Journal of Theoretical and Applied Information Technology
30th June 2022. Vol.100. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

THE SENTIMENT ANALYSIS OF THE INDONESIAN PALM OIL INDUSTRY IN SOCIAL MEDIA USING A MACHINE LEARNING MODEL

[1] RANGGA PRADIPTA, [2] RIYANTO JAYADI

[1,2] Information Systems Management Department, BINUS Graduate Program – Master of Information Systems Management, Bina Nusantara University, Jakarta, Indonesia

E-mail: [1]rangga.pradipta@binus.ac.id, [2]riyanto.jayadi@binus.edu

ABSTRACT

This research is a sentiment analysis of opinion data obtained from the Twitter social network about the palm oil industry in Indonesia. The opinion data used is text in Indonesian to show the palm oil industry in Indonesia and text in English to show issues of the palm oil industry happening globally. The palm oil industry in Indonesia is one of the strategic industries engaged in agriculture, so it is necessary to monitor public sentiment with information on the Internet, especially on social media. This research has produced a model for classifying public opinion on the palm oil industry from Twitter data. The data collection technique used is the Twitter Developer API, and it obtained 14,048 words for Indonesian and 12,421 words for English. This data collection begins in July and ends in September 2021. This study uses a machine learning model with the Naïve Bayes Classifier and Support Vector Machine algorithms to separate sentiment into two labeling classes, namely positive and negative, then compares which algorithm best serves to classify opinion data. The model generated by the Naive Bayes algorithm has the highest accuracy value, namely for domestic (Bahasa Indonesia) data of 82% and international (English) data of 85%.

Keywords: *Sentiment Analysis, Machine Learning, Naïve Bayes, Support Vector Machine (SVM), Palm Oil Industry*

1. INTRODUCTION

The palm oil industry is one of the strategic agricultural sectors for Indonesia. In the national economy, the palm oil industry has a strategic role as the largest foreign exchange earner, driving the national economy, building energy sovereignty, a populist economy, and employment [1]. One of the products produced from palm oil is crude palm oil (CPO). In 1980, Indonesia's CPO production had 700 thousand tons [2], then in 2019, Indonesia's CPO production reached 47.18 million tons [3].

In 2016, Indonesia dominated as the world's largest CPO producer by displacing Malaysia, with a 54% market share, while Malaysia was at 32% of world CPO production [2]. If seen from year to year, the palm oil industry in Indonesia continues to grow, and this has become a global competition between other vegetable oils such as coconut, soybean, rapeseed, and sunflower.

The Government of Indonesia has a Public Service Agency that focuses on supporting achievements in realizing sustainable palm oil. Palm Oil Plantation Fund Management Agency (BPDPKS) is a Public Service Agency under the Ministry of Finance that determines the organization and work procedures of the Agency through the Regulation of the Minister of Finance of the Republic of Indonesia Number 113/PMK.01/2015 dated 10 June 2015. The establishment of this BPDPKS implements the mandate of Article 93 of Law Number 39 of 2014 concerning Plantations to raise funds from plantation business actors, better known as the CPO Supporting Fund (CSF), which will support business continuity. Palm oil development program. There are many different goals for a sustainable palm oil development program. These include encouraging research and development, promoting business, improving industrial development infrastructure, supporting biodiesel programs, and replanting palm trees. It also aims to disseminate information about oil palm plantations and teach the community about oil palm plantations [3].

The development of the internet has influenced human life in interacting with each other, which occurs in disseminating information. Information on the Internet can be obtained through websites, blogs, microblogs, social media, etc. Social media is a

medium that currently brings up various kinds of information and news, and opinions that can form a belief related to the information conveyed. Especially if the person has large followers, the information shared can flow quickly in its distribution. Indonesian Social media users in 2021 have 170 million active users, which is 61.8% of the total population [4].

Twitter is a microblogging and social networking service where users can post messages known as "tweets". Registered users can post, like, and retweet tweets, but unregistered users can only read them. With today's big data technology, collecting thousands of tweets, retweeting, and applying them to data analysis on data becomes very easy. Retweets from followers can influence followers and allow for retweets to flow and spread quickly and widely.

To keep up with the current flow of information and news related to the palm oil industry, it is necessary to monitor the news media, knowledge, and opinions that arise from the data on social media and then identify the sentiment towards the data. It is hoped that having sentiment analysis data can provide excellent strategic direction in supporting a sustainable Indonesian palm oil industry and maintaining informed opinion. So that it is expected to provide positive information to explain the black campaign circulating in digital media, the effect of the Black Campaign on the palm oil industry will have an adverse impact, such as preventing the entry of palm oil products into foreign countries such as the European Union (EU) and others.

Digital online news media in Indonesia have published several titles discussing the issue of black campaigns against the palm oil industry in Indonesia. However, the Indonesian government has also tried to suppress the black campaign issues through diplomacy, advocacy, positive campaigns, and other efforts. So far, the strategy used in the black campaign against palm oil products has always been defensive, so more steps need to be made to discuss the positive things that palm oil can do [5].

Previous research has been conducted by testing and analyzing public opinion about Indonesian palm oil from domestic social media. Most of the opinions found from this study related to Sustainable Palm Oil, the Indonesian Economy, Good Palm Oil, Palm Oil Prices, Sustainability, FFB, and Palm Oil prices are ours. Retweets are sentences about how information from a tweet grows and can be identified by grouping positive, negative, and neutral statements [6]. Then another study also conducted a Digital Wallet Review, sentiment analysis of the Machine Learning approach with the Naïve Bayes model, Decision Tree, Support Vector Machine (SVM), and Neural Network. From the results of this study, the best performance was using the Naïve Bayes algorithm. [7]. The SVM approach provides a maximum level of accuracy and can be used as one of the criteria that need to be considered in carrying out stock activities [8]. In other research, by applying sentiment analysis, we will analyze data from social media, such as Twitter, in January 2021, by filtering out the word "COVID-19" and the phrase "vaccine" in Indonesian-language tweets. During that period, negative opinions developed because of the public's lack of confidence in the safety of vaccines. From these results, the government can see that, so far, it is still necessary to disseminate news about how good the vaccine is for the community. Using sentiment analysis in this research will be helpful in the preparation of strategies and steps needed in the future [9]. So that in this research will also be operated using Machine Learning with the Naïve Bayes model and Support Vector Machine (SVM) both for domestic and international issues.

In this case, the author conducts research on information or news related to Indonesian Palm Oil from social media more deeply in the domestic and international palm oil industry by using four keyword categories to be taken, namely "palm", "Biodiesel", "BPDPKS", and "International Palm Oil Issues" This is, of course, expected to be able to obtain more specific data regarding the palm oil industry. For example, on the "palm" aspect, one would see plantations, prices, factories, etc. On the "biodiesel" aspect, industrial information such as biodiesel and oil products would be viewed. For other palm oil companies, "BPDPKS" is expected to be able to view information related to the Public Service Agency (BLU) or Palm Oil Public Service Agency, which is a representative of the government that realizes sustainable palm oil, and "International Palm Oil Issues" is expected to be able to view information globally about information or issues that are happening globally related to the palm oil industry.

The research was conducted using Twitter social media data collected daily using the Twitter Developer API. After collecting, these CSV files will be combined into one file. Then the following process is to give a sentiment label to each tweet that is positive or negative, which will be the training data. The tools used are those written in the Python programming language using the required libraries such as Tweepy, sci-kit-learn, matplotlib, pandas, word cloud, etc. The algorithm models used are

Journal of Theoretical and Applied Information Technology
30th June 2022. Vol.100. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Naive Bayes and Support Vector Machine, imported from the sci-kit-learn library.

The expected result of this research is to produce a machine learning model that can be used to analyze data related to the palm oil industry by automatically giving a positive or negative label. By having this model, a report on the trend of issues that exist in real-time will be automatically generated so that it will be able to help provide insight to stakeholders in determining the future strategic direction of the palm oil industry in the short and long term.

2. RELATED STUDY

2.1 CRISP-DM

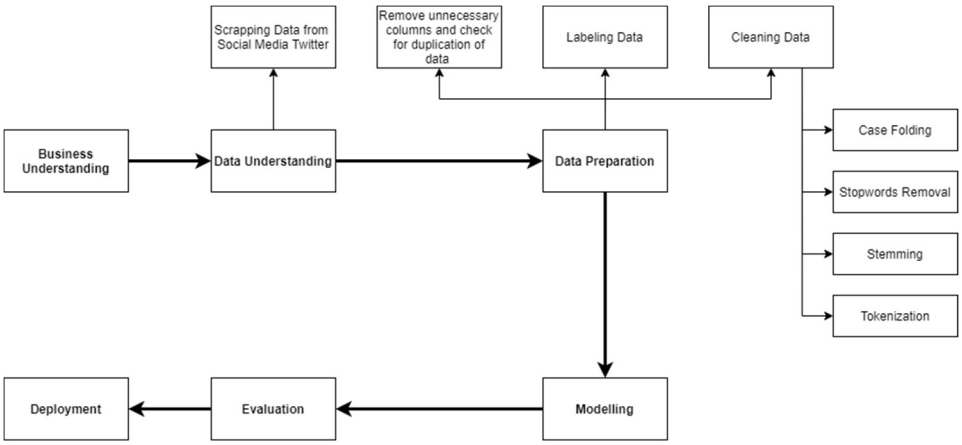


Figure 1: CRISP-DM Diagram process

CRISP-DM (Cross-Industry-Standard-Process model for Data Mining) is a generic data mining process model that offers an overview of a data mining project's life cycle. CRISP-DM is widely used in industry and other academia and is widely considered the de facto standard for knowledge and data mining projects based on user polls and surveys. Using CRISP-DM includes lower costs and time and fewer skill requirements to start data mining [10]. CRISP-DM has six main stages that need to be done, as shown in Figure 1 above.

2.2 Sentiment Analysis

Sentiment analysis is a large field of implementing natural language processes, computational linguistics, and text mining that can analyze one's perspective, sentiment, evaluation, behavior, decision, and emotions, no matter if the writer is discussing a theme, product, and services, institutions, individuals, or other certain activities [11].

Researchers conduct the sentiment analysis to determine what will be conveyed in the text and whether the expression reflects a positive (favorable) or negative (unfavorable) attitude towards a subject. It is essential to correctly determine the semantic relationship between sentiment expression and subject to improve the accuracy of sentiment analysis. The resulting precision results are excellent (75–95 percent, depending on the data) in identifying sentiments on Web pages and news [12].

2.3 Naïve Bayes Classifier

The Naïve Bayes classifier is popular because it is simple, efficient, and performs well in real-world problems. The term "Naïve" assumes that the features are completely independent. Even though the real world does not obey the above-mentioned constraints, this algorithm solves problems that conform to a normal distribution. This strategy combines supervised learning with statistical methods. It is based on a probabilistic model and makes it possible to capture uncertain characteristics in the text by estimating possible outcomes [13].

The following is the Bayes theorem equation:

$$P(C|X) = \frac{P(x|c)P(c)}{P(x)} \quad (1)$$

The Naïve Bayes classification technique is quite intriguing for analyzing large datasets. Improving

the performance of Naïve Bayes classifiers often requires deleting strongly correlated features. Another feature of the Naïve Bayes classifier is that the conditional probability calculations are highly parallelizable and suitable for distributed processing or instance-based processing in a MapReduce setting. Thus, the Naïve Bayes classifier is also helpful for extensive data analytics. This classifier outperforms more complicated models in various situations [14].

2.4 Support Vector Machine

Support Vector Machines (SVMs) are machine learning algorithms that use labeled training data to generate input-output mapping functions. The mapping function can be either a classification function or a regression (used to estimate the desired output). These functions are widely used in classification to convert input data (showing complex nonlinear connections) into high-dimensional feature spaces that are more separable. To classify training data, the most significant hyperplane margin is used. The distance between the two parallel hyperplanes is maximized to separate the data, where the more space between the parallel hyperplanes, the less likely it is to generalize errors. Various linear classifications (hyperplanes) may split data into classes. However, only one hyperplane reaches the maximum separation. SVM classifier data is represented as data points in the machine learning process. Multiple dimensions are possible for these data points. Finally, this raises whether data can be separated using an (n-1) dimensional hyperplane. Hyperplanes must be chosen to minimize the distance between the hyperplane and the nearest data point to get the most significant separation (margin) between two (or more) classes. [15].

SVM is one of the most extensively used supervised machine learning algorithms in sentiment analysis. Cross-validation, on the other hand, will be able to change the ratio of training and testing data until it finds a percentage that maximizes accuracy. Finally, the studied technique's performance was assessed by comparing it to the f-measure score obtained using SVM on the same dataset. It was found that the grid search method worked well, and it is recommended that this method be tested on more data sets to ensure it can return the appropriate values [16].

2.5 K-Fold Validation

A cross-Validation is a statistical approach for assessing a machine learning model's ability. It is often used in machine learning to compare and select models for predictable modeling problems because it is easy to learn, easy to apply, and generates skill estimates that have lower bias [17].

K-fold cross-validation is one of the model validation approaches used to assess the accuracy of a data categorization model. The dataset is separated into two sections: training the model and evaluating its performance. The data in the K-fold cross-validation is divided by k numbers, which correspond to the subset size. If there are k subsets, one is used for testing and k-1 is used for training. This procedure is repeated k times, with each subgroup serving as a test.

2.6 Confusion Matrix

The Confusion Matrix includes the categorization system's actual and predicted classifications. The data in the matrix is extensively used to evaluate the system's performance.

Table 1: The table shows the confusion matrix for the two-class classifiers [11]

	Classified Positive	**Classified Negative**
Actual Positive	TP	FN
Actual Negative	FP	TN

The classification of the confusion matrix is based on Table 1:

- TP : Set of positive examples that have been correctly classified (True Positive)
- FN : Set of examples of "negative" being misclassified (False Negative)
- FP : Set of examples of "positive" being misclassified (False Positive)
- TN : Set of negative examples that have been correctly classified (True Negative)

2.7 WordCloud

Wordcloud is a system that generates word order as a visual image related to the frequency of occurrence of words in a verbal text. The word cloud representation of the text will make it simpler for observers to see the text writer's thoughts and

Journal of Theoretical and Applied Information Technology
30th June 2022. Vol.100. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

viewpoints, allowing it to be used to evaluate written discourse [18].

Segmentation provides context from the word cloud and allows a deeper understanding of subtopics and conversational threads around popular Twitter topics. Specific clusters are displayed as word clouds to help users understand phrases without checking the entire collection of tweets. Thus, it enables users to comprehend better popular themes than other tools such as Wordle (single word cloud) and Infamous (clustered word cloud) [19].

3. METHODS & TOOLS

3.1 Business Understanding

Using the CRISP-DM method, the first step is to understand business. Using PESTEL analysis, this will explain the palm oil industry in Indonesia.

a) Politics

As one of the world's largest CPO producers, Indonesia already has export markets to countries that produce CPO-based products. In this case, of course, the bilateral relations between Indonesia and these countries are going well [20]. From 2015 until the first quarter of 2017, the European Union (EU) implemented the EU Directive on greenhouse emission regulations, seeking to suppress/slow or stop Indonesian CPO from entering the EU market. The enactment of the law is due to the greater dominance of the CPO market compared to other vegetable oils, such as soybean oil, sunflower oil and rapeseed. Negative issues that attack the palm oil industry are starting to be seen on Twitter social media, as in Figure 2.

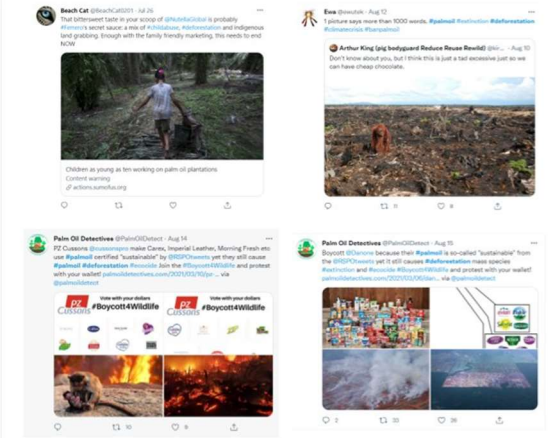


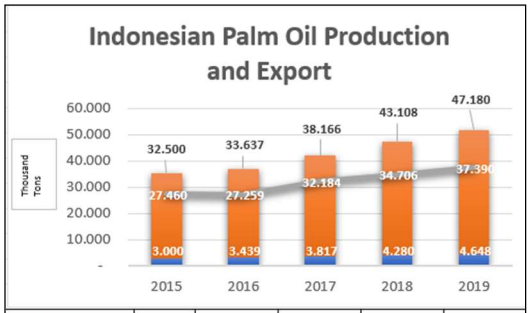
Figure 2: Negative Issues Against the Palm Oil Industry on Twitter

However, the effort to suppress the CPO market is not easy because several factors cause CPO imports to continue, as follows [2]:

- ✓ Lower CPO prices compared to other vegetable oil sources.
- ✓ Local vegetable oil production can only meet two-thirds of its consumption, so one-third is still highly dependent on imports.
- ✓ CPO is relatively more available than other vegetable oils in terms of supply. Support from the Indonesian government in responding to the sustainability issue of the national palm oil industry so that it can be sustainable.

b) Economy

The palm oil industry in Indonesia has a strategic role in the economy, namely as the largest foreign exchange earner in Figure 3; it can be seen that the export value of palm products continues to increase. Then it also affects energy sovereignty as a driver of the national economy, absorbing labor and driving the people's economic sector. All countries globally, including domestic consumption, have received economic benefits and consumption of products produced by the palm oil industry in oleochemical products, oleo food, and biodiesel [21]. The development of the production value of CPO and PKO of palm oil can be seen from 2015 to 2019 in Figure 3.



	2015	2016	2017	2018	2019
PKO Production	3.000	3.439	3.817	4.280	4.648
CPO Production	32.500	33.637	38.166	43.108	47.180
Export	27.460	27.259	32.184	34.706	37.390

Figure 3: Development of Indonesian Palm Oil Production and Exports 2015 – 2019 [22]

c) Social

The areas reliant on the palm oil industry have helped the area's social life. It all starts with rural development, which is about improving the quality of life, reducing poverty, and living with diverse ethnic groups in Indonesia. Small/local farmers work on oil palm plantations in Indonesia as part of the social function of oil palm plantations through People's Core Plantation (PIR), a mix of local

cultural values and modern management designs [21].

d) Technology

The palm oil industry has carried various technologies in its processing process, from upstream products to downstream products. Process technology converts raw materials into physical, chemical, biochemical, or biological processes. For example, there is a technology for processing palm oil into food products, oleochemical products, bioenergy products, etc.[23]. Starting with the use of many technologies in the industry 4.0 era, for example, blockchain technology can help the sustainability mechanisms of this industry through the Roundtable on Sustainable Palm Oil (RSPO) and Indonesian Sustainable Palm Oil (ISPO). The RSPO and ISPO are the only third-party methods for consumers and companies alike to verify and track the sustainability of palm oil consumption. Due to various variables, such as the entire complex value chain with many intermediaries, regulatory barriers, and lack of consumer knowledge, it isn't easy to establish and maintain palm oil traceability across the value chain. It is almost impossible to verify whether palm oil is produced responsibly without a clear and comprehensive traceability mechanism [24]. This needs to be done by implementing blockchain technology, with the advantages of transparency, security, traceability, efficiency, and cost-effectiveness. Then comes the advancement of IoT (Internet of Things) in drones to monitor plant health by taking photos of oil palm plantations with visible and infrared sensors. Given the extent of oil palm plantations, drone footage helps determine plant health [25]. Drones can also monitor oil palm plantations and create maps and information to make it easier for smallholders to obtain ISPO or RSPO certification. Drones can, among other things, detect land boundaries, identify the age and height of oil palm trees, count the number of trees, and estimate yields per hectare. In addition, the data can tell gardeners about the health of oil palm trees and the diseases and pests they might have. This allows them to apply fertilizer and pesticides more quickly.

e) Environment

Oil palm plantations have become contributors to oxygen production by absorbing carbon dioxide. Of course, the expansion of oil palm plantations is associated with deforestation issues, even though this is not supported. Because this expansion is a land-use change (reforestation). The environmental services provided by oil palm plantations, such as the long-term viability of the oxygen cycle, the long-term viability of the hydrological cycle, and the long-term viability of the carbon dioxide cycle, are critical to the overall health of global ecosystems. The more oil palm plantations there are and the more spread out they are, the more carbon dioxide they absorb, the more biomass they make, and the more oxygen they make [21].

f) Legal

Applicable rules and regulations must carry out all activities in oil palm plantations in Indonesia. Therefore, ISPO (Indonesian Sustainable Palm Oil), or the Indonesian Sustainable Palm Oil System, was created as a basic guideline for assessing palm oil growth. The aim is to increase public knowledge about the importance of producing palm oil sustainably, thereby increasing the value and competitiveness of palm oil produced in Indonesia in the world market. Several laws and regulations need to be considered by industry players, starting from the Law on Plant Cultivation Systems, Plantations, Land Legality, Environmental Impact Analysis, Production, Certification, and Seed Distribution, Location Permits, Provisions for Release of Forest Areas, and Granting of Cultivation Rights. for development, etc. [26].

3.2 Data Understanding

In this phase, the data collection process will be carried out in tweet data on Twitter with specific keywords. The method used is the Python programming language and the Tweepy library. Here the author uses a Twitter Developer account for free, so the query data can only be accessed seven days in advance. To overcome this, the author creates a small application script that automatically retrieves data daily from July 2021 until September 2021. This small program generates tweets data in the form of a CSV file.

The data is retrieved from tweets using the Twitter Developer API and categorized as domestic or international:

- ✓ Domestic Palm Issues:
  Query search used: *"sawit OR (tbs AND sawit) OR cpo AND (-filter:retweets AND -filter:replies) -meninggal -duren", "biodiesel OR b20 OR b30 OR d100 AND (-filter:retweets AND -filter:replies)", "bpdpks OR (peremajaan AND sawit) OR (bpdp AND sawit) AND (-filter:retweets AND -filter:replies) -meninggal -duren".*

✓ International Palm Oil Issue:
Query search used: *"palmoil OR (palm AND oil) AND (-filter: retweets AND -filter: replies)"*

3.3 Data Preparation

The approach will focus on gathering data to analyze the needed requirements. The following are the steps of data preparation:

a) Clean unnecessary data such as:
- Delete unneeded columns, and
- Remove duplication of data.

b) Labeling Data
Carry out labeling in grouping data into positive and negative labels. In the label above, we will give the label a number, where the positive label is worth 2, the negative label is 1, and 0 for neutral.

c) Cleaning data
Data cleaning is when the data will be modified to carry out the following method. Data cleaning is a procedure that determines whether data is incorrect, incomplete, or inappropriate and then improves quality by correcting errors and omissions. Data cleaning, in general, lowers errors and improves data quality [27]. The following are the four stages carried out in the data cleaning process:
i. Case Folding
This process is changing uppercase to lowercase.
ii. Stop words Removal
This process is to eliminate words that often appear and have no meaning. [28]
iii. Stemming
This process is to make changes to the word to return to the base word.
iv. Tokenization
This process separates a sentence into resulting words, which have different meanings. This word can later be referred to as a token.

3.4 Modeling

The process will focus on modeling the algorithm that will be used. Here the researcher will use Naïve Bayes classifier and Support Vector Machine (SVM). Using this algorithm is expected to produce the anticipated sentiment analysis.

The following two models are implemented in the following manner:

a) In the Naïve Bayes model, the data generated after the data processing process will be processed by separating the training and testing data. Then, the evaluation results will be made in the form of a confusion matrix consisting of accuracy, precision, recall, and f-measure.

b) In the Support Vector Machine (SVM) model, the modeling is done by splitting the data and then processing it at Term Frequency-Inverse Document Frequency (TFIDF). The TFIDF process assigns weights to the words in the float value that the classifier will use.

4. RESULT & DISCUSSION

4.1 Evaluation

The data may be seen from the model results produced using the Naïve Bayes approach and the Support Vector Machine (SVM).

4.1.1 Naïve Bayes

a) Domestic Twitter data modeling value of AUC: 0.804, and Figure 4 shows an example of a ROC curve.

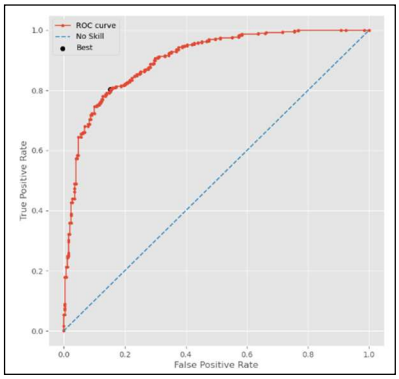


Figure 4: Domestic ROC Curve

b) International Twitter data modeling value of AUC: 0.851, and Figure 5 shows an example of a ROC curve.

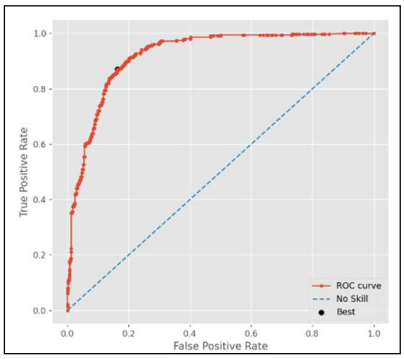


Figure 5: International ROC Curve

Journal of Theoretical and Applied Information Technology
30th June 2022. Vol.100. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.1.2 Support Vector Machine

a) Domestic Twitter data modeling value of AUC: 0.794, and Figure 6 shows an example of a ROC curve.

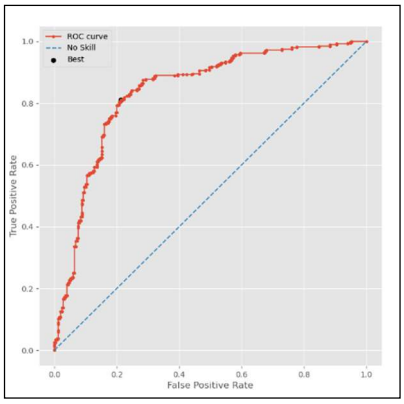


Figure 6: Domestic ROC Curve

b) International Twitter data modeling value of AUC: 0.822, and Figure 7 shows an example of a ROC curve.

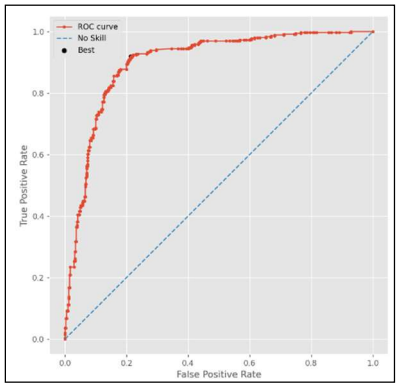


Figure 7: International ROC Curve

4.1.3 Comparison of the two algorithm models

The accuracy, precision, recall, and F-Measure scores are used to create the following comparison table based on the outcomes of this study. According to the findings of the examination of domestic Twitter data, as shown in table 2.

Table 2: Comparison of domestic dataset results

Model	Accuracy	Precision	Recall	F-Measure
Naïve Bayes	82%	80.72%	80.44%	80.57%
SVM	80%	78.84%	79.46%	79.09%

Then, we generated a model using international data from Twitter and found that the findings did not differ much from those obtained using domestic data. The results of the resulting values can be seen in Table 3.

Table 3: Comparison of International dataset results

Model	Accuracy	Precision	Recall	F-Measure
Naïve Bayes	85%	84.94%	85.10%	84.92%
SVM	83%	82.80%	82.25%	82.40%

Based on the table above, the Naïve Bayes algorithm model has an accuracy value of 82% for domestic datasets and 85% for international datasets. In comparison, SVM (Support Vector Machine) has an accuracy value of 80% for domestic datasets and 83% for international datasets. From this value, it can be said that the Naïve Bayes algorithm has a higher value than the Support Vector Machine (SVM). And from the accuracy value generated by the algorithm has produced a value above 80% to be used in the deployment process. Of course, the model that has been made also needs to be continuously provided with training data to accommodate new words found on social media. Or we can give another text source, such as a news site, Facebook, Instagram, etc., to make the resulting model richer.

4.2 Deployment

Each model created from each algorithm will be saved to a file in a *.sav file type. This model file can be used with the Python Pickle Library to make predictions about a sentence based on the machine learning process with the dataset and the data labels on data training. in Figure 8; we can see the snippet code on how to save the model.

```
#saving the model
import pickle

# save the model to disk
# pickling the vectorizer
pickle.dump(clf,
open('models/sawit_svm_vectorizer.sav',
'wb'))
# pickling the model
pickle.dump(SVMclf,
open('models/sawit_svm_classifier.sav',
'wb'))
```

Figure 8: Saving the model with Pickle

Based on the results of the analysis of a total of 12,421 Twitter datasets for English and 14,048 Twitter datasets for Indonesian, taken from July to September 2021, and with the process of making sentiment analysis models carried out using the Python programming language, this research has produced a sentiment analysis model that can be used to predict a sentence related to the Indonesian

Journal of Theoretical and Applied Information Technology
30th June 2022. Vol.100. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

palm oil industry. As for how to load the model with the Pickle library to implement the results of the machine learning model that has been created, such as the code snippet in Figure 9. And in this way, automatic predictions from Tweets can be implemented into an application, for example, a web-based application to automatically generate sentiments from social media Twitter.

```
import pickle

vectorizer =
pickle.load(open('models/sawit_svm_vectorizer
.sav', 'rb'))
classifier =
pickle.load(open('models/sawit_svm_classifier
.sav', 'rb'))

tweet = "Harga minyak goreng naikkkkkk"

text_vector = vectorizer.transform([tweet])
print(text_vector)
result = int(classifier.predict(text_vector))

ResStr = "Netral"
if (result == 0):
    ResStr = "Negatif"
elif (result == 1):
    ResStr = "Positif"

print(tweet, "| Sentiment: ", ResStr)
```

Figure 9: Load the model and predict the text

According to the output of negative issues from domestic issues, which is shown using a Word Cloud as shown in Figure 8, the following is the result of the analysis:

Figure 10: Negative Tweets on Domestic Issues

Within the Indonesian Palm Oil Industry, in the domestic data Word Cloud display, for example, there are the words "price," "garden," "business," "community," "forest," etc. For instance, in the word "price," which represents the selling price of fresh fruit bunches (FFB), which many people were currently discussing when this data was taken, there were issues related to prices here, there are several topics of discussion, for example, FFB prices increase, some areas still feel FFB prices are below standard, there is a decline in prices, then rising CPO prices will affect cooking oil prices. The word "community" appears in the word cloud because there is a problem discussed by many people, which involves local communities and palm oil business actors.

According to the output of negative issues from international issues, which is shown using a Word Cloud as shown in Figure 9, the following is the result of the analysis:

Figure 11: Negative Tweets on International Issues

Explanation, according to Figure 9, the Indonesian palm oil industry has significant challenges where there are various issues such as “orangutan”, “stop, “destruction”, “deforestation”, “boycott,” etc. For this reason, concrete steps are needed to suppress the case, for example, increasing positive content so that news trends can be balanced. For instance, related to land issues, it is necessary to socialize the rules regarding forest and land areas to preserve animals so that nature is well maintained. And then take strict action against anyone who commits a violation to create an industry that is friendly to the environment.

5. CONCLUSION

This study examines the problems of the Indonesian palm oil industry through Twitter data. This Twitter data will be classified based on sentiment analysis using machine learning methods using the Naïve Bayes Classifier algorithm and Support Vector Machine to find domestic and international issues to determine strategic steps forward for the national palm oil industry. Having a generated model is very helpful if you want to implement an automated program that can identify sentiment from Twitter opinion data that has a negative or positive meaning. From the resulting

model, with domestic data, the Naïve Bayes algorithm has a higher accuracy value of 82% than the Support Vector Machine (SVM) algorithm with 80%. Then for International data, the Naïve Bayes algorithm has a higher accuracy value of 85% than the Support Vector Machine (SVM) algorithm with 83%. Here it can be seen that there is a slight difference between Naïve Bayes and SVM.

In the future, further observations can be made to improve the accuracy of the resulting algorithm model continuously so that the training data needs to be studied regularly so that the desired model can be improved. Sentiment results may change depending on the training data and the words used. There is a basic need to understand the learning method to understand the text with unique expressions to make the correct classification. Need to improve query on Twitter Developer API in data retrieval to provide better context for Palm Industry, so Tweet data not related to Palm Industry is excluded from the dataset.

REFERENCES:

[1] T. Sipayung, "Myths and Facts of the Indonesian Palm Oil Industry in Global Social, Economic and Environmental Issues (Mitos vs Fakta Industri Minyak Kelapa Sawit dalam Isu Sosial, Ekonomi, dan Lingkungan Global)," *Bogor-Jawa Barat PASPI-Palm Oil Agribus. Strateg. Policy Inst.*, no. c, pp. 129–130, 2016.

[2] GAPKI, "Recent Developments in the Indonesian Palm Oil Industry (Perkembangan Mutakhir Industri Minyak Sawit Indonesia)," 2017. https://gapki.id/news/3971/perkembangan-mutakhir-industri-minyak-sawit-indonesia (accessed Mar. 28, 2021).

[3] BPDPKS, "BPDPKS Profile - Home," 2016. https://www.bpdp.or.id/sekilas-badan-pengelola-dana-perkebunan-kelapa-sawit (accessed Jun. 20, 2021).

[4] Data Reportal, "Digital in Indonesia: All the Statistics You Need in 2021 – Global Digital Insights," 2021. https://datareportal.com/reports/digital-2021-indonesia (accessed Jun. 20, 2021).

[5] Industri.kontan.co.id, "Facing the black campaign on oil palm, Indonesia will use an offensive strategy (Hadapi black campaign soal kelapa sawit, Indonesia akan gunakan strategi ofensif)," 2021. https://industri.kontan.co.id/news/hadapi-black-campaign-soal-kelapa-sawit-indonesia-akan-gunakan-strategi-ofensif (accessed Jul. 30, 2021).

[6] H. Irsyad and M. R. Pribadi, "Opinion Classification on Indonesian Palm Oil Agriculture Using Naïve Bayes (Klasifikasi Opini Terhadap Pertanian Sawit Indonesia Menggunakan Naive Bayes)," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 6, no. 2, pp. 230–239, 2020, doi: 10.35957/jatisi.v6i2.182.

[7] Yovin Sugandi, "Sentiment Analysis on Digital Wallet Reviews in Indonesia (Analisa Sentimen Terhadap Review Dompet Digital di Indonesia)," 2020. http://library.binus.ac.id/Collections/ethesis_detail/TSA-2020-0071 (accessed Jun. 25, 2021).

[8] N. A. Triany, "Analysis of Social Media Sentiment and Its Correlation to Stock Price Movements by Utilizing Data Mining (Analisis Sentimen Media Sosial dan Korelasinya terhadap Pergerakan Harga Saham dengan Pemanfaatan Data Mining)," 2020, [Online]. Available: http://library.binus.ac.id/Collections/ethesis_detail/TSA-2020-0140.

[9] Pristiyono, M. Ritonga, M. A. Al Ihsan, A. Anjar, and F. H. Rambe, "Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1088, no. 1, p. 012045, 2021, doi: 10.1088/1757-899x/1088/1/012045.

[10] W. Y. Ayele, "Adapting CRISP-DM for idea mining a data mining process for generating ideas using a textual dataset," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 6, pp. 20–32, 2020, doi: 10.14569/IJACSA.2020.0110603.

[11] B. Liu, *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. 2011.

[12] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," *Proc. 2nd Int. Conf. Knowl. Capture, K-CAP 2003*, no. March, pp. 70–77, 2003, doi: 10.1145/945645.945658.

[13] V. Patel, G. Prabhu, and K. Bhowmick, "A Survey of Opinion Mining and Sentiment Analysis," *Int. J. Comput. Appl.*, vol. 131, no. 1, pp. 24–27, 2015, doi: 10.5120/ijca2015907218.

[14] D. Berrar, "Bayes' theorem and naive bayes classifier," *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.*, vol. 1–3, no. September, pp. 403–412, 2018, doi: 10.1016/B978-0-12-809633-8.20473-1.

[15] M. Y. Harahap, "Sentiment Analysis on Online Ticket Booking Application User Satisfaction (Case Study: Traveloka) (Analisis sentimen terhadap kepuasan pengguna aplikasi

pemesanan tiket online),” 2020.
[16] M. Ahmad, S. Aftab, M. S. Bashir, N. Hameed, I. Ali, and Z. Nawaz, “SVM optimization for sentiment analysis,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 4, pp. 393–398, 2018, doi: 10.14569/IJACSA.2018.090455.
[17] Y. Nie, L. De Santis, M. Carratu, M. O’Nils, P. Sommella, and J. Lundgren, “Deep Melanoma classification with K-Fold Cross-Validation for Process optimization,” *IEEE Med. Meas. Appl. MeMeA 2020 - Conf. Proc.*, no. July, 2020, doi: 10.1109/MeMeA49120.2020.9137222.
[18] C. McNaught and P. Lam, “Using wordle as a supplementary research tool,” *Qual. Rep.*, vol. 15, no. 3, pp. 630–643, 2010, doi: 10.46743/2160-3715/2010.1167.
[19] N. Shahid, M. U. Ilyas, J. S. Alowibdi, and N. R. Aljohani, “Word cloud segmentation for simplified exploration of trending topics on Twitter,” *IET Softw.*, vol. 11, no. 5, pp. 214–220, 2017, doi: 10.1049/iet-sen.2016.0307.
[20] S. Silalahi, “Indonesia’s Economic and Political Interests in Cooperation with Slovakia in CPO (Crude Palm Oil) Exports (Kepentingan Ekonomi dan Politik Indonesia Bekerja Sama Dengan Slovakia Dalam Ekspor CPO),” vol. 2, no. 1, pp. 1–10, 2015.
[21] J. H. V Purba and T. Sipayung, “Indonesian Oil Palm Plantation in the Perspective of Sustainable Development (Perkebunan kelapa sawit indonesia dalam perspektif pembangunan berkelanjutan),” *Indones. J. Soc. Sci.*, vol. 43, no. 1, pp. 81–94, 2017, [Online]. Available: http://jmi.ipsk.lipi.go.id/index.php/jmiipsk/article/view/717/521.
[22] Gapki.id, “Indonesian Palm Oil Industry Performance (Kinerja Industri Kelapa Sawit Indonesia) | Indonesian Palm Oil Association (GAPKI),” Feb. 2020. https://gapki.id/kinerja-industri-sawit-indonesia (accessed Aug. 24, 2021).
[23] M. S. Prof. Dr. Ir. Erliza Hambali, “The Role of Agroindustrial Process Technology in the Development of Palm Oil Downstream Industry (Peran Teknologi Proses Agroindustri dalam Pengembangan lndustri Hilir Kelapa Sawit),” *Orasi Ilm. Guru Besar*, 2010.
[24] Indonesia Green Growth Program | Bappenas.go.id, “Vision: Blockchain Technology For Palm Oil Sector In Indonesia (Visi: teknologi blockchain untuk sektor minyak kelapa sawit di indonesia).” http://greengrowth.bappenas.go.id/visi-teknologi-blockchain-untuk-sektor-minyak-kelapa-sawit-di-indonesia/ (accessed Oct. 13, 2021).
[25] A. Stefano, “Utilization of GIS (Geographic Information System) to Monitor Oil Palm Plant Health,” *Bul. Loupe*, vol. 15, no. 02, p. 17, 2020, doi: 10.51967/buletinloupe.v15i02.35.
[26] Zaenal Abidin, “Law on Oil Palm Plantation (UU Perkebunan Kelapa Sawit) | PT Perkebunan Nusantara I,” 2017. http://ptpn1.co.id/artikel/undang-undang-tentang-perkebunan-kelapa-sawit (accessed Oct. 13, 2021).
[27] K. Natarajan, J. Li, and A. Koronios, “Data Mining Techniques For Data Cleaning.”
[28] S. Vijayarani and R. Janani, “Text Mining: open Source Tokenization Tools – An Analysis,” *Adv. Comput. Intell. An Int. J.*, vol. 3, no. 1, pp. 37–47, Jan. 2016, doi: 10.5121/acii.2016.3104.