Indonesian Journal of Computer Science
ISSN 2549-7286 (*online*)
Jln. Khatib Sulaiman Dalam No. 1, Padang, Indonesia
Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

Algorithm *Decision Tree* in Analysis Social Media Sentiment to Understand Consumer Views of Brands

Aggy Pramana Gusman, Harkamsyah Andrianof
apgusman@gmail.com, harkamsyah.andrianof@upiyptk.ac.id
Universitas Putra Indonesia YPTK Padang.

Article Information

Submitted : 20 Dec 2023
Reviewed : 27 Dec 2023
Accepted : 31 Dec 2023

Keyword

Cross validation, Decision Tree, Marketplace, Social Media, Rapidminer

Abstract

Online marketplaces in Indonesia are growing rapidly and have become one of the main destinations for internet and social media users. Various marketplace services are available and accessed by the majority of Indonesians. However, with this development, consumer satisfaction with marketplace services varies, from positive, to negative, to neutral. Many consumers express their reactions on social media, including Twitter. In this study, an analysis of opinions was conducted on posts by online business customers in Indonesia on Twitter from various consumers. However, due to the large number of comments, it is difficult to conclude the customer's opinion about online shopping sites that offer the best services. Even trending topics on Twitter only display hot topics that are widely discussed without clear conclusions. To classify general opinion data on Twitter from e-commerce sites, the first step is to process tweet data using Rapidminer tools to recognize the tweet data. Then, the decision tree algorithm is used to categorize opinion data. The results showed that using cross-validation, the decision tree algorithm achieved an accuracy of 70.27 percent, while using split validation, it achieved an accuracy of 66.95 percent. In this case, better accuracy was achieved using cross-validation. The results of this study can provide useful information for online businesses in Indonesia to improve the quality of their services and increase customer satisfaction. In addition, this study also provides an overview of the importance of utilizing the decision tree algorithm in categorizing opinion data on social media, especially on Twitter, as a tool for analyzing consumer sentiment towards a service or product.

.

A. Introduction

Many internet and social media users in Indonesia have made progress in accessing the marketonline. Marketonline it is a popular destination for social media users in Indonesia to purchase various goods and services. Apart from that, Twitter orplatform Social media is very popular in Indonesia. Many people use Twitter to talk about their experiences when making shopping transactions online[1]. In this social media, users can express their opinions directly and in real-time through posted messages or tweets. Due to its popularity, Twitter has become an important platform in evaluating and analyzing public opinion about the online market in Indonesia[2]. This research analyzed the opinions of online business customers in Indonesia through posts on Twitter [3]. Therefore, Twitter has become one of the most important social media in measuring customer satisfaction and the quality of online market services in Indonesia [3].

Many people use social media sites to express their opinions from somewhere. There are many popular social media sites like Facebook, Twitter, Instagram, and many more. One of the social media most often used to convey opinions is Twitter [4]. A wide range of consumers cannot conclude customer opinions about online shopping sites on Twitter Which E-commerce sites offer the best services, because of the large number of Comments on Twitter and even trending topics on Twitter only display hot topics that are widely discussed without conclusions..

Sentiment analysis is also known as a process, you will find user opinions about various topics or texts submitted by users or the process of determining whether the writing is positive, negative or neutral. This study is concerned with creating opinion analysis from customer opinion data online marketplace on Twitter is used to determine whether there is opinion data. Provide positive or negative feedback so that it can later be used to determine whether the purchase page is rated well online[5][6].

This research uses the decision tree algorithm because it is one of the most popular algorithms and is widely used by researchers. The decision tree algorithm is an algorithm that can be used to form decision trees [7]. A decision tree is a fairly easy method for people to interpret. Decision Trees are prediction models that use tree structures or hierarchical structures [8]. The decision tree concept transforms data into decision trees and decision rules. The main advantage is that damage complicates complex decision-making processes easier for decision makers. It is easier to find solutions to existing problems[9].

This research aims to implement the decision tree algorithm as a sentiment classification model for Twitter social media users in the context of shopping events in the marketplace, with positive, negative and neutral opinion categories [10]. The model developed in this research can provide an overview of public opinion trends in each market, so that it can help online business players understand customer perceptions of their services[11]. In this research, public opinion data was taken from Twitter and processed using the Rapidminer tool to identify emerging opinion patterns. The decision tree algorithm is used to classify these opinions into positive, negative or neutral categories. Thus, the developed model can provide useful information for online businesses in understanding

customer sentiment and taking appropriate actions to improve the quality of their services[12]. Apart from that, this research is also expected to provide insight into the use of decision tree algorithms as a tool for analyzing opinion data on social media, especially on Twitter, in order to understand public opinion trends regarding a product or service[13].

B. Research Methods

For the best sentiment analysis test results, there are several stages that must be carried out. First, collect public opinion or opinion data related to the topic you want to analyze. Second, data processing to clean the data from noise or information that is not relevant to the topic. Next, sentiment classification is carried out to determine whether public opinion is positive, negative or neutral. The fifth stage, the performance of the algorithm used is evaluated using evaluation metrics such as accuracy, precision and recall. Finally, the results of sentiment analysis are interpreted to draw conclusions and information that is useful in decision making or product or service development. It is very important for all of these stages to be carried out systematically and thoroughly to produce accurate and reliable sentiment analysis results.

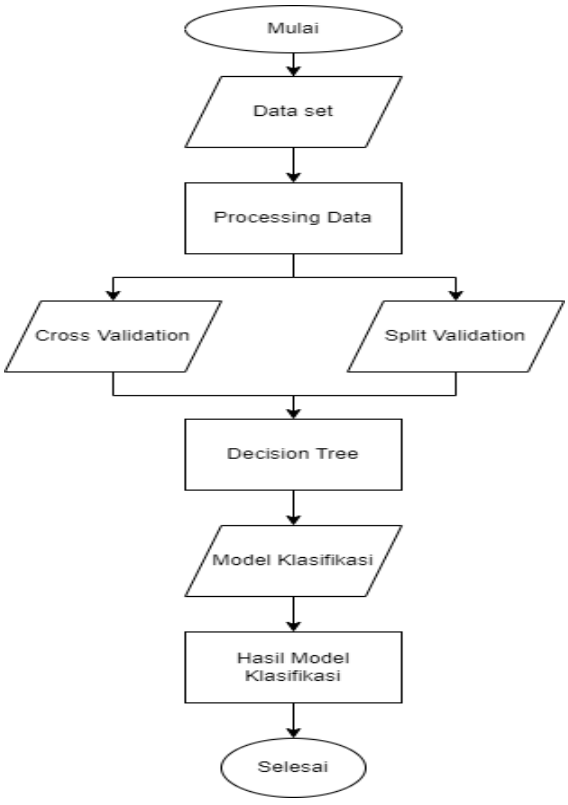


Figure 1. Method for classifying Public Sentiment towards Marketplaces in Indonesia

a. Data collection

The first step is to carry out the sentiment analysis process to collect information. In this research, the data comes from Twitter and obtains information from Twitter.

b. Data Labeling

After carrying out the sentiment analysis process to collect information, the next step is to collect data from various sources relevant to the research objectives. In this research, data was collected from Twitter using a random sampling technique from a number of tweets related to the theme of shopping events in the marketplace. This data collection process was carried out by utilizing the Twitter API and using certain keywords to identify tweets that were relevant to the research topic. The data obtained is then processed and stored in a format that can be accessed by the decision tree algorithm to carry out sentiment analysis. The sentiment categories that are widely used are three classes, namely positive, negative and neutral. The collected dataset can be seen in Table I.

Tabel 1. Dataset And Data Label

ID	ANALISIS SENTIMEN	POST TWEET
1	POSITIVE	Saran aku beli di *official* storenya jgn beli di marketplace karena ga bisa jamin keaslian dan *expdate* nya, mereka cuma jual official di http://hdistore.com atau di center HDI seluruh Indonesia.Googling aja sesuai dom kalian
2	POSITIVE	Sebagai perusahaan teknologi lokal buatan Indonesia, Tokopedia berusaha untuk mengakselerasi transformasi digital UMKM di tanah air, termasuk di daerah.
3	POSITIVE	Mari ramaikan dengan memberi dukungan kepada UMKM Indonesia lewat UMKM EXPO (RT) Brilianpreneur 2023 dari tanggal 14-18 Desember 2023 di JCC atau juga bisa beli produknya lewat *marketplace* loh seperti Blibli.
4	POSITIVE	Kartu kredit terbaru ini hadir untuk melengkapi kebutuhan alat pembayaran bagi penggunamarketplace di seluruh Indonesia
5	POSITIVE	Mewakili Bapak Presiden RI '@Jokowi saya hadir pada event UMKM EXPO (RT) Brilianpreneur 2023. Alhamdulillah, saat ini sudah lebih dari 21 juta total UMKM yang onboarding ke marketplace.
6	POSITIVE	Sobat, selain lebih praktis, perkembangan marketplace yang begitu pesat, dan beragam jenis pembayaran yang terjamin membuat aktivitas belanja *online* di Indonesia terus meningkat.
7	POSITIVE	Pos Indonesia punya kurir perempuan, namanya O Ranger Mawar. Gesit, cepat, aman & tepatwaktu. Komitmen @PosAjaOfficial utk membuat konsumen nyaman. Ada baiknya @tiktok@ShopeeID @LazadaID menjadikan Pos Aja sebagai mitra,
8	NEGATIVE	Sudah cukup kelihatan kalau partai ini tengah menjaring wajah-wajah "centang biru" baru. Kalaucuma berusaha menggaet suara baru, mungkin bisa dicoba cara-cara *marketplace online* di Indonesia: coba ajak member K-Pop atau aktor/aktris K-Drama.
9	POSITIVE	yuk mulai dari sekarang, kita dukung produk UMKM Indonesia dengan beli produk UMKM Brilianpreneur 2023 di *marketplace* kesukaan kalian!

10	POSITIVE	Lalu untuk produk *business to consumer* (B2C), yaitu aplikasi Goers dan *website* http://Goersapp.com, yang merupakan *marketplace* tiket destinasi wisata serta event sebagai *directory activity*, berbagai promo & tour yang dikurasi berdasarkan preferensi pada aplikasi
…	…	…
1165	POSITIVE	tokopedia agar memudahkanku mengelola jualan *online* di marketplace dan socmed, selain itu speknya Huawei juga no kaleng-kaleng udah gak perlu diraguin lagi ??

a. Data Analysis and Testing

After the data has been successfully processed, the next step in this research is to carry out experiments and tests using the decision tree method. In this test, two types of methods are used, namely cross validation and split validation to test the accuracy of the classification model used. The application used in this test is Rapidminer, which is a popular data analysis platform. By using the decision tree method and the Rapidminer application. It is hoped that the most accurate classification model can be found in classifying the sentiments of Twitter social media users regarding the theme of shopping events in the marketplace, with positive, negative and neutral opinion categories.

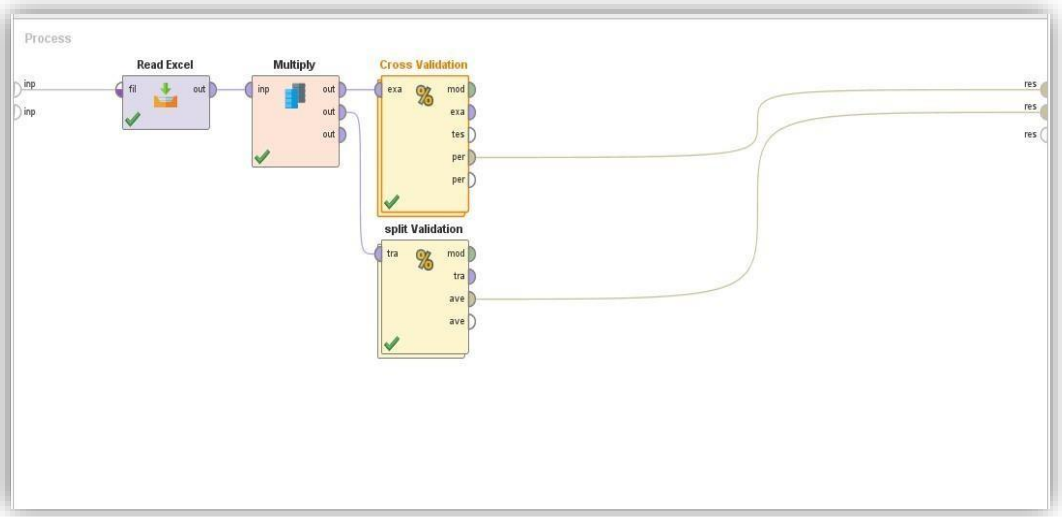


Figure 2. Testing with Cross Validation and Split Validation

C. Results and Discussion

In Figure 4, you can see the evaluation and validation process of the Decision Tree classification method which consists of two parts and three process views. The first part is Training, where this process uses the Decision Tree method to classify opinion data into positive, negative and neutral categories. The second part is Testing, where the trained model is tested and its performance is evaluated to obtain optimal accuracy values. The testing process is carried out using cross validation and split validation methods to ensure the reliability of the classification model that has been created.

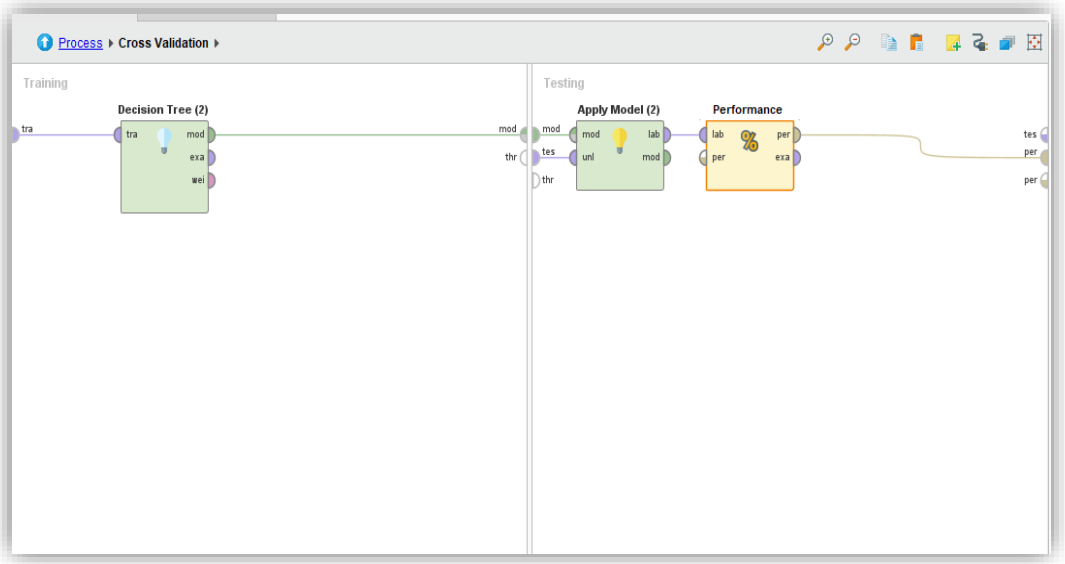


Figure 3. Modeling using the Decision Tree Method

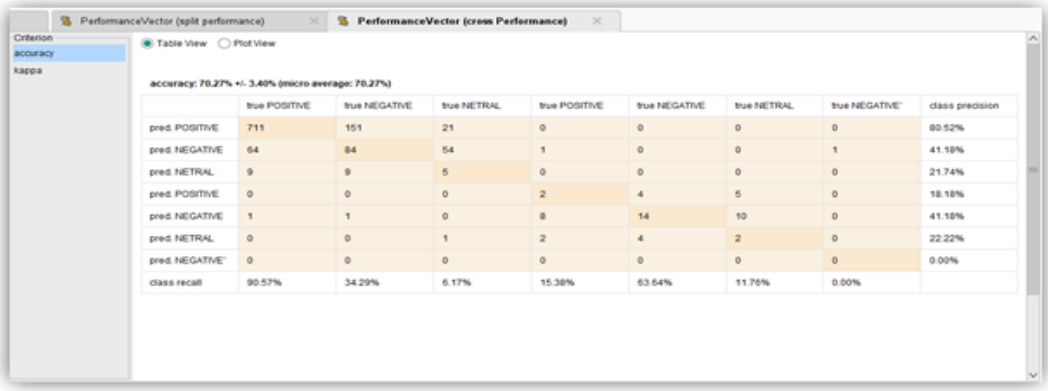
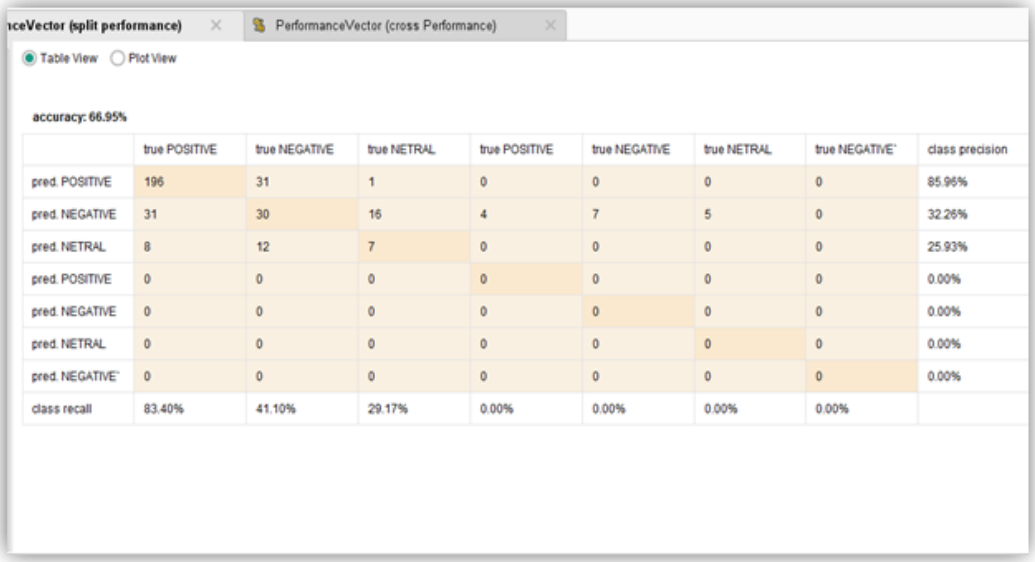


Figure 4. Cross Validation Accuracy Results



	true POSITIVE	true NEGATIVE	true NETRAL	true POSITIVE	true NEGATIVE	true NETRAL	true NEGATIVE'	class precision
pred. POSITIVE	196	31	1	0	0	0	0	85.96%
pred. NEGATIVE	31	30	16	4	7	5	0	32.26%
pred. NETRAL	8	12	7	0	0	0	0	25.93%
pred. POSITIVE	0	0	0	0	0	0	0	0.00%
pred. NEGATIVE	0	0	0	0	0	0	0	0.00%
pred. NETRAL	0	0	0	0	0	0	0	0.00%
pred. NEGATIVE'	0	0	0	0	0	0	0	0.00%
class recall	83.40%	41.10%	29.17%	0.00%	0.00%	0.00%	0.00%	

Figure 5. Split Validation Accuracy Results

Figures 4 and 5 show the accuracy results after testing using the Decision Tree method which produces an accuracy value in Cross Validation of 70.27% and an accuracy value in Split Validation of 66.95%. In this way, the accuracy value in Cross Validation is superior to Split validation.

D. Conclusion

That a marketplace is a market where sellers and buyers exchange services and goods for money or for other services and goods. Twitter is one of the social media most widely used to convey opinions about marketing transactions, this social media also allows users to post messages in real time. Twitter is one of the social media that is involved in a sentiment analysis process so that it can find users' opinions about various topics or texts sent by users or the process of determining whether the writing is positive, negative or neutral. Opinion data is the main step in classifying it so that the algorithm used is the decision tree algorithm. The results of the algorithm using Cross validation achieved an accuracy of 70.27% and using Split validation achieved an accuracy of 66.95%. The accuracy value in Cross validation is superior to Split validation.

E. Acknowledgment

We would like to thank the Padang Computer College Foundation and Putra Indonesia Padang University, which have fully funded this research in the form of a 2023 simlitupi research grant.

F. References

[1] T. F. Stafford and B. Q. Duong, "Social Media in Emerging Economies: A Cross-Cultural Comparison," *IEEE Trans. Comput. Soc. Syst.*, vol. 10, no. 3, pp. 1160–1178, Jun. 2023, doi: 10.1109/TCSS.2022.3169412.

[2] R. R. Ahmed *et al.*, "The Social Media Break-Up: Psycho-Behavioral Measures and Implications," *IEEE Access*, vol. 10, pp. 58116–58135, 2022, doi: 10.1109/ACCESS.2022.3178839.

[3] L. Wang, J. Niu, and S. Yu, "SentiDiff: Combining Textual Information and Sentiment Diffusion Patterns for Twitter Sentiment Analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 10, pp. 2026–2039, Oct. 2020, doi: 10.1109/TKDE.2019.2913641.

[4] F. Alattar and K. Shaalan, "Using Artificial Intelligence to Understand What Causes Sentiment Changes on Social Media," *IEEE Access*, vol. 9, pp. 61756–61767, 2021, doi: 10.1109/ACCESS.2021.3073657.

[5] P. Durga and D. Godavarthi, "Deep-Sentiment: An Effective Deep Sentiment Analysis Using a Decision-Based Recurrent Neural Network (D-RNN)," *IEEE Access*, vol. 11, pp. 108433–108447, 2023, doi: 10.1109/ACCESS.2023.3320738.

[6] P. Vyas, M. Reisslein, B. P. Rimal, G. Vyas, G. P. Basyal, and P. Muzumdar, "Automated Classification of Societal Sentiments on Twitter With Machine Learning," *IEEE Trans. Technol. Soc.*, vol. 3, no. 2, pp. 100–110, Aug. 2021, doi: 10.1109/TTS.2021.3108963.

[7] C. H. Hsu, "Optimal Decision Tree for Cycle Time Prediction and Allowance Determination," *IEEE Access*, vol. 9, pp. 41334–41343, 2021, doi: 10.1109/ACCESS.2021.3065391.

[8] K. Saengtabtim *et al.*, "Predictive Analysis of the Building Damage from the 2011 Great East Japan Tsunami Using Decision Tree Classification Related Algorithms," *IEEE Access*, vol. 9, pp. 31065–31077, 2021, doi: 10.1109/ACCESS.2021.3060114.

[9] J. Zhang, H. Jia, and N. Zhang, "Alternate Support Vector Machine Decision Trees for Power Systems Rule Extractions," *IEEE Trans. Power Syst.*, vol. 38, no. 1, pp. 980–983, Jan. 2023, doi: 10.1109/TPWRS.2022.3220088.

[10] J. Singh and P. Tripathi, "Sentiment analysis of Twitter data by making use of SVM, Random Forest and Decision Tree algorithm," *Proc. - 2021 IEEE 10th Int. Conf. Commun. Syst. Netw. Technol. CSNT 2021*, pp. 193–198, 2021, doi: 10.1109/CSNT51715.2021.9509679.

[11] M. Adnan, R. Sarno, and K. R. Sungkono, "Sentiment Analysis of Restaurant Review with Classification Approach in the Decision Tree-J48 Algorithm," *Proc. - 2019 Int. Semin. Appl. Technol. Inf. Commun. Ind. 4.0 Retrosp. Prospect. Challenges, iSemantic 2019*, pp. 121–126, Sep. 2019, doi: 10.1109/ISEMANTIC.2019.8884282.

[12] E. Daniati and H. Utama, "Decision Making Framework Based on Sentiment Analysis in Twitter Using SAW and Machine Learning Approach," *2020 3rd Int. Conf. Inf. Commun. Technol. ICOIACT 2020*, pp. 218–222, Nov. 2020, doi: 10.1109/ICOIACT50329.2020.9331998.

[13] Rizkiansyah, A. Herliana, D. P. Alamsyah, and T. F. Tjoe, "Comparison of the K-Nearest Neighbor and Decision Tree algorithm to the Sentiment Analysis of Investment Applications Users in Indonesia," *2022 7th Int. Conf. Informatics Comput. ICIC 2022*, 2022, doi: 10.1109/ICIC56845.2022.10006970.