PUBLIC SENTIMENT ANALYSIS THROUGH TWITTER ON SALES TAX ON LUXURY GOOD INCENTIVES USING MACHINE LEARNING METHOD

Algra[a], Amy Fontanella[*b], Hendrick[c],
[abc]Politeknik Negeri Padang

ABSTRACT

The government's policy on Incentives on Sales Tax on Luxury Goods on the Delivery of Taxable Goods Classified as Luxury in the Form of Certain Motorized Vehicles Covered by the Government for the 2021 fiscal year during the pandemic has invited pros and cons among the public, including twitter users. This study aims to identify and analyze public sentiment toward the government's policy. The method used is machine learning, namely the Natural Language Toolkit (NLTK) analysis. The Results of this study obtained 49 positive sentiments and 65 negative sentiments that describe support, opposition and critical discourse on the policy with a data accuracy rate of 94.23%. The government needs to consider public sentiment and can be taken into consideration in determning the impact of a policy on the economy from all walks of life.

INTRODUCTION

The Coronavirus Disease (COVID-19) pandemic that has hit the world and put pressure on various aspects, including the global economy. All countries in the world are economically affected, including Indonesia. Indonesia's economic indicators have experienced a significant decline, accompanied by the cessation of various economic activities as a result of the pandemic. Various efforts have been made by the government to revive the economic sector, one of which is in the form of tax incentives. One of the steps released by

the Directorate General of Taxes with the aim of easing the burden on the community and facilitating the continued operation of the economy is the provision of the incentives for sales tax on luxury goods on Motor Vehicles as stated in Minister of Finance Regulation No. 20/PMK.010/2021.

Based on the regulation, the purpose of this policy are 1) to increase people's purchasing in the motor vehicle industry sector in order to encourage national economic growth, it is necessary to provide government support for the motor vehicle industry sector 2) to realize the government's support, it is necessary to provide an incentive for Sales Tax on Luxury Goods for Delivery of Taxable Goods classified as luxury in the form of certain motorized vehicle borne by the government, 3) whereas there is no regulation on it so that it needs to be regulated in a Regulation of the Minister of Finance.

In this regulation, the government provides a tax discount for the ≤1.500 cc segment of the sedan category and one drive axle (4x2) which has a local purchase of at least 70 percent. The incentive will be caried out in stages over nine months, lasting in three stages and each stage is valid for 3 months. The sales tax on luxury goods exemption will be granted in the first stage. Then, the second stage is given incentive at 50 percent and 25 percent of the rate will be given in the third stage. The amount of this incentive will be evaluated every 3 months. Policy instruments will be used "PPnBM DTP '' (handled by the government) through the revision of the Regulation of the Minister of Finance which was enacted on March 1, 2021.

This incentive policy requires in-depth analysis and study to obtain an idea of whether the policy is urgent and needed by the stakeholders. The provision of tax incentives, on the other hand, will increase tax expenditure (tax spending), however the effectiveness of this tax incentive on improving the national economy will only be achieved if the policy

is right on target. Tax incentive policies as the main component in tax spending, must go through a selective, measurable, structured process, through in-depth and comprehensive studies and ensure the creation of equality between local and foreign business actors in the business world. Do not let the provision of incentives actually injure the current conditions that are conducive to the business world (Safrina, N., 2019).

The policy of providing tax incentives has received various responses from various groups. The public expresses the pros and cons of this emergence of PMK No.20/PMK.010/2021 in various ways, one of which is by giving opinions on social media. The impact of public opinion on government policies conveyed through social media is very important to study and research. Since the use of social media is currently seen as more effective and efficient in conveying aspiration

The growth of social media in recent years is very rapid. Public has freedom to express their views on almost anything in discussions on social media. There are many micro-blogging websites, such as Twitter, Facebook, Tumbler etc. that can be a medium for delivering opinions. One of the most popular forms of communication on social media is Twitter. Twitter is the most open and simplest platform for sharing public opinion on various topics. The outgrowth of social media in expression of thoughts has resulted in the availability of huge volumes of data from masses. These social networking data and government policy data can be combined to observe some useful public sentiment. The availability of a large amount of data has opened up an arena for conducting research in domain politics and social both (Pabreja, 2017).

This study aims to determine and analyze public responses to government policies related to government policy on Sales Tax on Luxury Good Incentives through twitter. The findings of this study will contribute to the government to reanalyze policies that

are pro and con in the public. Good policy formulation will certainly lead to an increase in the image of the government in the eyes of the public as a government that is responsive to public dynamics.

LITERATURE REVIEW

Research that takes data from social media using data mining techniques, a technique that combines several types of analysis to explore the meaning of structured raw data that can be text, images and videos (Shayaa et al., 2018). For data collection in the form of text, text mining techniques are used, namely the activity of processing text documents using programming languages, data mining and machine learning (F.R. Lucini et al., 2017). The purpose of text mining or text analytics is to analyse public sentiment on a problem or phenomenon and to assess public opinion about products, events, or topics from the perspective of users or groups (Shayaa et al., 2018). The advantage of using text mining is that the data collection process will be faster through the user of social media such as Twitter.

There are several approaches that can be used to perform social media sentiment analysis such as machine learning, lexicon-based and hybrid methods that combine the two previous methods. Sentiment analysis using machine learning conducts training and grouping data based on polarization of text on social media, while sentiment analysis techniques use lexicon, calculate scores and explain how words and sentences in the text correspond to lexicons. The following is an image of the flow of data mining process from social media.

Several previous studies have used the data available on social media to explore, estimate and analyse various things such as the popularity of politicians, public sentiment towards government regulations etc. Alizah, et al., (2020) conducted a sentiment analysis related to the lockdown on Twitter with labelling method. This research produces a

model to predict public sentiment by using Naïve Bayes and Support Vector Machine. The evaluation results from the two algorithms are more than 80%.

Haryani, et al., (2018) analyse Twitter user comments related to determining factors of customer satisfaction with E-commerce using the Lexicon classification method. The result of this study found that the most dominant factors affect customer satisfaction. Uses a specific word dictionary to perform data collection in twitter app using R studio.

Pabreja (2017), focused on the question of whether data from social networking sites can be used to interpret the attitudes of citizens of a country towards various policies. The study used a sentiment analysis of 5.000 tweets with keyword "GTS" using the R language Technique, which is a powerful language widely used for data analysis and statistical computing. The research findings show that social media is a strong and reliable source of public opinion.

Das, et al., (2020), propose a new Twitter-based polarity-polarity framework and words cluster for identifying most popular GST-exclusive word-counts and sentiment analysis within a given range of tweets, which can be the good indicators of word occurrence probability specific to such an event regarding this topic. As A result, this work gives a much more detailed aspect of the lexical level analysis of tweets from several directions, along with the future improvement prospect.

Unlike previous studies, this study uses the Natural Language Toolkit (NLTK) sentiment analysis method. Utilization of this method uses a multidisciplinary approach of science in providing recommendations in policy. Natural language is one of the science parts of computer science, which is a more specific subsection of artificial intelligence. This method allows computers to understand, process and manipulate the language used by humans. The ability of computers to

understand language is not only in the form of audio but also able to translate text or writing. All responses in the form of text or writing, will be analysed with sentiment analysis methods. Sentiment analysis will group all text into 3 types, namely positive sentiment, negative sentiments and neutral. Therefore, by using this method, it can be known how social media users respond to government policies.

Machine Learning

Machine learning is a method that consists of data collection, pre-processing, feature extraction and training. In general, machine learning is divided into three groups in the training process, namely supervised learning, unsupervised learning and reinforcement learning. Machine learning develops to be able to perform classification such as writing, and sound to be specific to text mining. Text mining, also known as text data mining, is the process of converting unstructured text into structured text to recognize concepts, patterns, topics and attributes in the data. One method of text mining is sentiment analysis that has the ability to analyse a problem or product based on text (Shayaa et al., 2018)

Natural Language Toolkit

Natural Language Toolkit or commonly abbreviated as NLTK is a python-based platform developed for NLTK text data processes equipped with more than 50 corpora and lexical resources. In addition, NLTK also provided a library for text mining ranging from classification, tokenization, stemming, tagging and parsing. One of the features that NLTK has is sentiment analysis (Elbagir &Yang, 2019)

NLTK provides various functions which are used in pre-processing of data so that data available from twitter becomes fit for mining and extracting features. NLTK supports various machine learning algorithms which are used for training classifiers and to calculate the accuracy of different classifiers.

Sentiment analysis

Sentiment analysis is the process of determining sentiment and grouping the polarity of text in a document or sentence so that categories can be defined as positive, negative or neutral sentiments. Researchers are currently widely using sentiment analysis to determine public perception. Sentiment analysis can also be equated with opinion mining, because it focuses on opinions that express positive or negative. In sentiment analysis data mining is done to analyse, process and extract textual data within an entity, such as services, product, individuals, phenomenon or topic. The analysis process can include the text of reviews, forums, tweets or blogs with pre-processing data covering the process of tokenization, stop word, deletion, stemming, sentiment identification and sentiment classification.

Nowadays people share their ideas, thoughts or opinions on various social media about various things, ranging from product reviews, political talks, movies, music, sport and others. By conducting sentiment analysis, manufacturers can find out how consumers respond to their products, political parties can find out how the image of their cadres in the eyes of the public, the government can know people's opinions about government policies, and so on. The result of sentiment analysis believed to be more accurate and describe the real situation that the survey methods that have been conducted, because this sentiment analysis is based on public opinion that develops realistically and is natural without being engineered, so that decision making based on sentiment analysis is expected to be more logical, rational and has a relatively small risk of rejection.

Twitter

Twitter is a social media platform that first launched in July 2006 under the name "twitter". Twitter combines two formats, namely social networking media format and blogs or

known as microblogging. One can follow other profile accounts, or called followers. To be able to see the content produced by the profile account without being followed back by the account that is followed. A person cannot comment on content produced by a particular profile account.

Twitter limited the amount of content to 140 characters which are to be known as "tweets". With the format syntax "@username" to convey messages to other accounts or hashtags "#" to signify a specific topic so that others can follow the discussion around the topic. A person can also reply or forward the message using the "retweets ("RT") format. Trending topics on Twitter become parameters for the dissemination of information in the media. A topic can be a trending topic depending on how many Twitter users retweet, reply or mention. The age of trending topics is diverse and cannot be determined (31 percent last for one day, 7 percent last for more than 10 days, "big brother" is a trending topic that lasts up to 76 days) (Rumata, 2017).

METHODOLOGY

This study uses a sentiment analysis approach to see the public response to government policies on Sales Tax on Luxury Good Incentives with the rapid miner application version 7.1. Broadly speaking, the process carried out in this research is data pre-processing and data processing. More complete stages and research processes can be seen in Figure 1.
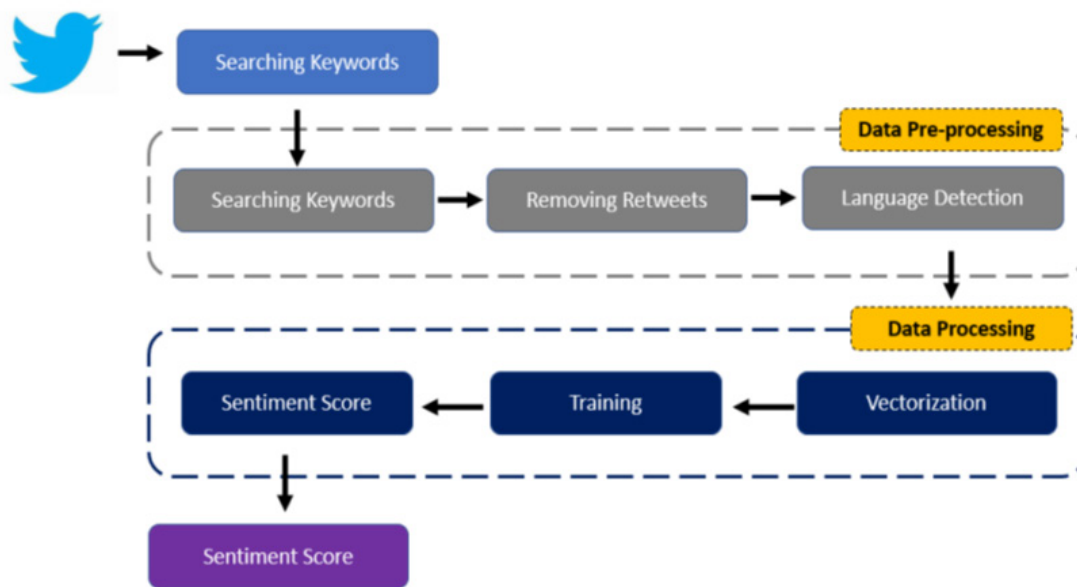

Figure 1 Research Block Diagram

Data collection

Data collection is carried out through several stages:
- Twitter API registration. The difference between the Twitter API and regular twitter lies in the access it has. The Twitter API account can be used through the program so that the text collection process can be done easily.
- Searching keywords using search engines in the NLTK library. The keywords that are searched use Indonesian language settings and try to use words that are easily recognizable on Twitter. This research uses "PPnBM ".

Pre-Processing Data

Pre-processing is the initial process carried out to obtain information about Twitter user responses about the keywords mentioned in the previous sub-chapter. All Twitter user responses will be saved in excel format. Furthermore, this file will be referred to as a dataset. The next pre-processing stage is to remove retweets from Twitter users. This is used to remove responses that sometimes have nothing to do with the topic being discussed on Twitter. The last stage in pre-processing is language detection. This process serves to narrow down the comments obtained on Twitter so that the comments that will be taken are only in Indonesian.

Data Processing

All comments obtained in the data collection process are called datasets. The dataset obtained from the data collection and pre-processing process will be labelled according to the desired target such as positive sentiment, negative sentiment and neutral sentiment. After all datasets are labelled, then the modelling process is carried out using machine learning methods. There are several machine learning methods commonly used in sentiment analysis modelling such as Support Vector Machine (SVM), K Nearest Neighbour (KNN), Naïve Bayes and Backpropagation. After carrying out a series of training processes, a model will be generated that is able to give sentiment to words that appear on Twitter based on the desired topic.

RESULTS

This study uses public sentiment data through the official website www.twitter.com with the Twitter API using the "PPnBM" Keyword. Data collection is carried out randomly after the government releases the Sales Tax on Luxury Good Incentives policy and after the extension of the policy is determined by the government. After the policy was issued by the government, 157 tweets were obtained. Sentiment labels are divided into positive and negative which are done manually. This is because the rapid miner version 7.1 application can only detect words using English. After the data has a sentiment label, the dataset is grouped into testing data and training data. The testing data consist of 78 sentiments, while the training data consists of 98 sentiments. The cleansing process is carried out by tokenizing, transform cases, stop word filters and comparison of classification models so that the design model process is obtained as shown in Figure 2
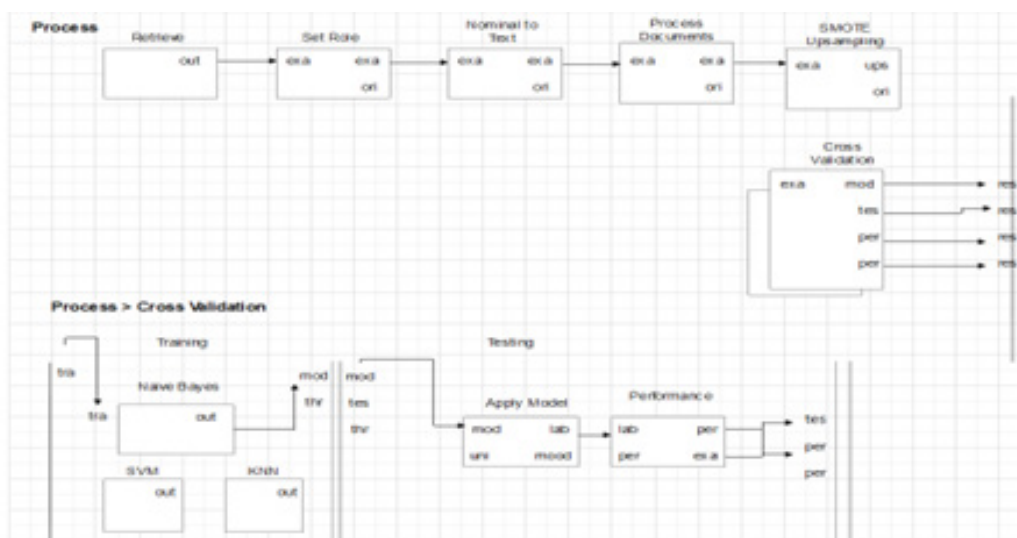

Figure 2 Design Model Process

The result of training and testing show that by using the Naïve Bayes method, an accuracy rate of 84.89% is obtained in analyzing public sentiment. The test data obtained were 122 tweets consisting of 49 positive sentiments and 65 negative sentiment. The testing data was reprocessed using the same design model process after policy was established and an accuracy rate of 94.23% was obtained as shown in Table 1 below:

Accuracy: 94.23% +/-1.04% (micro average: 94.23%)

	true Negatif	true Negatif	Class precision
pred Negatif	582	7	98.81%
pred Positif	68	643	90.44%
class recall	89.54%	98.92%	

Table 1   The accuracy level of the Naïve Bayes Model after data retrieval.

Sentiment results on 122 tweets mentioning the keyword "PPnBM" indicate support for the policy. Positive responses emerged from the automotive industry sector, manufacturers and dealers who flooded social media with attractive sales schemes at low prices. The emergence of the policy in the midst of a pandemic, according to experts, can revive the automotive industry by boosting sales again. On the other hand, there is opposition to the Sales Tax on Luxury Good Incentives policy, tweets depicted in this category in the form of calls, doubts and negative opinions. One of them comes from the sales of used cars. They commented that the issuance of this policy was actually very detrimental. Sales of used cars are declining because the public will delay buying, because the prevailing incentives will lower the price of new cars. The same thing was conveyed by the middle class with low purchasing power. These people commented that this policy was issued at the wrong time because it was issued amidst the controversy over a plan to collect Value Added Tax (VAT) for education services and basic necessities. Basic things such as education and

basic necessities cannot be exploited through tax policy. Unlike the case with a car that from the start was not a primary need. The critical discourse on the incentive policy is reflected in the tweets that suspect that this program is only for the interests of certain groups related to political interest. This policy is considered to want to reduce movements that can affect government power

Based on the results of public sentiment above, it can be analysed that the large number of negative sentiments from the public give signals that there is not enough information provided by the government regarding the importance of the Sales Tax on Luxury Good Incentives policy issued during the current pandemic. The official website of the Ministry of Industry of the Republic of Indonesia states that: in line with the development of the implementation of the policy, the performance of the automotive industry and car sales in the country shows a positive trend. In March 2021, when the Sales Tax on Luxury Good discount was first applied, there was an increase in new car sales of up to 28.85 percent. In fact, in April 2021, the sales surge reached 227% compared to the same period in 2020 (year in year / yoy). Referring to data from the Association of Indonesian Automotive Industry (GAIKINDO), there is an increase in retail sales, from January to April 2021 by 5.9 percent yoy to 257.953 units cumulatively. On a monthly basis retail sales volume has approached the normal level or around 80.000 per month.

This sales spike data reflects that the Sales Tax on Luxury Good incentive program is running successfully, with all parties benefiting, both in terms of automotive business, consumers and the government. The government has also succeeded in obtaining VAT and Income Tax revenues from increased car sales. On the other hand, consumers get a new vehicle at a more affordable price.

The government's active role in providing clear information on social media is very much needed, both before and after

a policy is established. This aims to minimize negative sentiment from the public. This study also shows the importance of government consideration in calculating the impact of policy on the economy from all walks of life. The imposition of taxes on inappropriate objects or the imposition of taxes on certain objects with inappropriate rates will actually have a negative impact on state revenues and/or the national economy at large.

CONCLUSION

The objective of this study is to analyze public sentiment on government policies regarding the provision of Sales Tax on Luxury Good incentives during the Corona Virus (COVID 19) pandemic. The results show that using the Natural Language Toolkit (NTLK) method (data accuracy rate is 94.23%), public responses on social media need to get the government's attention and consideration in making policy. The application of sentiment analysis can also be useful to minimize negative sentiment from the public on a policy, so as to make the government responsive to the dynamics that occur in the public. Further research can be developed by adding data from various social media sources to increase accuracy.

REFERENCES

Safrina, N., Soeharto, A., & Asrina, A. S. (2019). "Menjaga Marwah" insentif perpajakan yang berdampak pada penerimaan pajak di Indonesia tahun 2019. *Jurnal Riset Terapan Akuntansi, 4*(1), 2020

Pabreja, Kavita (2017). Sentiment analysis using twitter data. *International Journal of-Applied Research*, *3*(7), 660-662.

Haryani, C. A., Tohari, H., & Nurrahman, Y. A. (2019). Sentimen Analisis Kepuasan Pelanggan E-commerce menggunakan Lexicon Classification dengan R. Konferensi Nasional Sistem Informasi 2018.

Alizah, M. D., Nugrogo, A., Radiyah., U., & Gata, Windu. (2020). Sentimen Analisis Terkait Lockdown pada Social Media Twitter. Indonesian Journal on Software Engineering.

Rumata, V. M., (2017). Analisis Isi Kualitatif Twitter "#TaxAmnesty" dan "#AmnestyPajak": An approach for sentiment analysis of GTS Tweets Using Words Popularity Versus Polarity Generation Sourav Das1, Dipankar Das2, and Anup Kumar Kolya3

Shayaa, S., Jaafar, N. I., Bahri, S., Sulaiman, A., Seuk Wai, P., Wai Chung, Y., & …Al-Garadi, M.

A. (2018). Sentiment analysis of big data: Methods, applications, and open challenges.*IEEE Access*, *6*, 37807–37827. https://doi.org/10.1109/ACCESS.2018.2851311

Anchal Kathuria Avinash Sharma. (2020). A comparative study of various approaches for sentiment analysis of twitter data.