Journal of Theoretical and Applied Information Technology
15th September 2021. Vol.99. No 17


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

# SENTIMENT ANALYSIS OF INDONESIAN E-COMMERCE PRODUCT REVIEWS USING SUPPORT VECTOR MACHINE BASED TERM FREQUENCY INVERSE DOCUMENT FREQUENCY

[1] **SITI FIDYANTI NURFADILA HADJU,** [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]siti.hadju001@binus.ac.id, [2]riyanto.jayadi@binus.edu

**ABSTRACT**

E-commerce transaction has grown eight times every year in Indonesia. Meanwhile, the number of Indonesian e-commerce customers grows twice every year. The Indonesian e-commerce market has attracted prospects both in terms of management strategies and customer opinions. Therefore, this study aims to provide another perspective to e-commerce management that product reviews reviewed by users can be parameters in determining management strategy. Existing reviews can be collected, processed, and predicted into product segmentation data using the CRISP-DM method. To achieve this goal, the authors compare several machine learning algorithms to determine the best way for sentiment analysis on product reviews contained in five E-Commerce. in Indonesia. Data were collected from product reviews in 3 broad categories: telephone, groceries, and fashion. Then this data is placed by comparing four algorithms, i.e., Decisions Tree, Random Forest, Gradient Boost, and Super Vector Machine). It can be denied that the SVM method is the best method with an accuracy rate of 95.875%. After the prediction model is modelled, the deployment involves connecting the prediction to the dashboard software to display it to E-commerce management and related sellers.

**Keywords:** *Sentiment Analysis, Product Review, E-commerce.*

## 1. INTRODUCTION

Indonesia's e-commerce market is expanding. Each year, the number of active internet vendors in Indonesia doubles, reaching 4.5 million in 2017. According to Mckinsey's report [1], at least $8 billion in consumer spending occurs in Indonesia through online commerce. Additionally, online trading sales will increase eightfold, reaching $40 billion in formal e-commerce by 2022.

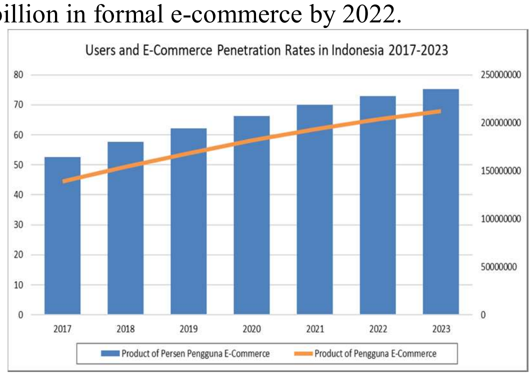


*Figure 1. E-Commerce Growth Projected Data In Indonesia[2]*

Next is the projected data of e-commerce users reported by the databooks.katadata.co.id, which until 2023, the e-commerce business will continue to experience an increase in users[2]. So the author sees that e-commerce in Indonesia is an exciting thing to be analysed further, especially judging from the user's point of view, which is how the user's opinion through a review of the products. The reviews are opinions, views and opinions submitted by users on product assessment based on their experience in purchasing products. This review can be viewed on the e-commerce products page.

Another thing that can be seen from the data projection data of e-commerce users is that there is a strategic need from business people to make sales in e-commerce. This fact is strengthened by research conducted by the University of Amsterdam which shows that a review can increase conversion by as much as 270%. Other data is also demonstrated by the Bright Local Consumer Survey [3] where:

- The average consumer reads an average of ten online reviews before forming an opinion about a local business.
- 57% of consumers will only patronize a business with four or more stars.
- 89% of consumers read business responses to reviews

Because strategy is required to fulfil the demands of a large number of e-commerce consumers, user evaluations of products may be one of the criteria used to determine the source management strategy. Where existing reviews may be gathered and analysed to provide data for product categorization. E-commerce is a commercial activity that involves the sale or purchase of an item or product online or via the use of Internet media. E-commerce has developed into a comprehensive business strategy, providing a diverse array of services and prospects that seem to be disrupting established company structures [4].

Sentiment analysis, often known as opinion mining, is a methodology for determining customers' sentiments, emotional reactions, and perspectives towards a company's goods, brands, or services. Sentiment analysis detects changes in the public opinion of a brand that may support or contradict the company's strategy. When teams should be reformed or new inventive tactics established, sentiment ratings indicate [5].

Machine learning (ML) is a branch of artificial intelligence on algorithmic and statistical models used by computer systems to perform certain activities based on patterns or conclusions. In its learning method, machine learning is divided into three namely(Ella Hassanien Ashraf Darwish Editors, n.d.):

- *Supervised Learning*: ML learns *the mapping of input* and output data with the correct value given by the supervisor to be applied to the new data.
- *Unsupervised Learning*: ML learns *input* without data from *supervisors* by looking for patterns and regularity of existing information.
- *Reinforcement Learning*: ML evaluates the output in the form of provisions or sets of actions that the system has

Numerous previous studies have examined this issue using a machine learning approach to identify Amazon product review sentiment analysis. The machine learning approach divides the product review process into three stages: positive, negative, and neutral[6]. The system employs a similar algorithm for the opinion keyword and an SVM algorithm to determine the document's similarity, allowing it to determine the product review category [7]. In addition, the study of[8], [9], and [10] have studied similar problem including the product rating [11]. Product data used on the previous studies taken from various online stores like Amazon and E-bay.

In this research, we are leveraging data from Indonesian e-commerce. The dataset is enhanced to improve the performance of the program throughout the training phase. During the modeling phase, decision tree and random forest techniques are applied. The algorithm has been fine-tuned to achieve the highest level of accuracy. The evolution of previously produced algorithm models allows the generation of more compelling findings and a comparison of the various classifications employed.

This research adheres to the Cross-Industry Standard Data Mining Process (CRISP-DM). CRISP-DM is a data mining technique that standardizes the strategy and procedure for data mining operations. Six separate processes comprise CRISP-DM. The first component is Business Understanding, which is discussed in detail in Sections 1 and 2. Second, Section 2 addresses Data Understanding, which contains a description of the data source and a tag cloud representation. Thirdly, Section 3 delves into the process of Data Preparation. The fourth step, which was described in Section 3, is the Modeling phase. The fifth phase is Evaluation (Section 3), which involves the presentation of the findings. Meanwhile, Section 5 discusses the research's consequences and deployment suggestions.

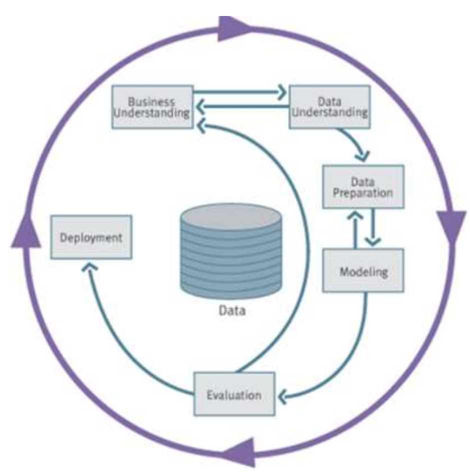

*Figure 2. CIRSP-DM*

Journal of Theoretical and Applied Information Technology
15th September 2021. Vol.99. No 17

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

## 2. METHODOLOGY

This model uses based term frequency inverse document frequency method. according to the CRISP-DM framework that has been discussed previously, the next step is to prepare several things that are needed in processing the model, namely what data and algorithms will be tested.

### 2.2 Data

In term of data preparation with data collection, we determine the data source and process the data to be used by the selected model. The e-commerce sites are JDID, Shopee, Tokopedia, Bukalapak and Lazada with 14742 reviews. Figure 4 show the example of the e-commerce site featuring product reviews.

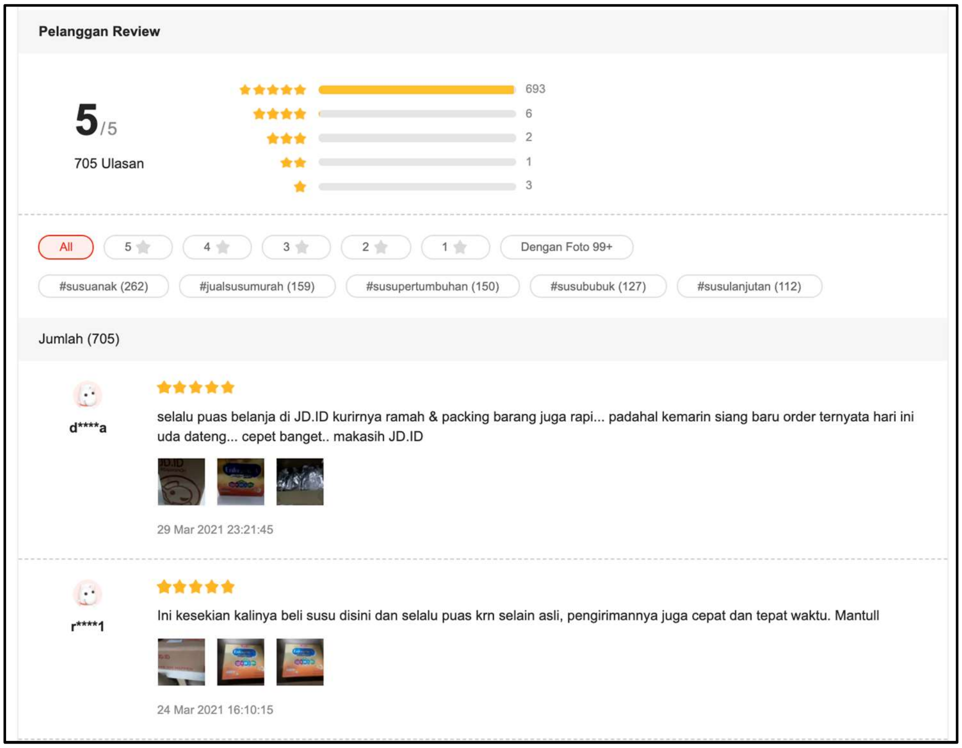


*Figure 3. Product Review On E-Commerce*

Based on observations from e-commerce sites can be explained product review data consists of:

- E-Commerce Name - Text
- Category - Text
- Review – Text
- Category Comment – Text
- Contain Seller – Text
- Contain E-Commerce – Text
- Contain Delivery – Text

Next, Data Exploration and Collection this exploration is carried out to conduct data collection based on the results of exploration. Data is collected using the extension webscrappper.io. Here is the data collected in the Figure 4.

Journal of Theoretical and Applied Information Technology
15th September 2021. Vol.99. No 17


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

| Nama Produk | E-Commerce | URL | Review | Rating | Response |
|---|---|---|---|---|---|
| Enfagrow A+ 3 Original 400 gr - Susu Formula Pertumbuhan Batita | Shopee | https://shopee.co.id/Enfagrow-A-3-Original-400-gr-Susu-Formula-Pertumbuhan-Batita-i.46796711.1396845481 | Cepet sampaaaiiiiiii, terima kasiiihhhhhhhhhhhhhhhhhhhhhhhh | 5 | Positif |
| Enfagrow A+ 3 Original 400 gr - Susu Formula Pertumbuhan Batita | Shopee | https://shopee.co.id/Enfagrow-A-3-Original-400-gr-Susu-Formula-Pertumbuhan-Batita-i.46796711.1396845481 | Expired november 2021, pengiriman cepat | 5 | Positif |
| Enfagrow A+ 3 Original 400 gr - Susu Formula Pertumbuhan Batita | Shopee | https://shopee.co.id/Enfagrow-A-3-Original-400-gr-Susu-Formula-Pertumbuhan-Batita-i.46796711.1396845481 | Sering2 kasi gimmick yg menarik ya. Ini bagus permainan edukasi dan besar | 5 | Positif |
| Enfagrow A+ 3 Original 400 gr - Susu Formula Pertumbuhan Batita | Shopee | https://shopee.co.id/Enfagrow-A-3-Original-400-gr-Susu-Formula-Pertumbuhan-Batita-i.46796711.1396845481 | Kali ini pengiriman cepat. Terimakasih ya enfa ???????????????????? | 5 | Positif |
| Enfagrow A+ 3 Original 400 gr - Susu Formula Pertumbuhan Batita | Shopee | https://shopee.co.id/Enfagrow-A-3-Original-400-gr-Susu-Formula-Pertumbuhan-Batita-i.46796711.1396845481 | Terima kasih, cepat banget. Packing jg aman???? | 5 | Positif |
| Enfagrow A+ 3 Original 400 gr - Susu Formula Pertumbuhan Batita | Shopee | https://shopee.co.id/Enfagrow-A-3-Original-400-gr-Susu-Formula-Pertumbuhan-Batita-i.46796711.1396845481 | Pesanan sudah diterima, packaging rapih, pengiriman standard tidak terlalu lama, respon penjual baik & ramah, kurir baik & ramah, next mau order lagi, terimakasih shoppee, terimakasih seller, sukses selalu! | 5 | Positif |

*(a) Example Of Data Reviews*

| Row Labels | Count of Category |
|---|---|
| **Fashion** | **6362** |
| Negatif | 121 |
| Netral | 320 |
| Positive | 5921 |
| **Groceries** | **4810** |
| Negatif | 80 |
| Netral | 369 |
| Positive | 4361 |
| **phone** | **3570** |
| Negatif | 52 |
| Netral | 129 |
| Positive | 3389 |
| **Grand Total** | **14742** |

(b) *Total Data Collected For Training And Evaluation*
*Figure 4. Data Review On Product Ecommerce*

A tag cloud is often used to envision the most frequently used words in collected product reviews. In this instance, buyers frequently use the terms when writing reviews for e-commerce products. According to the results of the cloud tag visualization, the most prevalent positive words are Bagus, Suai, Cepat, Price, and Delivery. The following are the results returned by the cloud tag:

Journal of Theoretical and Applied Information Technology
15th September 2021. Vol.99. No 17

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


*Figure 5. Tag Cloud For Positive Product Reviews*


*Figure 6. Tag Cloud For Negative Product Reviews*

## 2.2 Algorithms

A decision tree classifier algorithm is a classification algorithm included in supervised machine learning for incremental decision making where data will continue to be divided based on specific parameters. Decision tree classifier has successfully implemented various fields, including signal radar classification, character recognition, medical diagnosis, and speed recognition. The advantage of the decision tree is its ability to parse complex decisions into simple so that it is easier to understand. Some popular decision tree algorithms include ID3, C4.5, and CART[12]. We are employing C4.5. Each root, node, branch, and leaf contribute to the tree's structure. The leaf is the most basic component of the tree structure, while the root is the most complicated. Each attribute in the data collection represents a node. A branch is the segment of the graph that links nodes. The following are the formulae for the decision trees 1-4[13]:

$$Pi = \frac{Ci}{|T|} \quad (1)$$

$$H\ (T) = -\sum_{i=1}^{M} P_i log2P_i(2) \quad (2)$$

$$IG(Y,T) = H(T) - \sum_{i=1}^{M} \frac{|Ti|}{|T|} H(Ti) \quad (3)$$

$$DI(Y = \ -\sum_{i=1}^{M} \frac{|Ti|}{|T|} log_2\left(\frac{|Ti|}{T}\right) \quad (4)$$

If classes and their associated values are expected to be repeated, the probability value of a class is calculated using Equation (1). Where is the total number of connected class values for a certain class. Calculate the entropy of these classes using the equation (2). Given that the class values in the dataset are subsets of their attribute values, the information gained may be determined by dividing the class values by their attribute values. Equation (3) is the outcome. While Equation (4) provides the starting point for calculating the value of the dataset's attribute.

The Random Forest classifier is a classification system that uses multiple tree classifications to forecast which labels each tree will vote for. A random forest is a collection of any number of decision trees. Each decision tree model is trained on a unique collection of records and columns. The latter of which may also be a bit- or byte-vector. Bootstrapping generates row sets equivalent in size to the original input table for each decision tree. At Equation, we have the random forest method (5) [14]:

$$Gini = 1 - \sum_{i=1}^{C} (P_i)^2 \quad (5)$$

Equation (5) determines the Gini for each branch on a node using the class and probability, thereby determining which branches are more likely to occur. Here, pi denotes the class's relative frequency in the dataset, and c denotes the total number of classes [14].

$$Entropy = \sum_{i=1}^{C} - P_i log_2(P_i) \quad (6)$$

Equation (6) is entropy, in which the likelihood of a particular outcome dictates how well the node should branch. In comparison to the Gini index, it requires a greater amount of mathematical computation due to the logarithmic function used to calculate it[14].

Gradient boosting is a learning procedure that combines many simple predictors to produce a powerful committee with performances improved over the single members. The gradient boost tree model has a few key parameters to tune for best fitting the specific problem. Parameter values are carefully tuned according to our problem size, unique features and subsequent constraints to produce the best results for our experiments[15]. The following is the formula for the Gradient Boost:

$$Loss = MSE = \sum (y_i - y_i^p)^2 \quad (7)$$

Equation (7) have $y_i$ as the target value, $y_i^p$ is the prediction, and $L(y_i, y_i^p)$ is the loss function. The predictions should be made in such a way that our loss function (MSE) is as small as possible. We can find the values with the lowest MSE using gradient descent and updating our predictions at a learning rate.

SVM is used to extract unique patterns from acquired signals, which are then classified according to the location of the machine's faults. The SVM's fundamental concept is that it can determine the optimal hyperplane(s) for classifying data into two different classes. SVMs can perform linear or nonlinear classifications based on the features of the data. Numerous SVM classifiers can be used in conjunction to solve multiclass classification problems. The primary support vector machine classifier is built from a simple linear maximum margin classifier[16]. In this case, we are using support vector machines: SMO stands for Sequential Minimal Optimization. A complete quadratic programming (QP) optimization issue must be addressed in order to train an SVM. SMO breaks the issue down into its smallest possible subproblems, which are subsequently solved analytically. Due to the linear equality restriction associated with Lagrange multipliers, 1 is the smallest possible issue. The constraint is then then reduced to for any two multipliers $\alpha_1$ and $\alpha_2$:

$$0 \leq \alpha_1, \alpha_2 \leq C, \quad (8)$$

$$y_1 \alpha_1 + y_2 \alpha_2 = k_1 \quad (9)$$

The issue is simplified to the following analytically by Equations (8) and (9): determine the minimal value of a one-dimensional quadratic function. K is the inverse of the sum of the remaining terms in the equality constraint, which remains constant between iterations.

## 3. RESULT AND DISCUSSION

To identify the Product Review category, several criteria for positive, negative, and neutral reviews are labeled. This criterion is used as training data for the criterion to learn how to differentiate between different types of test review.

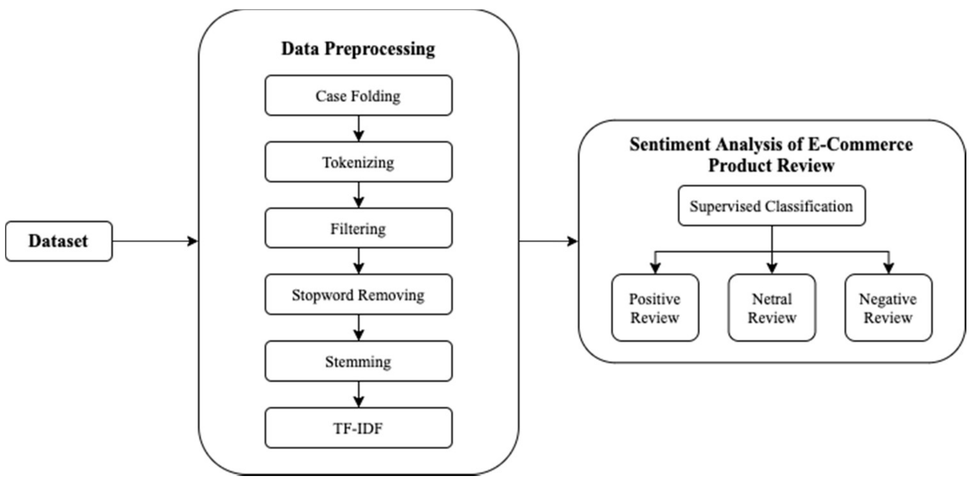


*Figure 7. Stages Of The Product Review Detection Process*

Unneccessary phrases such as numerals, accentuation, conjunctions, abbreviated forms, uppercase and lowercase characters, and skipping the opening and finishing phrases will be deleted during the data preparation step. Preprocessing removes all product review data and as much language as feasible. This step trains the classifier on 80% of the dataset and tests it on 20%. The yield will be calculated using the dataset, which will be produced using decision trees and irregular timberland approaches. The decision tree is a flow diagram in the form of a tree, with each internal node representing an attribute test. Each branch corresponds to the test's outcome. Within the Decision Tree, the leaf node indicates the classes or distribution of classes. Several parameters are used, including quality degree, which is determined by the Gini file and indicates the degree of disparity, and not using the pruning strategy, which has the capacity to reduce overfitting and so increase the quality of expectations. Reduced Error Pruning, keeping the number of records per hub to a minimum of two, the number of records to store for a total of 10,000, the average part point, the number of threads to four, and skipping putative columns that lack space information.

Several parameters employed in each algorithm. The Support Vector Machine makes use of numerous factors, including choosing the document class as the objective, applying a 1.0 intersecting penalty to decide the number of sentences returned for each erroneously categorized point (best value 1), and using a polynomial kernel with total control 1.0, bias

Journal of Theoretical and Applied Information Technology
15th September 2021. Vol.99. No 17

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

1.0, and gamma 1.0. Gradient Boosted Trees acquire the framework via the use of variables, which include the research target column, attribute selection through column attributes, and standard column attributes (string, double, integer). During lessons, this dialog box enables us to manually shift columns from the left pane to the right pane.

The assessment makes use of a number of different assessment measures. The bulk of current systems include the product review categorization, which indicates whether a review is positive, negative, or neutral; however, unlike binary types, there are no positive or negative classifications; hence, what we need to do here is identify TP, TN, FP, and FN for each category. True Positive (TP) refers to instances where positive product review articles are accurately expected to be positive. True Negative (TN) is term for no positive product reviews found. False Negative (FN) is a term that refers to instances where good product review articles are wrongly expected to be positive. False Positive (FP) is a term that refers to instances where positive product review components are wrongly projected to be positive.

Accuracy measures the degree of successful prediction between predicted review categories and the real category. Meanwhile, Precision quantifies the proportion of all detected product reviews that are predicted to be product reviews, resolving the critical issue of determining which review belongs to which category. Precision and accuracy are indicators of performance. By constructing a probability curve, the AUC is utilized to assess and quantify performance on a classification issue. The closer the ROC curve is to the 45-degree angle, the less accurate the model test. On the other side, the higher the AUC, the more accurate the model. The following are the result of analyzing the sentiment of e-commerce product reviews

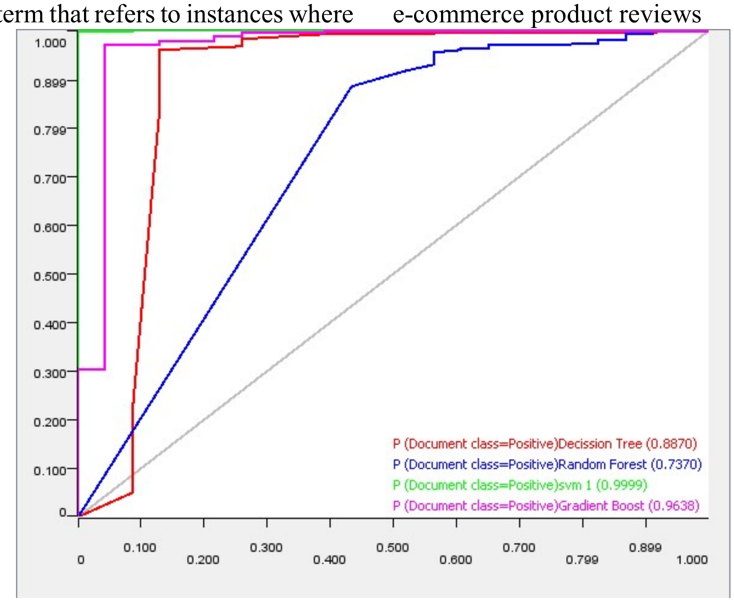

*Figure 8. ROC Graph For Decision Tree, Random Forest, SVM, And Gradient Boost*

*Table 1. Confusionn Matrix Decission Tree*

| Document | Positive | Negative | Netral |
|---|---|---|---|
| Positive | 2536 | 17 | 25 |
| Negative | 24 | 14 | 0 |
| Netral | 36 | 3 | 26 |

*Table 2. Confusionn Matrix Random Forest*

| Document | Positive | Negative | Netral |
|---|---|---|---|
| Positive | 2578 | 0 | 0 |
| Negative | 47 | 1 | 0 |
| Netral | 65 | 0 | 0 |

*Table 3. Confusion Matrix Super Vector Machine*

| Document | Positive | Negative | Netral |
|---|---|---|---|
| Positive | 2663 | 1 | 0 |
| Negative | 15 | 0 | 0 |
| Netral | 12 | 0 | 0 |

*Table 4. Confusion Matrix Gradient Boost*

| Document | Positive | Negative | Netral |
|---|---|---|---|
| Positive | 2576 | 1 | 1 |
| Negative | 32 | 16 | 0 |
| Netral | 62 | 1 | 2 |

*Table 5. Model Result*

| Model | Correction Classified | Accuracy | Wrong Classified | Error |
|---|---|---|---|---|
| Decission Tree | 2.567 | 95.762% | 115 | 4.274% |
| Random Forest | 2.579 | 95.838% | 112 | 4.162% |
| Gradient Boost | 2.578 | 95.801% | 113 | 4.119% |
| SVM | 2.580 | 95.875% | 111 | 4.125% |

As shown in Table 1, the decision tree technique achieved 95.762 percent accuracy, 115 incorrect classifications, and a ROC of 88.7 percent. Meanwhile, table 2 indicates that the random forest technique achieved an accuracy of 95.838 percent, 112 incorrect classifications, and a ROC of 73,70 percent. Meanwhile, as shown in Table 3, the SVM Predictor approach achieved 95.875 percent accuracy, 111 incorrect classifications, and a ROC of 99,9 percent. As shown in Table 4, the Gradient Boost Trees approach achieved an accuracy of 95.801 percent with 113 incorrect classifications, and a ROC of 96,3 percent. The SVM Predictor approach outperformed the other approach.

The machine learning model with the best performance is selected and ready to be deployed. The deployment of this machine learning model will make the model available on the production system, which means the model can be available to provide sentiment prediction results based on actual opinion data sent to the model. Through the application of this machine learning model, businesses can take value or benefit from the model that has been built in order to gain insight into sentiment. The process of implementing the selected model is by publishing the model as a Dashboard Tableau where the algorithm results with the highest prediction score are SVM which will then be entered in the data in the tableau using connecting nodes.

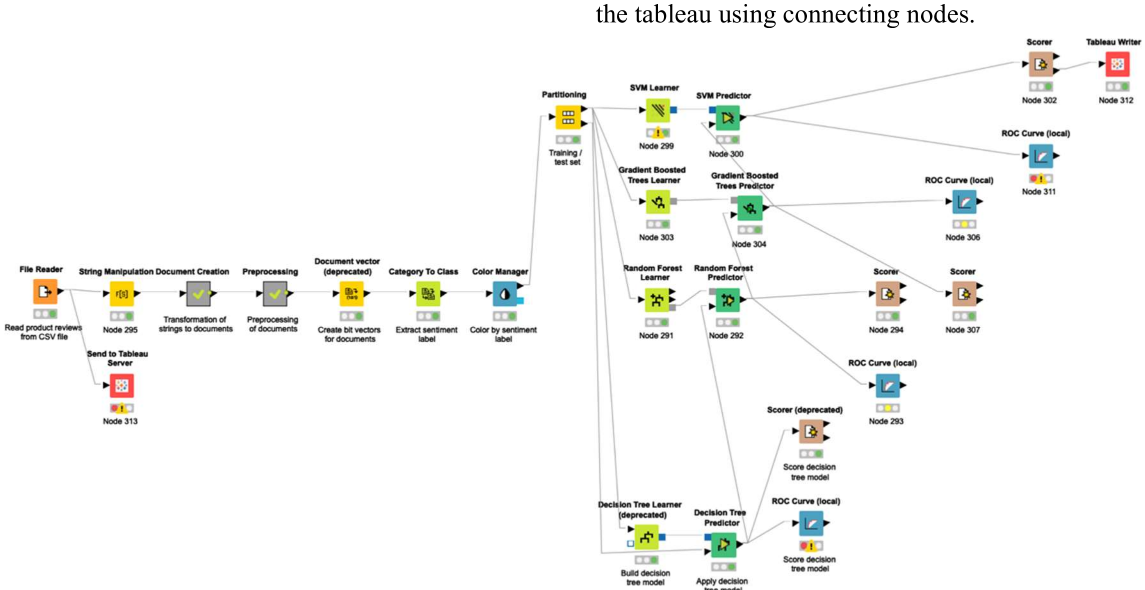


*Figure 8. ROC Graph For Decision Tree, Random Forest, SVM, And Gradient Boost*

Next, After Connected to Tableau, the next step is to create a Tableau dashboard. The first step is to confirm the data set.

After that, create a simple Dashboard Tableau that will display information on the number of reviews, Classification of review categories based on E-commerce, and Classification of review categories based on product categories. After all the required graphs are complete, the last step is to create a dashboard by combining existing graphs, as shown in the image of Figure 9.

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

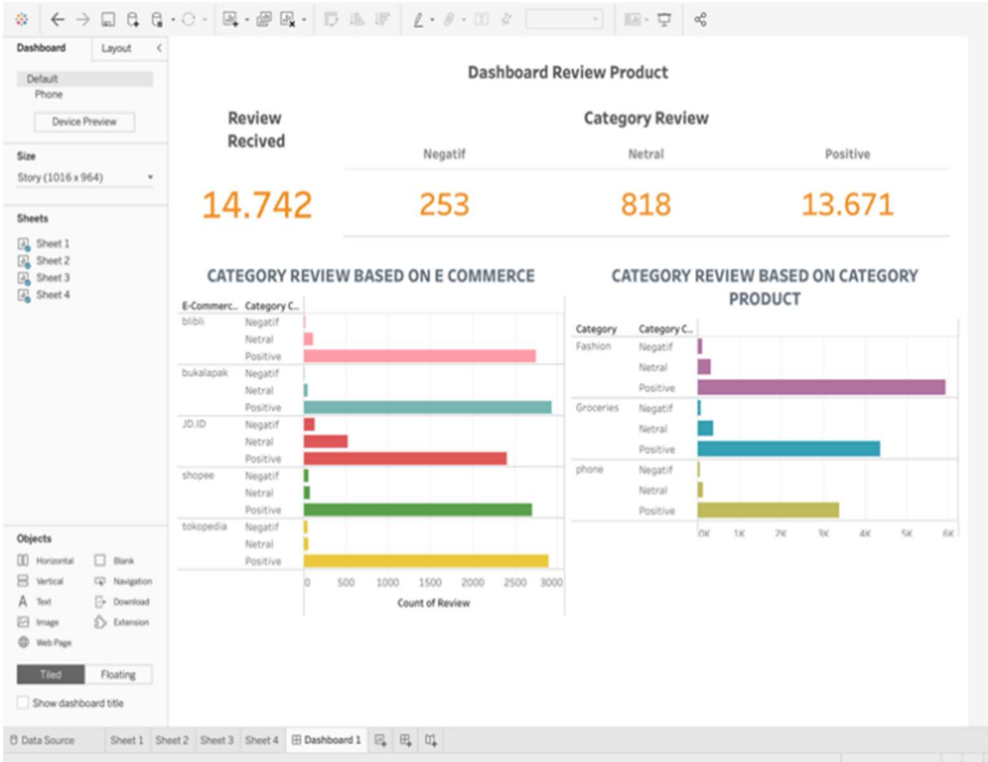


*Figure 9. Tableau Dashbpard For Product Review On E-Commerce*

## 4. CONCLUSION

Sentiment analysis for Indonesian e-commerce product evaluations was created in this work employing machine learning using four distinct kinds of algorithms, namely Decision Tree, Random Forest, SVM Predictor, and Gradient Boosted Tree. Additionally, this study was tokenized, case-folded, normalized, filtered, stop words were eliminated, stemming was done, and TF-IDF analysis was conducted.

The results indicate that product review sentiment analysis is accomplished by computing the weights ascribed to positive, neutral, and negative sentiments in an Indonesian language product review. Following that, classification is used to estimate the weight of words utilizing pre-processing, sentiment analysis, and TF-IDF. These weights will be used to determine the similarity of the term to the test and training data. We can examine the attitudes expressed in product reviews by evaluating the words used in the product review. The accuracy of the sentiment analysis done using SVM techniques was 98.9 percent, the number of wrong classifications was 28, and the ROC value was 0.98, according to the findings.

Based on this result, we can conclude that it is possible to infer that this modeling methodology may be utilized to aid in the identification of product reviews in modern Indonesia. It is envisaged that this model may also be utilized to aid e-commerce enterprises in performing in-depth evaluations of client behavior and trust, resulting in increased revenue and the potential to expand the market.

For further research, we can use more advanced algorithm. In addition, the data set used can be more and more which makes the results more concrete, and lastly this model may be applied to other relevant fields.

## REFERENCES

[1] K. Das, T. Tamhane, B. Vatterott, P. Wibowo, and S. Wintels, "The Digital Archipelago Executive Summary," *McKinsey & Company*, no. August, 2018.

[2] D. H. Jayani, "Tren Pengguna E-Commerce Terus Tumbuh | Databoks," *Databoks*, 2019, [Online]. Available: https://databoks.katadata.co.id/datapublish /2019/10/10/tren-pengguna-e-commerce-2017-2023

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

[3] R. Murphy, "Local Consumer Review Survey: How Customer Reviews Affect Behavior," *BrightLocal*, 2020, [Online]. Available: https://www.brightlocal.com/research/local-consumer-review-survey/
[4] A. S. Punita Duhan, "M-Commerce: Experiencing the Phygital Retail - Google Libros," 2015.
[5] P. Nakov, "Semantic Sentiment Analysis of Twitter Data," *arXiv*, no. May, 2017, doi: 10.1109/CSCI.2018.00251.
[6] T. U. Haque, N. N. Saber, and F. M. Shah, "Sentiment analysis on large scale Amazon product reviews," *2018 IEEE International Conference on Innovative Research and Development, ICIRD 2018*, no. June, pp. 1–6, 2018, doi: 10.1109/ICIRD.2018.8376299.
[7] R. S. Jagdale, V. S. Shirsat, and S. N. Deshmukh, *Sentiment analysis on product reviews using machine learning techniques*, vol. 768, no. January. Springer Singapore, 2019. doi: 10.1007/978-981-13-0617-4_61.
[8] Z. Singla, S. Randhawa, and S. Jain, "Sentiment analysis on product reviews using machine learning techniques," *2017 International Conference on Intelligent Computing and Control (I2C2)*, 2017.
[9] Y. Yang, *Behavioral Pattern-Based Customer Segmentation*. 2011. doi: 10.4018/978-1-60566-010-3.ch023.
[10] E. Tyagi and A. K. Sharma, "Sentiment Analysis of Product Reviews using Support Vector Machine Learning Algorithm," *Indian Journal of Science and Technology*, vol. 10, no. 35, pp. 1–9, 2017, doi: 10.17485/ijst/2017/v10i35/118965.
[11] Anshuman, S. Rao, and M. Kakkar, "A rating approach based on sentiment analysis," *Proceedings of the 7th International Conference Confluence 2017 on Cloud Computing, Data Science and Engineering*, no. January, pp. 557–562, 2017, doi: 10.1109/CONFLUENCE.2017.7943213.
[12] Z. ÇETİNKAYA and F. HORASAN, "Decision Trees in Large Data Sets," *Uluslararası Muhendislik Arastirma ve Gelistirme Dergisi*, vol. 13, no. 1, pp. 140–151, Jan. 2021, doi: 10.29137/umagd.763490.
[13] M. M. Mijwil and R. A. Abttan, "Utilizing the Genetic Algorithm to Pruning the C4.5 Decision Tree Algorithm," *Asian Journal of Applied Sciences*, vol. 9, no. 1, Feb. 2021, doi: 10.24203/ajas.v9i1.6503.
[14] H. Deng, "Interpreting tree ensembles with inTrees," *International Journal of Data Science and Analytics*, vol. 7, no. 4, pp. 277–287, Jun. 2019, doi: 10.1007/s41060-018-0144-8.
[15] T. Pranckevičius and V. Marcinkevičius, "Comparison of Naive Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification," *Baltic Journal of Modern Computing*, vol. 5, no. 2, 2017, doi: 10.22364/bjmc.2017.5.2.05.
[16] L. Zanni, T. Serafini, and G. Zanghirati, "Parallel software for training large scale support vector machines on multiprocessor systems," *Journal of Machine Learning Research*, vol. 7, pp. 1467–1492, 2006.