

# Sentiment Analysis of Shopee Food Application User Satisfaction Using the C4.5 Decision Tree Method

Fersellia<sup>1)\*</sup>, Ema Utami<sup>2)</sup>, Ainul Yaqin<sup>3)</sup>

<sup>1,2,3)</sup>Universitas AMIKOM Yogyakarta, Indonesia

<sup>1)</sup>[fersellia.18@students.amikom.ac.id](mailto:fersellia.18@students.amikom.ac.id), <sup>2)</sup>[ema.u@amikom.ac.id](mailto:ema.u@amikom.ac.id), <sup>3)</sup>[email@amikom.ac.id](mailto:email@amikom.ac.id)

**Submitted** : Jun 9, 2023 | **Accepted** : Jun 19, 2023 | **Published** : Jul 1, 2023

**Abstract:** Sentiment analysis on public opinion regarding the Shopee food application is an interesting topic in the context of evaluating service quality in the shopee food application. In this digital era, user opinion has a very important role in shaping public perception of the application. Therefore, sentiment analysis is needed to understand user opinion about the Shopee food application. This study uses Decision Tree C4.5 to analyze public sentiment on the use of the Shopee Food application on Twitter users. However, beforehand it is necessary to overcome the problem of data imbalance which is common in datasets, where the number of positive, negative, and neutral sentiments is not balanced. To overcome this problem, three different techniques are used, namely SMOTE, undersampling, and a combination of oversampling and undersampling. The results of this study indicate that the SMOTE technique provides better results in overcoming data imbalances and increasing prediction accuracy. With an accuracy of 0.88. the SMOTE technique can provide more accurate sentiment predictions than the undersampling technique and the combination of oversampling and undersampling. This is because SMOTE can synthetically expand the number of minority samples, thereby preventing the loss of information and maintaining variation in the dataset. In conclusion, sentiment analysis on the Shopee Food application on Google Play using the Decision Tree C4.5 algorithm and the SMOTE technique can overcome data imbalances with a prediction accuracy of 0.88. This technique is more efficient than the undersampling technique and the combination of oversampling and undersampling. These results can provide developers with valuable insights to improve app quality and user satisfaction.

**Keywords:** Shopee Food, Sentiment Analysis, Decision Tree C4.5, Smote, Undersampling, Combine Oversampling & Undersampling.

## INTRODUCTION

Shopee is an e-commerce platform that supports the development of the digital economy and holds the largest market share in Indonesia. One of the services offered by Shopee is Shopee Food, a food delivery service. Shopee Food was first launched in April 2020, initially offering frozen food, snacks, cakes, and various ready-to-eat meals (Mufidah et al. 2022). By using Shopee Food, users can easily enjoy a variety of cuisines from their favorite restaurants. During the COVID-19 pandemic in

\*name of corresponding author



the second quarter of Ramadan, Shopee Food experienced a 25-fold increase in visits compared to regular days (Pratama 2020). Moreover, Shopee Food offers various promotions, discounts, vouchers, and special offers, which further attract users to utilize the application (Vania and Simbolon 2021).

However, every user has a varied experience when using the Shopee Food application, particularly in terms of service quality, delivery speed, menu variety, and satisfaction with the ordered food. While some users have positive experiences and enjoy the convenience of ordering their favorite meals with a wide selection of restaurants, many others encounter negative experiences such as delivery delays, order mistakes, or unsatisfactory food quality. If these issues are not addressed, it may lead to a decline in public interest in using the application.

To analyze public opinions effectively, sentiment analysis is required as a text mining technique capable of automatically understanding, processing, and extracting data by computers. Sentiment analysis is often associated with social media data analysis to obtain both positive and negative sentiments. (Sari et al. 2020). This process falls within the domain of Natural Language Processing (NLP), which is responsible for recognizing and extracting language meaning from textual data (Nurzahputra and Muslim 2016). In this study, a text classification method is needed to facilitate sentiment analysis and achieve accurate and optimal results. The decision tree C45 algorithm will be utilized, as previous research on sentiment analysis (Bibi 2019) has demonstrated high accuracy levels

Therefore, the problem to be addressed is how to perform sentiment analysis on public opinions regarding the usage of the Shopee Food application, with a specific focus on service quality, delivery speed, menu variety, and user satisfaction.

## LITERATURE REVIEW

Previous research conducted by (Bibi 2019) by proposing a two-step method, namely by removing hashtags and stop words in the preprocessing process and the second is using the feature vector designed. Where feature vectors are formulated by identifying the number of positive words, negative words, and the presence of negations and the use of POS tags, and the C45 decision tree algorithm is used as a classification algorithm. The experimental results of the proposed method show significant success in terms of accuracy, namely obtaining an accuracy of 90%.

To improve service quality, Bank BCA also conducts a sentiment analysis consisting of 4 attributes (time, accuracy, focus, and satisfaction) with each attribute having a weight value of 1 to 5. However, this research is still being done manually. Therefore (Lia Hananto, Sofiah Hilabi, and Noviani 2022) conducted research using the C.45 algorithm by carrying out 3 tests to produce an accuracy of 88.75% with an AUC value of 0.744 and testing on applications that were made to produce 0.722.

Besides that (Pattihha and Hendry 2022) also conducted research using the Naïve Bayes, K-NN, and Decision Tree methods to make comparisons by looking at the level of accuracy of the three methods used. The results of this study show that the naïve Bayes method has an accuracy of 84.08% with class precision for pred. Positive is 83.65%, pred. Neutral is 97.06%, negative is 100%, K-NN method is 83.38% with class precision for pred. positive is 83.05%, neutral pred is 96.43, negative pred is 0.0% and the decision tree method gets an accuracy of 81.09 with class precision for pred. positive is 81.09%, pred. neutral is 0.0% and pred. negative 0.0%. With the conclusion that Naïve Bayes has the highest accuracy with 84.08%.

## METHOD

In every research, there is always a method or model used to explain the research flow used. Fig. 1 describes the stages that will be carried out in this study starting from dataset collection, preprocessing, word weighting, and the data balancing process using three methods, namely SMOTE, Oversampling, and undersampling, Random Undersampling, then classification will be carried out using a Decision Tree. And finally, testing or evaluation is carried out.

\*name of corresponding author



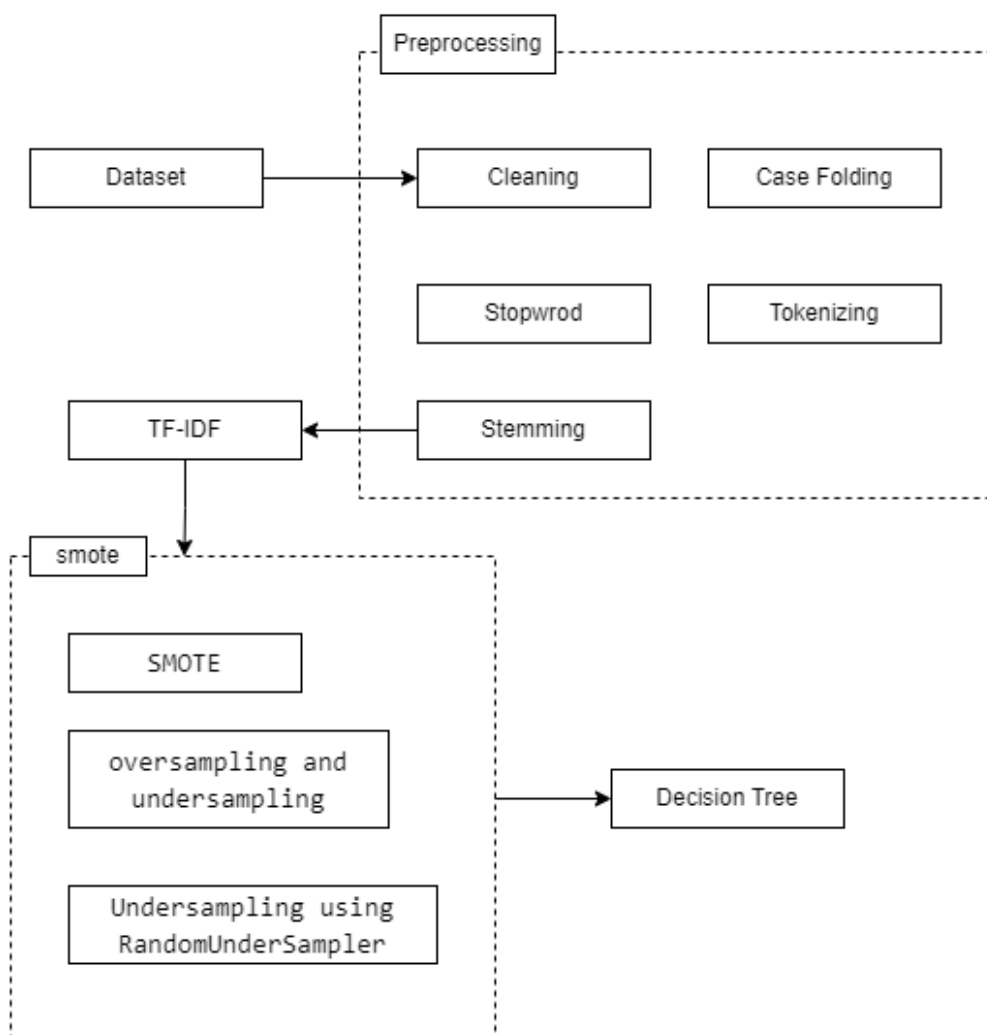


Fig 1. Research Flow

**Dataset**

This data set comes from public opinion on the Shopee Food application. This dataset was taken from Twitter over a span of one month, from May 2023 to June 2023 with a total of 1005 tweets using the keyword shop food. Table 1 is an example of the dataset that will be used.

Table 1. Dataset

No	Tweet
1	need spay ovo 363k need bgt buat shopeefood dm aja yg mau tukeran apa

\*name of corresponding author



2	Resto Middle East Food by Kebuli-Q\nPusat (08.00 - 21.30)\n ♀ Jl.Jatiluhur Raya, Baranangsiang Indah, C.8 No.1, Kel.Katulampa, Bogor\n <a href="https://t.co/4pq2Ygi31e">https://t.co/4pq2Ygi31e</a> \n (GoFood/GrabFood/ShopeeFood)\n 📞 Telp/WA: 0818106881 / 08196000030\n Ig : @kebuli_q / @KebuliQ\n Fb : Kebuli-q\n <a href="https://t.co/ObAkhwvznb">https://t.co/ObAkhwvznb</a>
3	<a href="https://t.co/kk0VUw9dnX">https://t.co/kk0VUw9dnX</a> link grabfood dagangan ayah.. Ayukk yg laper sini ada sate padang soto padang yuk order cobain yukk 🙌🙌\n Di gofood dan shopeefood juga ada #jualan #jajanan #makanan #kuliner #satepadang #sotopadang #grabfood #gofood #depok #sedap #online #anak #shopeefood.
4	Lahiya baru sadar hari ini tanggal kembar pantesan tadi promo shopeefood 30rebay wkwkkkk
5	KOK FITUR SHOPEEFOOD GUE ILANG?!! GUE LAPER NIH BOS
6	@euphxriax_ @euphxriax_ kak aku need spay buat shopeefood, tukeran aja aku ada saldo 363k kalo mau tuker semua, kalo engga berapa aja tukeran yuk kak need urgent
7	Lisensi Franchise Ayam Geybok Bang Jarwo #Promo #GoFood #GrabFood #ShopeeFood #AirAsiaFood #AyamGeybokBangJarwo #GeybokBangJarwo #Lunch #MakanSiang #AyamGeprek #Franchise #Ayam #KulinerNusantara #Kuliner #Review #Meal #Lunch #Jakarta\n <a href="https://t.co/evxa9nOsCr">https://t.co/evxa9nOsCr</a>
8	percuma ajg masak seharian tau tau tengah malam malah shopeefood 🙌🙌 kalau kaya gini unsur hematnya dimana dahhh
9	/pamer sekalian curhat. ternyata diskon shopeefood lumayan bagus jg ya menurut sender, tp nunggunya lama banget sejam lebih wkwk sender udah ga laper lg, malah ngantuk 🙌 <a href="https://t.co/L1DZdRYWP7">https://t.co/L1DZdRYWP7</a>
10	Wah abang shopeefood nya tadi ganteng \n 😊 😊 🙌

### Preprocessing

Preprocessing is an important task and step in text mining, natural language processing (NLP), and "IR" information retrieval. Text mining uses data pre-processing to extract interesting and important information from unstructured text data. Information Search (IR) is used to decide which documents in a collection should be retrieved to satisfy a user's need for information (Gurusamy and Kannan S 2014). Therefore, search decisions are made by comparing the search terms to the index terms (important words or phrases) that appear in the document itself. In preprocessing there are several steps that must be carried out such as cleaning, case folding, tokenizing, stopwords, and stemming.

The dataset that has been collected does not rule out the possibility that it will contain a lot of irrelevant or even missing data. To overcome this problem, a cleaning process is needed. At this stage, data cleaning is carried out where this stage aims to avoid data duplication, inconsistent data, correcting data errors, or adding data to support the system created. (Budiman and Ramadina 2015).

After cleaning the data, the process of generalizing the use of capital letters will be carried out. For example, if there is the letter "CasE Folding", then with case folding we will change all letters to lowercase.

Next, the tokenizing process is carried out, tokenizing is a process for dividing or breaking text or sentences into tokens or certain parts such as data sets (Rofiqi et al. 2019).

\*name of corresponding author



The next stage is the stopword, where the stopword is generally used in the information retrieval task. In the world of programming, especially in the field of data classification, stopwords are needed because they are very useful for reducing the number of words or text that must be processed.

The final stage of data preprocessing is stemming, where stemming is used to return words that have affixes to become basic words such as “belajar” become “ajar” and “membaca” become “baca” in the stemming process is very necessary because it can increase accuracy (Mufidah et al. 2022)(Agastya 2018).

### TF-IDF

TF-IDF is a way to weight the relationship of a word (term) to a document. This method combines two weighting concepts, namely the frequency of occurrence of a word in certain documents and the frequency of documents containing inverted words. The number of occurrences of a word in a particular document indicates how important the word is in that document. The frequency of documents containing a word indicates how often the word occurs. Thus, the weight of the word-document relationship is high when the frequency of words in the document is high and the frequency of all documents containing the word in the document is low. TF-IDF formula:

$$tf = 0,5 + 0,5 x \frac{tf}{\max (tf)}$$

$$idf_t = \log \left( \frac{D}{df_t} \right)$$

$$W_{d,t} = tf_{d,t} x IDF_{d,t}$$

Information:

$tf$  = number of words to search for in a document.

$\max(tf)$  = the highest number of occurrences of a term in the same document.

Mark  $D$  = document totals

$df_t$  = number of documents containing the term  $t$ .

$IDF$  = *inversed Document Frequency* ( $\log_2 (D/df)$ )

$d$  =  $d$ -th document

$t$  = the  $t$ -word of the keyword

$W$  = the weight of the  $d$ -th document to the  $t$ -word

### SMOTE

SMOTE (Synthetic Minority Over-sampling Technique) is an approach to balancing sample data on classes with excessive imbalance (majority) with a focus on minority classes to increase the efficiency of the classification method. In SMOTE there is a possibility of matching, i.e. information from the minority class, being copied (Bias 2021), (Bisri and Rachmatika 2019).

\*name of corresponding author



## Decision Tree

Decision Tree is a popular and effective prediction and classification method. Decision Trees can transform facts to form decision trees that can represent rules. The regulations that apply are easy to understand. The processes carried out in the decision tree method are converting data into a decision tree form, changing the tree form into a roll form and simplifying roles (Prasetyo and Pahlevi 2019). In the decision tree method, internal nodes, root nodes, and terminal nodes are part of the tree. For some time, variables or properties were root nodes and internal nodes, and class names were leaf nodes. The classifier keeps track of root nodes and internal nodes with query data to end nodes. The class identifier information in this query is based on the existing identifier in the internal node (Tri Romadloni, Santoso, and Budilaksono 2019).

## Classification Evaluation

Classification evaluation is based on testing the correct object and the wrong object. This validation determines the best type of model from the classification results (Siringoringo 2018). Evaluation in this study uses the confusion matrix. The confusion matrix is information about the actual classification results that can be predicted by the classification system. From the results of the confusion matrix, the value of accuracy, precision, and recall will be determined.

Accuracy is the determination of the system in carrying out the classification process correctly.

### *Akurasi*

$$= \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

Accuracy is the ratio of the number of relevant documents from the classification system to the total number of documents found.

### *Presisi*

$$= \frac{TP}{TP + FP} \quad (3)$$

Recall is the ratio of the number of documents returned by the classification system to the total number of relevant documents.

### *Recall*

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

## RESULT

In this sentiment analysis process, the dataset was labeled using the TextBlob library. It was found that reviews about the Shopee Food application tended to be more neutral than positive and negative. For more detailed information, please refer to Fig 2.

\*name of corresponding author



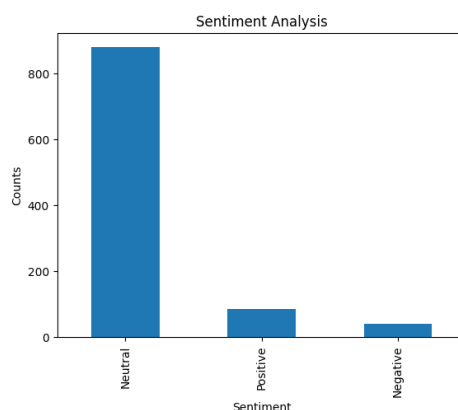


Fig. 2 Sentiment Analysis Visualization

Based on Figure 2, it can be observed that the number of neutral reviews is more dominant compared to positive and negative reviews. There are 879 neutral reviews, 86 positive reviews, and 41 negative reviews. However, the data is still imbalanced. To address this issue, the implementation of techniques such as SMOTE, oversampling, and undersampling is necessary, particularly undersampling, to achieve the desired balance in the dataset.

After balancing the data, classification is performed using the decision tree C45 model. Decision tree C45 is an algorithm that works by constructing a decision tree based on the features present in the dataset. In this stage, the dataset is divided into two parts, with 80% for training data and 20% for testing data. In the classification stage, the decision tree C45 utilizes learning rules based on information gain and split criterion. The decision tree is built by splitting the dataset based on attributes that have the highest information gain, creating an optimal separation between different classes. After constructing the decision tree, classification is performed by following the branches in the decision tree. Once the classification process is complete, performance evaluation is conducted using several metrics such as accuracy, precision, recall, and F1-score. The evaluation results are shown in the following table, providing an overview of how well the decision tree C45 model performs in classifying sentiments in the Shopee Food application.

Table 2. Performance Evaluation

Teknik	Accuracy	Recall	Precision	F1-Score
SMOTE	0.88	0.86	0.88	0.86
Undersampling	0.40	0.4	0.80	0.51
Combine oversampling dan undersampling	0.87	0.87	0.83	0.85

From the table above, it can be concluded that the SMOTE technique outperforms other imbalance techniques. The results show that SMOTE achieves an accuracy of 0.88, while the combination of oversampling and undersampling achieves an accuracy of 0.87, with a very small difference of 0.01. On the other hand, the undersampling technique yields the lowest accuracy with a value of 0.4.

## DISCUSSIONS

In this study, the SMOTE technique is better than the three proposed techniques because using the SMOTE technique produces new synthetic samples using an interpolation technique between existing minority samples. Adding this composite sample expands the minority sample size, giving the model

\*name of corresponding author



more information to examine and make better decisions. While undersampling is lower because undersampling eliminates a large number of majority samples to achieve class balance. This can result in the loss of important information from an omitted majority sample, thereby weakening the model's ability to learn significant patterns from the data. While it is lower than the SMOTE technique because the combination of oversampling and undersampling is eliminating a large number of majority samples and some minority samples and repeating the remaining minority samples. This can result in the loss of important information from the data set, both from the majority and minority classes. This loss of information can affect the model's ability to learn meaningful patterns from the data

### CONCLUSION

For the three techniques proposed for imbalance data, the SMOTE technique obtains better accuracy than the others where the SMOTE technique obtains an accuracy of 0.88 from the combined oversampling and undersampling which is 0.01 lower than the SMOTE technique, which obtains an accuracy of 0.87. while undersampling has the lowest accuracy, namely 0.4.

### REFERENCES

- Agastya, I. Made Artha. 2018. "Pengaruh Stemmer Bahasa Indonesia Terhadap Peforma Analisis Sentimen Terjemahan Ulasan Film." *Jurnal Tekno Kompak* 12(1):18. doi: 10.33365/jtk.v12i1.70.
- Bias, Introduced Sentiment. 2021. "JURNAL RESTI Sentiment Classification for Film Reviews by Reducing Additional." (10):4–10.
- Bibi, Raheela. 2019. "Sentiment A Nalysis for Urdu N Ews Tweets U Sing Decision T Ree." *2019 IEEE 17th International Conference on Software Engineering Research, Management and Applications (SERA)* 66–70.
- Bisri, Achmad, and Rinna Rachmatika. 2019. "Integrasi Gradient Boosted Trees Dengan SMOTE Dan Bagging Untuk Deteksi Kelulusan Mahasiswa." *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTETI)* 8(4):309. doi: 10.22146/jnteti.v8i4.529.
- Budiman, Irwan, and Retma Ramadina. 2015. "Penerapan Fungsi Data Mining Klasifikasi Untuk Prediksi Masa Studi Mahasiswa Tepat Waktu Pada Sistem Informasi Akademik Perguruan Tinggi." *Ijccs* x, No.x(1):1–5.
- Gurusamy, Vairaprakash, and Kannan S. 2014. "Preprocessing Techniques for Text Mining." *International Journal of Computer Science & Communication Networks* 5(1):7–16.
- Lia Hananto, April, Shofa Sofiah Hilabi, and Detrie Noviani. 2022. "Design of Customer Satisfaction Application at BCA Kcp Rengasdengklok Using C.45 Algorithm Method." *Buana Information Technology and Computer Sciences (BIT and CS)* 3(1):11–16. doi: 10.36805/bit-cs.v3i1.2048.
- Mufidah, Farah Syadza, Sri Winarno, Farrikh Alzami, Erika Devi Udayanti, and Ramadhan Rakhmat Sani. 2022. "Analisis Sentimen Masyarakat Terhadap Layanan Shopeefood Melalui Media Sosial Twitter Dengan Algoritma Naïve Bayes Classifier." *JOINS (Journal of Information System)* 7(1):14–25. doi: 10.33633/joins.v7i1.5883.
- Nurzahputra, Aldi, and Much Aziz Muslim. 2016. "Analisis Sentimen Pada Opini Mahasiswa Menggunakan Natural Language Processing." *Seminar Nasional Ilmu Komputer (Snik)*:114–18.
- Pattiiha, Franly Salmon, and Hendry Hendry. 2022. "Perbandingan Metode K-NN, Naïve Bayes, Decision Tree Untuk Analisis Sentimen Tweet Twitter Terkait Opini Terhadap PT PAL Indonesia." *JURIKOM (Jurnal Riset Komputer)* 9(2):506. doi: 10.30865/jurikom.v9i2.4016.

\*name of corresponding author



- Prasetyo, Karno Ganjar, and Said Mrza Pahlevi. 2019. "Analisis Perbandingan Algoritma Decision Tree Dengan Support Vector Machine Untuk Mendeteksi Kompetensi Mahasiswa Konsentrasi Informatika Komputer Studi Kasus : Politeknik Lp3I Jakarta, Kampus Depok." *Lentera ICT* 5(5):11–26.
- Pratama, Gama. 2020. "130-64-555-1-10-20200901." 1:21–34.
- Rofiqi, Moh Afif, Abd. Charis Fauzan, Afivatu Pratama Agustin, and Ahmad Agung Saputra. 2019. "Implementasi Term-Frequency Inverse Document Frequency (TF-IDF) Untuk Mencari Relevansi Dokumen Berdasarkan Query." *ILKOMNIKA: Journal of Computer Science and Applied Informatics* 1(2):58–64. doi: 10.28926/ilkomnika.v1i2.18.
- Sari, Dita Novita, Dita Novita Sari, Firda Adelia, Fita Rosdiana, Belsana Butar Butar, and Muhadi Hariyanto. 2020. "Analisa Sentimen Terhadap Review Produk Kecantikan Menggunakan Metode Naive Bayes Classifier." *JIKA (Jurnal Informatika)* 4(3):109. doi: 10.31000/jika.v4i3.3086.
- Siringoringo, Rimbun. 2018. "Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE Dan K-Nearest Neighbor." *Jurnal ISD* 3(1):44–49.
- Tri Romadloni, Nova, Imam Santoso, and Sularso Budilaksono. 2019. "Perbandingan Metode Naive Bayes, Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl Commuter Line." *Jurnal IKRA-ITH Informatika* 3(2):1–9.
- Vania, Izella, and Remista Simbolon. 2021. "Pengaruh Promo ShopeeFood Terhadap Minat Beli Pengguna Shopee (Di Daerah Tangerang Selatan)." *Fakultas Ekonomi Universitas Advent Indonesia* 46–58.
- Watrianthos, Ronal, Sudi Suryadi, Deci Irmayani, Marnis Nasution, and Elida F. S. Simanjorang. 2019. "Sentiment Analysis of Traveloka App Using Naïve Bayes Classifier Method." *International Journal of Scientific and Technology Research* 8(7):786–88.
- Agastya, I. Made Artha. 2018. "Pengaruh Stemmer Bahasa Indonesia Terhadap Peforma Analisis Sentimen Terjemahan Ulasan Film." *Jurnal Tekno Kompak* 12(1):18. doi: 10.33365/jtk.v12i1.70.
- Bias, Introduced Sentiment. 2021. "JURNAL RESTI Sentiment Classification for Film Reviews by Reducing Additional." (10):4–10.
- Bibi, Raheela. 2019. "Sentiment A Nalysis for Urdu N Ews Tweets U Sing Decision T Rec." *2019 IEEE 17th International Conference on Software Engineering Research, Management and Applications (SERA)* 66–70.
- Bisri, Achmad, and Rinna Rachmatika. 2019. "Integrasi Gradient Boosted Trees Dengan SMOTE Dan Bagging Untuk Deteksi Kelulusan Mahasiswa." *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi (JNTEI)* 8(4):309. doi: 10.22146/jnteti.v8i4.529.
- Budiman, Irwan, and Retma Ramadina. 2015. "Penerapan Fungsi Data Mining Klasifikasi Untuk Prediksi Masa Studi Mahasiswa Tepat Waktu Pada Sistem Informasi Akademik Perguruan Tinggi." *Ijccs* x, No.x(1):1–5.
- Gurusamy, Vairaprakash, and Kannan S. 2014. "Preprocessing Techniques for Text Mining." *International Journal of Computer Science & Communication Networks* 5(1):7–16.
- Lia Hananto, April, Shofa Sofiah Hilabi, and Detrie Noviani. 2022. "Design of Customer Satisfaction Application at BCA Kcp Rengasdengklok Using C.45 Algorithm Method." *Buana Information Technology and Computer Sciences (BIT and CS)* 3(1):11–16. doi: 10.36805/bit-cs.v3i1.2048.

\*name of corresponding author



- Mufidah, Farah Syadza, Sri Winarno, Farrikh Alzami, Erika Devi Udayanti, and Ramadhan Rakhmat Sani. 2022. "Analisis Sentimen Masyarakat Terhadap Layanan ShopeeFood Melalui Media Sosial Twitter Dengan Algoritma Naïve Bayes Classifier." *JOINS (Journal of Information System)* 7(1):14–25. doi: 10.33633/joins.v7i1.5883.
- Nurzahputra, Aldi, and Much Aziz Muslim. 2016. "Analisis Sentimen Pada Opini Mahasiswa Menggunakan Natural Language Processing." *Seminar Nasional Ilmu Komputer (SNIK)*:114–18.
- Pattiiha, Franly Salmon, and Hendry Hendry. 2022. "Perbandingan Metode K-NN, Naïve Bayes, Decision Tree Untuk Analisis Sentimen Tweet Twitter Terkait Opini Terhadap PT PAL Indonesia." *JURIKOM (Jurnal Riset Komputer)* 9(2):506. doi: 10.30865/jurikom.v9i2.4016.
- Prasetyo, Karno Ganjar, and Said Mrza Pahlevi. 2019. "Analisis Perbandingan Algoritma Decision Tree Dengan Support Vector Machine Untuk Mendeteksi Kompetensi Mahasiswa Konsentrasi Informatika Komputer Studi Kasus : Politeknik Lp3I Jakarta, Kampus Depok." *Lentera ICT* 5(5):11–26.
- Pratama, Gama. 2020. "130-64-555-1-10-20200901." 1:21–34.
- Rofiqi, Moh Afif, Abd. Charis Fauzan, Afivatu Pratama Agustin, and Ahmad Agung Saputra. 2019. "Implementasi Term-Frequency Inverse Document Frequency (TF-IDF) Untuk Mencari Relevansi Dokumen Berdasarkan Query." *ILKOMNIKA: Journal of Computer Science and Applied Informatics* 1(2):58–64. doi: 10.28926/ilkomnika.v1i2.18.
- Sari, Dita Novita, Dita Novita Sari, Firda Adelia, Fita Rosdiana, Belsana Butar Butar, and Muhadi Hariyanto. 2020. "Analisa Sentimen Terhadap Review Produk Kecantikan Menggunakan Metode Naive Bayes Classifier." *JIKA (Jurnal Informatika)* 4(3):109. doi: 10.31000/jika.v4i3.3086.
- Siringoringo, Rimbun. 2018. "Klasifikasi Data Tidak Seimbang Menggunakan Algoritma SMOTE Dan K-Nearest Neighbor." *Jurnal ISD* 3(1):44–49.
- Tri Romadloni, Nova, Imam Santoso, and Sularso Budilaksono. 2019. "Perbandingan Metode Naive Bayes, Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl Commuter Line." *Jurnal IKRA-ITH Informatika* 3(2):1–9.
- Vania, Izella, and Remista Simbolon. 2021. "Pengaruh Promo ShopeeFood Terhadap Minat Beli Pengguna Shopee (Di Daerah Tangerang Selatan)." *Fakultas Ekonomi Universitas Advent Indonesia* 46–58.
- Watrianthos, Ronal, Sudi Suryadi, Deci Irmayani, Marnis Nasution, and Elida F. S. Simanjourang. 2019. "Sentiment Analysis of Traveloka App Using Naïve Bayes Classifier Method." *International Journal of Scientific and Technology Research* 8(7):786–88.

\*name of corresponding author

