ILKOM Jurnal Ilmiah Vol. 14, No. 3, December 2022, pp. 339-347
Accredited 2nd by RISTEKBRIN No. 200/M/KPT/2020; E-ISSN 2548-7779 | P-ISSN 2087-1716

Research Article **Open Access (CC–BY-SA)**

Sentiment analysis of customer satisfaction levels on smartphone products using Ensemble Learning

Muhammad Ma'ruf [a,1]; Adam Prayogo Kuncoro[a,2]; Pungkas Subarkah[a,3,*], Faridatun Nida[a,4]
[a] *Universitas Amikom Purwokerto, Jl. Letjend Pol. Soemarto No.127, Banyuma and 53127, Indonesia*
[1] *muhmaruf354@gmail.com;* [2]*adam@amikompurwokerto.ac.id;* [3*]*subarkah@amikompurwokerto.ac.id,*
[4]*faridatun.nida@amikompurwokerto.ac.id*
* *Corresponding author*

Article history: Received August 24, 2022; Revised October 12, 2022; Accepted November 29, 2022; Available online December 20, 2022

Abstract

Technological developments create new ways for people to trade, especially the discovery of new technologies that make people want to have the technology to trade. Technological sophistication creates a new way of trading, namely with e-commerce applications, but there are conditions where the seller cannot know the level of satisfaction from his customers and also the problems experienced by his customers if only seen based on the rating in the case. in buying and selling smartphones. From these problems, a solution emerged to create a system that can filter out negative and positive comments, this study uses machine learning using the K-Nearest Neighbors, SVM, Naive Bayes algorithm with hyperparameters taken from previous research. The researcher uses the ensemble learning method with the Voting Classifier technique, which is an algorithm to combine several algorithms made. From the results of testing the highest accuracy was obtained by SVM with an accuracy value of 91.18% while the ensemble learning method got an accuracy value of 89.22% but for SVM the difference in the value of training and testing accuracy was 7.1% while for the ensemble learning method the difference in training accuracy and testing is 4%. The conclusion of the research objective is that the ensemble learning method can help improve the performance of the algorithm for commenting sentiment analysis on smartphone products.

Keywords: E-commerce; Ensemble Learning; K-Nearest Neighbors; Machine Learning; Naive Bayes; Smartphone; SVM; Voting Classifier.

Introduction

The rapid technological development due to globalization has spread worldwide and it changes human life. The presence of this technology makes everything easy and fast because many things can be done instantaneously. Many jobs have changed because of the technology, it even takes over some professions, research says that globalization related to the economic field can increase the unemployment rate, especially in Indonesia[1]. Even so, people can still take advantage of technological developments as a place to earn additional income.

Many economic interactions have changed due to the technological advancement, one of which is trade. In the present time, many people offer their wares through social media platforms. Utilization of technology as a large platform to improve business strategies, one of which is by utilizing social media which is very superior in buying and selling products and services [2]. In its implementation, it is like traditional buying and selling process. The difference is only in the online interaction. Sellers and buyers do not meet in person. There are only pictures provided by the seller. If both agree, the goods will be sent immediately to the buyer. The use of technology to sell in this way has proven to be effective and is increasingly favored by many people, because they can buy goods such as smart phones only from home.

The innovation of technological developments creates a special buying and selling platform or what is often called e-commerce which is a market share in the digital world. The presence of e-commerce that is not existed in conventional business practices offer many people to find it easy to sell their wares through e-commerce applications [3].

E-commerce supports economic development, but it has a weakness. One of them is that feedback given by buyers determines a business operation. In an e-commerce there is such a thing as a feedback and rating that can assist recommendations for other buyers. This will help the seller to be able to see feedback from his customers, on the other hand sometimes a rating assessment with comments does not match with the reality. There are many people who take advantage of it by flooding bad feedback to lower the rating of a certain business entity. This is mostly done in a smartphone.

So from these problems the author is interested in analyzing comment sentiment on e-commerce applications. In the previous research on sentiment analysis on the Google Play Store with the SVM classification method (support vector machine) and maximum entropy classification indicated good accuracy on the SVM classification method[4]. Another research about sentiment analysis using SVM obtained good accuracy, so the authors were interested in using SVM classification for sentiment analysis[4]–[7].

This system development method also used several algorithms such as Naïve Bayes and K-nearest neighbor. In the previous studies Naïve Bayes worked well with small training data sizes[8], [9]. In addition, research on classifying tweet data, it was concluded that K-nearest neighbor had better accuracy than SVM[10].

For the current study, Ensemble Learning method with the Voting Classifier technique were combined with several classification systems to produce optimal results[7], [8], [11]. Thus, the results of the research model created can predict positive and negative sentences and separate the comments to calculate the percentage between negative and positive comments.

Method

In the system design process there were several steps that must be taken, in the sentiment analysis process flow shown in **Figure 1**.

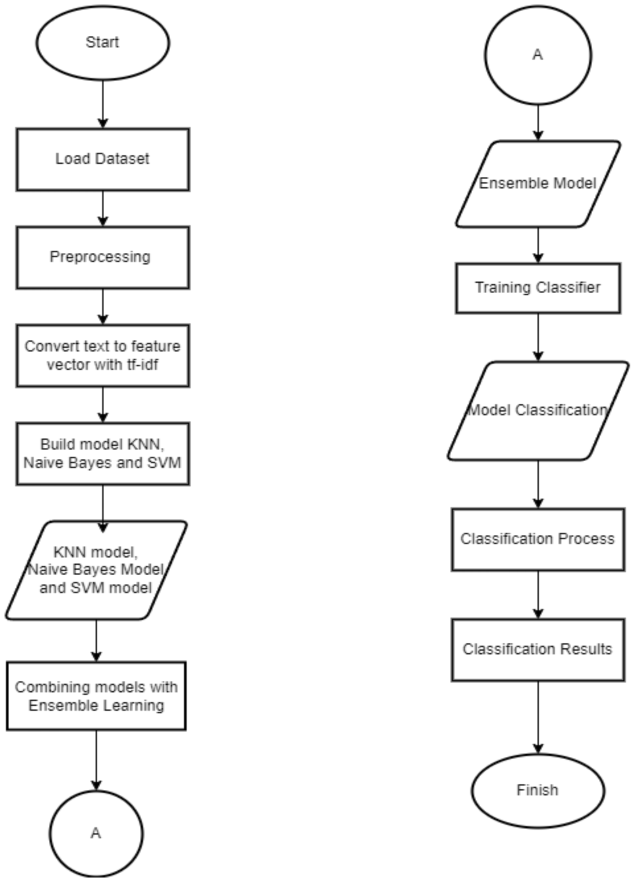

Figure 1. Sentiment Analysis Flow

A. Data Collection and Exploration

The data collection can be taken from various sources such as sensor data, data collection from an institution or from the internet.

In this study, data collection was taken from comments on e-commerce platforms, namely Shopee and Tokopedia, the data taken were 510 comments consisting of negative and positive comments. The number of data for each negative and positive was 255. In exploration, the chosen data was clear so it can be processed.

B. Preprocessing

Before the data was processed, the comments would go through a preprocessing stage. The first step was case folding in which the equalization of text into lowercase letters, then punctuation elimination to remove unimportant characters and punctuation marks. The last was cleaning words that did not have much impact on the sentiment of the sentence and this study used an Indonesian list. This study did not use stemming and lemmatization or changing words into basic words, because in the previous studies, the use of stemming and lemmatization did not provide a stable accuracy increase[12]. In fact, the time to complete a sentiment analysis increased up to 310 times. This was very bad because stemming can reduce the efficiency of sentiment analysis.

C. Feature Extraction

Feature extraction or vectorization is the stage to convert words into vector form, integers or floats. In word weighting, one of the popular techniques used is TF-IDF. The use of TF-IDF weighting could help in the classification process[13][14][15].

TF-IDF is taken from TF and IDF. It is a method to determine how far the terms are related to the document by giving each word a weight[15]. This weighting is the result of multiplying term frequency and inverse document frequency[16]. Term Frequency is equal to the number of times a word appears in a particular document. It is calculated as an **Equation 1**:

$$TF = \frac{Word\ frequency\ in\ document}{Total\ words\ in\ document} \tag{1}$$

Inverse Document Frequency is for a certain word equal to the total number of documents, divided by the number of documents containing a certain word. The entire term log is calculated to reduce the impact of division. It is calculated in the **Equation 2**:

$$IDF = \log \frac{Total\ number\ of\ documents}{Number\ of\ documents\ containing\ the\ word} \tag{2}$$

D. Building Algorithm Model

Making the model for this classification system used several predetermined algorithms, namely K-Nearest Neighbor, Naive Bayes and SVM. After making the three models, all models were combined with the ensemble learning method using the Voting Classifier technique to get the base majority of the three algorithms.

- Super Vector Machine (SVM)
  SVM is an algorithm that is included in supervised learning. It works by looking for the best hyperplane by maximizing the distance between classes[5]. Hyperplane is a function that can be used to separate between classes. In this sentiment analysis research, SVM produced the best hyperplane when separating two categories, namely positive and negative. From previous research[5], SVM had high accuracy with linear kernel.
- K-Nearest Neighbor (KNN)
  KNN is an algorithm that functions to classify based on learning data taken from the value of k closest neighbors[17]. Where k is the nearest distances. This algorithm works by finding a sample of k values that are closest to the k input sampel, and determining the decision results from the largest number of samples from k samples[5]. According to previous research[5], the value of k that is quite good to use is 20.
- Naive Bayes
  Naive Bayes Classifier is classification method based on Bayes' theorem. It is a classification method using probability and a statistical method proposed by British scientists. It is to predict future odds based on past experience [18]. The advantage of using this method is that only a small amount of training data is required to determine the estimation parameter during the classification process [18]. The basic theory of the Bayes method can be formulated as follows **Equation 3**:

$$P(A|B) = \frac{P(B|A)P(A)}{PB} \quad (3)$$

Description:
P(A|B) = occurrence probability of A if B occurs.
P(B|A) = occurrence probability of B if A occurs.
P(A) = occurrence probability of A regardless of any event.
P(B) = occurrence probability of B regardless of any event.

- Ensemble Learning – Voting Classifier
  Ensemble learning is a way of how an algorithm learns data using a combination of several algorithms or models to obtain output with greater accuracy [7][19]. Some ensemble learning methods that can be utalized are Voting, bagging, Boosting, and Stacking. In this research, the author used Voting. The Voting classification method is available in the scikit-learn framework. The method produced an output that would be combined with the average rule of probolity voting. Then, the class with high average score would be selected for the final result[19].

Results and Discussion

At the testing stage, it was used to determine the performance of several models. The best performance can be implemented to filter negative and positive comments in e-commerce applications. According to previous research on making the KNN model, the best k value of 20 and for SVM was a linear kernel[5].

A. Dataset

The data was taken from comments in Shopee and Tokopedia platforms about smartphone products. 510 data were collected, of which 255 were positive comments and 255 were negative comments. Dataset was publicly accessible on (https://github.com/notfound313/sentimen-analysis/blob/main/HP_K.csv). In this study, we tested the model with 80:20 and 70:10 ratios of training and test data. Then the performance of each ratio of data was tested. An example of the data used is shown in **Table 1**.

Table 1. Dataset Samples

Label	Komentar
1	Mantapppp Original Masih segel Sesue deskripsi Fast respon Rekomended
1	packing aman respon cepat
0	tidak ada plastik buat hp nya garansi sdh sobek
0	kecewa sama kurirnya aja kelamaan
1	mantep puas banget

B. Evaluationg

To find out the performance, accuracy, recall, precesion and F1 Score were measured by the following equation. Based on **Table 2** is Confusion Matrix.

Table 2. Confusion Matrix

		Predicted Class	
		Yes	No
Actual Class	Yes	True Positive	False Negatif
	No	False Positif	True Negatif

Accuracy is one of the performance measurements using intuition. It is simply the ratio of correctly predicted observations to total observations[20]. Accuracy is a good measure, but it requires a symmetric dataset with approximately the same false positive and false negative values[13][20]. The formulation can be seen as follows **Equation 4**:

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (4)$$

Precision is the ratio of the correctly predicted positive observations to the total predicted positive observations[13][20]. It can be formulated as follows **Equation 5**:

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

Recall the number of correctly assigned users to the class the total number of users in the class[13][20]. Formulated as follows **Equation 6**:

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

F1 Score is the weighted average of Precision and Recall[13][20]. Formulated as follows **Equation 7**:

$$F1\ Score = 2\frac{Recall\ X\ Precision}{Recall + Precision} \quad (7)$$

In the test scenario, the data was divided into train data and test data with the ratio 80:20 and 70:30. After the data was divided according to the determined proportions, the data were weighted or transformed into vectors using the TD-IDF method. After that, the data was trained with existing models, namely SVM, KNN, Naive Bayes and combined with the Voting Classifier. Next, the metric accuracy, recall accuracy, and F1 score as scores for the built model's performance evaluation. Below were the performance results of the algorithm.

Table 3. Dataset Ratio 80:20

Algorithm Models	Accuracy	Recall	Precision	F1 Score
SVM	91.18 %	94.12 %	88.89 %	91.43 %
KNN	85.29 %	80.39 %	89.13 %	84.54%
Naive Bayes	89.22 %	84.31 %	93.38 %	88.66 %
Voting Classifier	89.22 %	86.27 %	91.67 %	88.89 %

In **Table 3** the dataset is divided by a ratio of 80:20, the highest accuracy was by SVM with 91.18%. Naive Bayes and Voting Classifier obtained the same accuracy of 89.22% but Voting classfier getting F1 Score slightly higher than Naive Bayes and KNN obtained the smallest accuracy with 85.29%.

Table 4. Dataset Ratio 70:30

Algorithm Models	Accuracy	Recall	Precision	F1 Score
SVM	94,77 %	96,05 %	93.59 %	94.81 %
KNN	87,58 %	81.58 %	92.54 %	86.71%
Naive Bayes	90,2 %	84.21 %	95.52 %	89.51 %
Voting Classifier	92,16 %	89.47 %	94.67 %	91.89 %

In **Table 4**, the dataset ratio was 70:30, the pseudo-model increased accuracy and the highest accuracy was SVM with 94.77% and also the F1 score was almost the same, 94.81%. The Voting Classifier model obtained a higher accuracy than the previous Naive Bayes in the first test with 92.16% accuracy while Naive Bayes's acuracy was 90.2%. Although KNN's performance was increase but the accuracy was lower than the previous three models, with only 87.58 % but its F1 score was almost the same as the accuracy value, 86.71%.

The next performance test would be seen from the accuracy score of the training process with training and testing data to evaluate the underviting or overviting occurrence in the designed model. The performance test can be seen in **Table 5** and **Table 6** as follows:

Table 5. Score Accuracy with Ratio 80:20

Algorithm Models	Training set score	Testing set score
SVM	0.9828	0.9118
KNN	0.8676	0.8529
Naive Bayes	0.9142	0.8922
Voting Classifier	0.9338	0.8922

Table 6. Score Accuracy with Ratio 70:30

Algorithm Models	Training set score	Testing set score
SVM	0.9828	0.9118
KNN	0.8676	0.8529

Algorithm Models	Training set score	Testing set score
Naive Bayes	0.9142	0.8922
Voting Classifier	0.9338	0.8922

In the tests listed in **Table 5** with a ratio of 80:20 SVM had a high training score, 98.28% and low testing score, 91.11%. The gap was 7.17% whereas the gap of other models was approximately 1-4%. KNN obtained the smallest value of the three models. The results of the next test shown in **Table 6** with a ratio of 70:30. Some models' performance increased but experienced underfitting where the training accuracy was smaller than testing. Compared to SVM, the overviting was higher in the second test than the first test. The overfitting decreased but of all the tested models, the Voting Classifier had almost the same training and testing scores with less than 1% difference.

Visualization of the performance testing results between training and testing with a dataset ratio of 80:20 and 70:30 can be seen in **Figure 1** and **Figure 2** below.

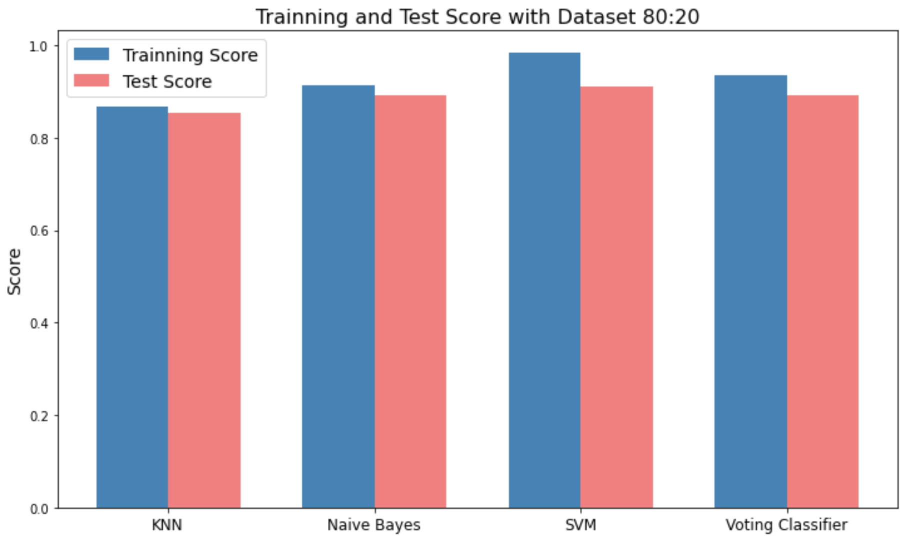

Figure 1 Visualization of Score Accuracy Ratio 80:20

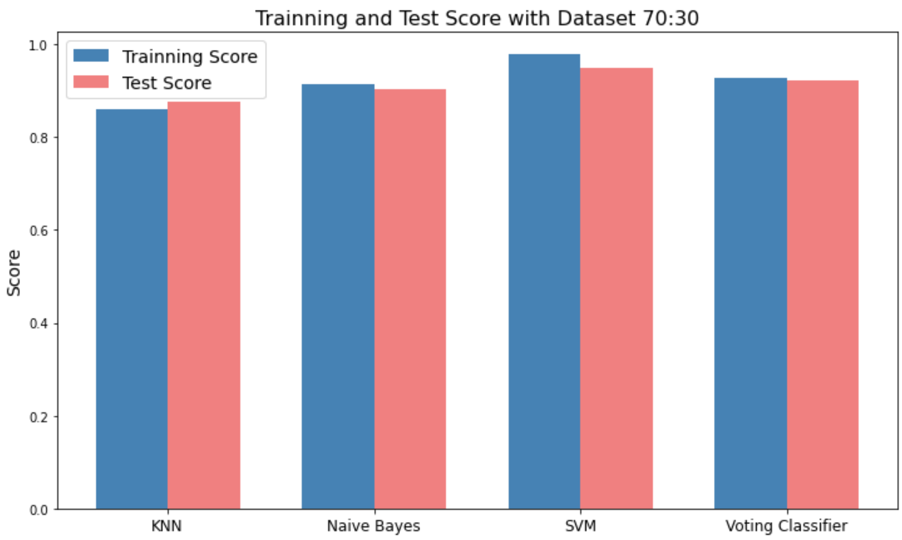

Figure 2 Visualization of Score Accuracy Ratio 70:30

As can be seen in **Figure 1** and **Figure 2**, the performance of the Voting Classifier indicated a high and stable performance as shown in the training scores and test scores at ratios of 80:20 and 70:30 which were almost the same. SVM on the other hand achieved the highest performance of all algorithms but unstable whereas Naive Bayes, whose performance scores were still below the Voting Classifier.

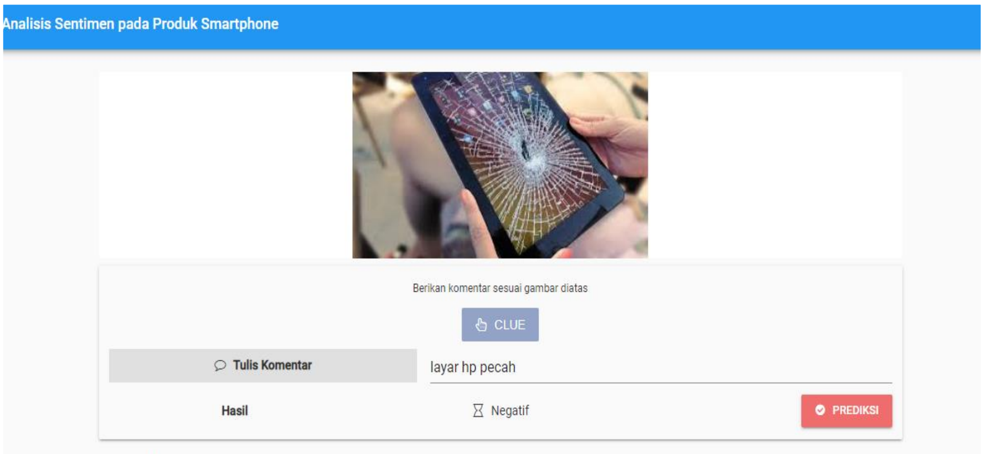

Figure 3. Result of Voting Classifier

The implementation of the designed system can be seen in **Figure 3** generated from the Voting Classifier model. In **Figure 4** the implementation of the program was successful in filtering comments and collecting them into positive and negative categories and calculating all comments based on positive and negative comments.

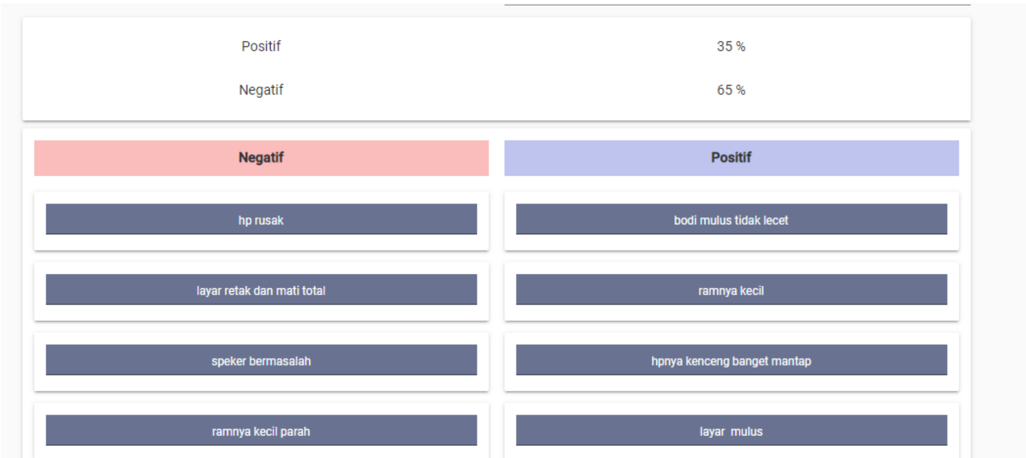

Figure 4. Comment Filtering Results

Conclusion

This research designed a KNN model with a k value of 20 and for SVM using a linear kernel to classify negative and positive comments. It can be concluded that the accuracy of SVM had a higher value compared to KNN, Naive Bayes and Voting Classifier in the use of datasets with a ratio of 80:20 and 70:30. Its highest accuracy value reached 94.77% in a ratio of 70:30. And for each model indicated an accuracy increase in the 70:30 ratio, where KNN got an accuracy of 87.58%, Naive Bayes got 90.2% and Voting Classifier had a big enough increase with 92.16%. In the test with a ratio of 80 :20 voting classifier accuracy had the same value as Naive Bayes with an accuracy of 89.22%. Although SVM showed high accuracy, SVM had a large overfitting with the gap difference between the training and testing accuracy scores was 7.17%. Compared to other models, the gap difference was below 4% in the ratio 80:20. At the ratio of 70:30, Voting Classifier show the accuracy difference between training and testing scores was below 1%. This indicates that the Voting Classifier`s performance is better in studying data and helps in improving the algorithm model for classifying negative and positive comments on smartphone products compared to SVM, KNN and Naive Bayes. The system is also successful in examining customer satisfaction.

Acknowledgement

The researcher would like to thank Amikom University Purwokerto for the support given to researchers through the MSIB Kemendikbudristek 2022 Program, both materials and non-materials. So that this research can be completed on time.

References

[1] A. Setyawan, I. W. Suparta, and N. Aida, “Globalisasi Ekonomi Dan Pengangguran: Studi Kasus

Indonesia," *J. Ekon. Bisnis dan Kewirausahaan*, vol. 10, no. 3, p. 263, 2021, doi: 10.26418/jebik.v10i3.49278.

[2] N. S. B. Maria and T. Widayati, "Dampak Perkembangan Ekonomi Digital terhadap Perilaku Pengguna Media Sosial dalam Melakukan Transaksi Ekonomi," *J. Konsep Bisnis dan Manaj.*, vol. 6, no. 2, pp. 234–239, 2020, doi: 10.31289/jkbm.v6i2.3801.

[3] R. P. D. Yulida Army Nurcahya, "ANALISIS PENGARUH PERKEMBANGAN FINTECH DAN E-COMMERCE TERHADAP PEREKONOMIAN MASYARAKAT," vol. 2019, no. 2157, pp. 1–57, 2019.

[4] L. B. Ilmawan and M. A. Mude, "Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 154–161, 2020, doi: 10.33096/ilkom.v12i2.597.154-161.

[5] F. Sodik and I. Kharisudin, "Analisis Sentimen dengan SVM , NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter," *Prisma*, vol. 4, pp. 628–634, 2021.

[6] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *Edutic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.

[7] A. Andreyestha and A. Subekti, "Analisa Sentiment Pada Ulasan Film Dengan Optimasi Ensemble Learning," *J. Inform.*, vol. 7, no. 1, pp. 15–23, 2020, doi: 10.31311/ji.v7i1.6171.

[8] R. Kohavi, "Scaling Up the Accuracy of Naive-Bayes Classifiers: A Decision-Tree Hybrid," *Proc. Second Int. Conf. Knowl. Discov. Data Min.*, vol. 7, no. Utgo 1988, pp. 202–207, 1996, [Online]. Available: http://www.aaai.org/Library/KDD/1996/kdd96-033.php.

[9] P. Domingos and M. Pazzani, "On the Optimality of the Simple Bayesian Classifier underZero-One Loss. Machine Learning," *Mach. Learn.*, vol. 29, pp. 103–130, 1997, [Online]. Available: https://link.springer.com/article/10.1023/A:1007413511361.

[10] M. Rezwanul, A. Ali, and A. Rahman, "Sentiment Analysis on Twitter Data using KNN and SVM," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 6, pp. 19–25, 2017, doi: 10.14569/ijacsa.2017.080603.

[11] A. Saifudin and R. S. Wahono, "Penerapan Teknik Ensemble untuk Menangani Ketidakseimbangan Kelas pada Prediksi Cacat Software," *J. Softw. Eng.*, vol. 1, no. 1, 2015.

[12] I. M. A. Agastya, "Pengaruh Stemmer Bahasa Indonesia Terhadap Peforma Analisis Sentimen Terjemahan Ulasan Film," *J. Tekno Kompak*, vol. 12, no. 1, p. 18, 2018, doi: 10.33365/jtk.v12i1.70.

[13] A. N. Assidyk, E. B. Setiawan, and I. Kurniawan, "Analisis Perbandingan Pembobotan TF-IDF dan TF-RF pada Trending Topic di Twitter dengan Menggunakan Klasifikasi K-Nearest Neighbor," *e-Proceeding Eng.*, vol. 7, no. 2, pp. 7773–7781, 2020.

[14] A. Kanitkar, "Bollywood Movie Success Prediction using Machine Learning Algorithms," *2018 3rd Int. Conf. Circuits, Control. Commun. Comput.*, pp. 1–4, 2018, doi: 10.1109/CIMCA.2018.8739693.

[15] B. Herwijayanti, D. E. Ratnawati, and L. Muflikhah, "Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity," *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 1, pp. 306–312, 2018, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/796.

[16] R. S. Perdana and M. A. Fauzi, "Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter dengan Metode Support Vector Machine dan Lexicon Based Features Twitter event detection View project Human Detection and Tracking View project," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1(12), no. October, pp. 1725–1732, 2017, [Online]. Available: https://www.researchgate.net/publication/320234928.

[17] R. Puspita and A. Widodo, "Analisis Sentimen terhadap Layanan Indihome di Twitter dengan Metode Machine Learning," vol. 6, no. 4, pp. 759–766, 2021, [Online]. Available: http://openjournal.unpam.ac.id/index.php/informatika759.

[18] Mochammad Haldi Widianto, "Algoritma Naive Bayes | BINUS UNIVERSITY BANDUNG - Kampus Teknologi Kreatif," 2019. https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes/ (accessed Aug. 09, 2022).

[19] Gandhi and M. Pandey, “Hybrid Ensemble of classifiers using voting,” *2015 Int. Conf. Green Comput. Internet Things*, pp. 399–404, 2015, doi: 10.1109/ICGCIoT.2015.7380496.

[20] J. Han, M. Kamber, and J. Pei, “Data mining concepts and techniques,” *MK*, 2011.