiment analysis and modeling.

4. Conclusion

Based on the results of the research conducted, the following conclusions can be drawn:

1. After processing 4,000 raw comment data sourced from application x, the data was obtained through a crawling process and then preprocessed, resulting in 3,553 clean comments. The results of lexicon-based automatic labeling show that public opinion on the increase in special VAT on luxury goods shows a high positive sentiment of 2,605 (73.3%) and 947 (26.7%) negative.
2. The sentiment classification results regarding the VAT increase using the support vector machine (SVM) algorithm with a polynomial kernel and a training data to test data ratio of 80:20 yielded an accuracy rate of 76.65%. Performance evaluation of the model shows that SVM is capable of predicting positive classes well but performs poorly on negative classes, with a precision value of 76.18% for positive classes, 100% for negative classes, recall of 100% for positive classes and 7.78% for negative classes, and an F1-score of 86.51% for positive classes and 14.44% for negative classes.

References

[1] W. Aprilia, N. U. Agun, L. K. Datrini, and A. A. B. Amlayasa, "KEPATUHAN WAJIB PAJAK DALAM MEMENUHI KEWAJIBAN PERPAJAKAN ORANG PRIBADI," *Jurnal Lingkungan & Pembangunan*, vol. 6, no. 1, 2022, [Online]. A. Junaidi, "Internet of Things , Sejarah , Teknologi Dan Penerapannya : Review Internet of Things , Sejarah , Teknologi Dan Penerapannya : Review," *Jitter*, vol. I, no. AUGUST 2015, pp. 62–66, 2016.

[2] P. Khairunnisa and T. Herning Sitabuana, "TINDAKAN KORUPSI OKNUM DITJEN PAJAK MEMPENGARUHI PERSEPSI WAJIB PAJAK ATAS PEMUNGUTAN PAJAK," *SIBATIK JOURNAL: Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, dan Pendidikan*, vol. 1, no. 7, pp. 1179–1194, May 2022, doi: 10.54443/sibatik.v1i7.139.

[3] G. Samuel, “Analisis Yuridis Tingkat Kepatuhan Membayar Pajak Masyarakat Indonesia,” *Risalah Hukum*, vol. 18, no. 1, pp. 63–70, 2022, doi: 10.30872/risalah.v18i1.650.

[4] R. Hakim, “Konsep Pajak Dalam Kajian Al-Qur’an Dan Sunnah,” *TAFAKKUR (Jurnal Ilmu Al-Qur’an dan Tafsir)*, vol. 2, pp. 36–48, 2021.

[5] T. H. Salsabila, T. M. Indrawati, and R. A. Fitrie, “Meningkatkan Efisiensi Pengambilan Keputusan Publik melalui Kecerdasan Buatan,” *Journal of Internet and Software Engineering*, vol. 1, no. 2, p. 21, Apr. 2024, doi: 10.47134/pjise.v1i2.2401.

[6]J. Kristovani Siagian, “ANALISIS SENTIMEN MASYARAKAT INDONESIA TERHADAP RENCANA KENAIKAN PPN MENJADI 12% DI MEDIA SOSIAL X DENGAN METODE NAÏVE BAYES,” *5th Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, vol. 3, no. 2, 2024.

[7] N. Rahayu, S. Indri Yani, M. Marwah, and A. Pratama, “Analisis Sentimen terhadap Terorisme pada Platform Twitter menggunakan Support Vector Machine,” *Jurnal Manajemen Informatika, Sistem Informasi dan Teknologi Komputer (JUMISTIK)*, vol. 4, no. 1, pp. 430–441, Jun. 2025, doi: 10.70247/jumistik.v4i1.152.

[8] Salam, R. R., Jamil, M. F., & Ibrahim, Y. (2023). Sentiment Analysis Of Cash Direct Assistance Distribution For Fuel Oil Using Support Vector Machine Analisis Sentimen Terhadap Bantuan Langsung Tunai (Blt) Bahan Bakar Minyak (Bbm) Menggunakan Support Vector Machine. 3(April), 27–35.

[9]Amaliah, F., & Nuryana, I. K. D. (2022). Perbandingan Akurasi Metode Lexicon Based dan Naive Bayes Classifier pada Analisis Sentimen Pendapat Masyarakat terhadap Aplikasi Investasi pada Media Twitter. Journal of Informatics and Computer Science (JINACS), 3(3), 384–393. https://doi.org/10.26740/jinacs.v3n03.p384-393