Indonesian Journal of Electrical Engineering and Computer Science
Vol. 39, No. 1, July 2025, pp. 364~373
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v39.i1.pp364-373

A sentiment analysis on skewed product reviews: Ben & Jerry's ice cream

Nabilla Nurulita Dewi[1], Sekar Gesti Amalia Utami[1], Shalsabila Aura Adiar[1], Hasan Dwi Cahyono[2]
[1]Department of Informatics, Faculty of Information Technology and Data Science, Universitas Sebelas Maret, Surakarta, Indonesia
[2]Department of Data Science, Faculty of Information Technology and Data Science, Universitas Sebelas Maret, Surakarta, Indonesia

Article Info

Article history:

Received Jun 24, 2024
Revised Dec 2, 2024
Accepted Mar 25, 2025

Keywords:

Logistic regression
Naive Bayes
Product reviews
Sentiment analysis
Support vector machine

ABSTRACT

Sentiment analysis of product reviews offers valuable insights into consumer perspectives, which can inform product development and marketing strategies. Given the growing importance of user-generated content like product reviews, this study explored sentiment classification in online reviews of Ben & Jerry's ice cream. We designed and evaluated three machine learning algorithms for sentiment classification: Naïve Bayes (NB), logistic regression (LR), and support vector machine (SVM). The dataset exhibited a significant class imbalance, with substantially more positive than negative reviews. We employed two oversampling techniques: the synthetic minority oversampling technique (SMOTE) and the adaptive synthetic sampling approach (ADASYN). With the original skewed data, NB, LR, and SVM achieved accuracies of 91.90%, 93.77%, and 95.09%, respectively. While SMOTE did not improve performance in some scenarios, ADASYN yielded positive results and generally enhanced model reliability across all algorithms. Post-balancing with ADASYN, the sentiment distribution became less skewed, and accuracies shifted to 92.04% for NB, 94.96% for LR, and 95.23% for SVM. The combination of SVM and ADASYN demonstrated promising results, suggesting this approach may offer robust and efficient performance for binary sentiment classification, especially with imbalanced datasets.

Corresponding Author:

Hasan Dwi Cahyono
Department of Data Science, Faculty of Information Technology and Data Science
Universitas Sebelas Maret
Surakarta, Central Java, Indonesia
Email: hasan.dwi.cahyono@gmail.com

1. INTRODUCTION

Customer satisfaction, product reviews, and user preferences are examples of user-generated information rapidly growing and becoming a valuable source of business and marketing intelligence [1]. These customer reviews have become crucial for product manufacturers, service providers, and end-users to understand public opinion and make concrete decisions. Customer reviews on e-commerce platforms are a significant factor in consumer decision-making in the modern digital age [2]. These reviews are not only to provide prospective customers with helpful information but also to enable manufacturers to get insightful criticism that helps them improve the standard of their goods and services. Thus, sentiment analysis is emerging as a crucial tool for deriving conclusions from the content of the reviews due to the growing number of reviews that are accessible online [3]. Every industrial company faces various challenges, especially the large number of competitors in the same sector, such as in the food and beverage (F&B) industry.

The manufacturing sector must continuously improve its service quality. Service quality improvement refers to the ability of a process to produce a product or service according to the specifications desired by the customer [4]. However, sometimes situations occur where the products must meet customer expectations when handed over to them. A fundamental idea in service management and marketing is customer happiness. Numerous theories exist concerning customer satisfaction. Expectation-discrepancy theory, equity theory, attribution theory, dissonance theory, and contrast theory have all been applied to the study of consumer happiness in the past [5]. Before purchasing a product or service, customers have quality expectations regarding the product or service. After making a purchase, they will compare actual perceptions with their expectations. Positive mismatch occurs when the actual perceptions are higher than the expectation. When actual perceptions meet expectations, there is no mismatch of expectations. Conversely, a negative mismatch means that actual perceptions fall below expectations.

Sentiment analysis has increasingly become crucial in determining market trends and consumer feedback by identifying and categorizing thoughts within a text. Past studies have improved sentiment analysis accuracy and efficiency through various categorization methods [6]. Furthermore, a study from Bahtiar *et al.* [7] used sentiment analysis on Google Play Store app reviews to understand users' feelings about commercially sold programs using Naïve Bayes (NB) and logistic regression (LR). The tone of each review was established by comparing it to the application's rating. Two conditions were applied to the dataset: two labels (positive and negative) and three (positive, neutral, and negative). LR classification yielded the best results for the Shopee dataset with two labels, with 84.58% accuracy, 84.66% precision, and 84.63% recall. The study showed that datasets with two labels generally yielded more accurate results than those with three. Based on the previous studies, NB and LR show promising results in sentiment analysis. However, both algorithms suffered performance challenges, especially when the data used for investigation has imbalanced distributions (skewed). Support vector machine (SVM) is a practical machine-learning approach to regression and classification problems. SVM identifies the perfect hyperplane separating all classes from most isolated data points [8]. As a result, SVM performs well in high-dimensional settings like text data with thousands of features and complex decision boundaries. Also, kernel functions can transform the incoming data into higher dimensional space, thereby improving classification performances [9], [10]. SVM has been used in cases where the separability of classes is not linear [11]. Analyzing customer reviews can help improve product offers, increase customer satisfaction, and offer insightful information about consumer preferences. Understanding sentiment in ice cream reviews can also help merchants and producers spot trends, resolve problems, and adjust their marketing tactics to satisfy customers better.

This study aims to improve sentiment analysis performance, especially for imbalanced data. We focus on customer reviews of ice cream products. Our approach involves careful data cleaning and using oversampling techniques based on synthetic minority oversampling technique (SMOTE) and the adaptive synthetic sampling approach (ADASYN) oversampling methodologies to achieve dataset balancing. We then evaluate the performance of NB, LR, and SVM using several metrics. Our goal is to provide practical guidance for choosing the best sentiment analysis method for imbalanced datasets.

2. METHOD

This study was conducted in multiple steps, as Figure 1 illustrates. In this work, we integrated different approaches and phases of the research as outlined in the following steps: to guarantee the dataset's quality and algorithmic compatibility, we first gathered and preprocessed the combination of product names and customer reviews to predict the sentiments of customers. Thus, the consideration of multicollinearity of variables can be relaxed. Lastly, we tested and compared the robustness of each model in analyzing the sentiment of customer reviews by assessing its performance using criteria including accuracy, precision, recall, and F1-score. We used a Windows 10 Pro 64-bit PC with 13th Gen Intel Core i9 3.00 GHz, 128 GHz of RAM, and NVIDIA GeForce RTX 4070 Ti to investigate.

2.1. Dataset

A dataset related to customer reviews in F&B was investigated using multinomial Naive Bayes (MNB), LR, and SVM. The distinct data was collected from September until October 2020 from Ben & Jerry’s ice cream. This timeframe was to ensure that the data reflects current consumer opinions and responses to any potential product launches or marketing campaigns by Ben & Jerry’s. The dataset has 7943 records with considerably higher positive reviews than negative ones. We used two sections in the dataset: the product review and the product name. Only positive or negative reviews were considered to concentrate the sentiment analysis on distinct consumer happiness or dissatisfaction markers. Excluding neutral reviews, which frequently provide hardly comprehensive input, made the study simpler and guaranteed that the conclusions drawn were both applicable and accurately represented the strong opinions of the customers.

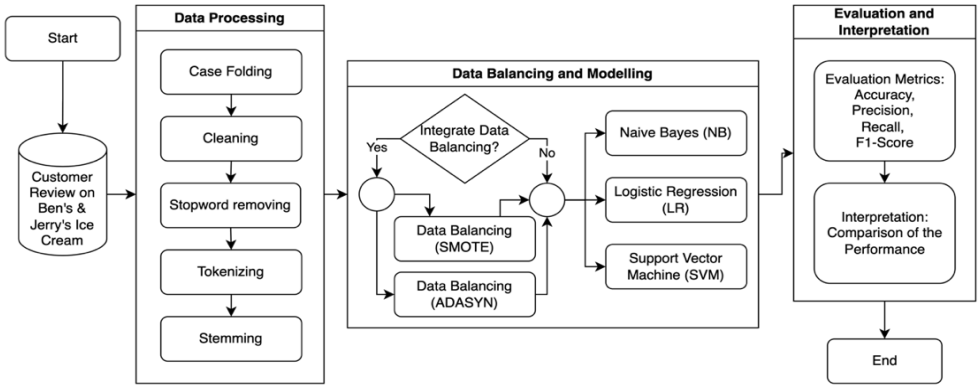


Figure 1. The flowchart of our research stages

2.2. Data processing

Customer reviews are generally in the form of unstructured text data that often contains noise, such as spelling errors and symbols. The preprocessing stage was purposed to eliminate needless words and decrease the dataset's dimensionality, making it easier to process in the following step. This study used several preprocessing techniques, including case folding, cleaning, stopword removal, tokenization, and stemming. Case folding is responsible for converting all text to lowercase to ensure consistency of all words in the text. Cleaning means cleaning text from excessive use of letters and symbols by correcting contractions, removing word repetitions, and excessive punctuation. Stopword removal eliminates less essential words (stopwords) to reduce the dimensionality of the dataset. Tokenization splits texts or sentences into individual words. Stemming is transforming words in the text to their base form.

2.3. Dataset balancing

In this study, the dataset used experienced data imbalance, where the number of samples in one category (negative category) is much less than the positive category. While there are 6,401 good ratings, there were 1,135 negative ones. These proportions between a dataset's positive and negative ratings were typically an imbalance case. The imbalance condition may affect the model's performance by distorting prediction outcomes as the model tends to be more accurate on the majority class and ignore the minority class. To investigate the imbalance impact of this issue, this research used the oversampling approach in combination with the SMOTE and ADASYN for imbalance learning.

The SMOTE technique increases the sample size in the minority class by generating new instances based on existing data. Using minority examples as a starting point, this method generates new synthetic cases comparable to yet distinct from the original examples. This method helps to balance the majority and minority classes, which is expected to improve the classification model's capacity to distinguish between positive and negative ratings more accurately [12]. On the other hand, ADASYN is a sampling approach involving learning steps. ADASYN can adjust the weight distributions of minority classes based on their learning difficulties. Subsequently, more data are synthesized for the minority class, which has higher learning difficulties. This learning-based synthesized approach is considerably beneficial in the presence of high bias, which is typical for imbalanced datasets [13].

2.4. Research methods

SVM, LR, and NB are the classification techniques employed in this research. NB is a straightforward probabilistic machine learning technique based on the ideas of feature independence and the Bayes theorem. NB is a popular solution for classification issues and excels in high-dimensional feature spaces. A response variable with two or more categories, y, and one or more continuous predictor variables, x, are related. This relationship can be explained using logistic regression. Regression analysis and classification are two common uses for SVM, a well-known machine learning technique. SVMs perform exceptionally well in high-dimensional domains and can handle challenging classification tasks. To divide the borders of various classes, SVM creates a hyperplane in multidimensional space. The scikit-learn machine learning toolkit and the Python programming language are used to implement these techniques.

2.4.1. Naive Bayes

The most substantive probability, however, is established through a technique called the NB strategy that also categorizes test data into the most appropriate classes. Being a simple probabilistic machine learning

technique, the NB classifier is based on feature independence and the Bayes theorem. A well-known NB probability classifier that applies Bayes' theorem is popularly used due to its simple applicability and effectiveness on texts [14]. The assumption that one feature does not affect another within the same class simplifies the computation process. Despite having high independence requirements, NB remains attractive for sentiment analysis since it often performs strongly in text classification tasks [15].

Moreover, NB is particularly suited for real-time applications requiring swift responses due to its processing efficiency [16]. NB is commonly applied for classification problems and performs well in high-dimensional feature space [17]. The classifier's name is derived from the "naive" assumption that every feature assigned a class label is independent of every other feature. NB uses relatively little training data relative to other classifiers and is a computationally efficient classifier. They work well when the independence criterion is met, or the approximate correlation between characteristics can be determined. These classifiers have been shown to perform strongly in almost all real-world applications, which include sentiment analysis, spam filtering, text categorization, and medical diagnosis. The MNB technique is one of the successful variants of this algorithm that works by using word frequency statistics — for instance, a binary vector in the word space [18]. Unlike the multivariate Bernoulli event model, this MNB strategy assumes that document durations are independent of the class labels within the documents.

2.4.2. Logistic regression

The application of the LR method in the analysis is a way to establish a relation between one or more continuous predictor variables, x, and a response variable that has two or more categories, y, [19]. Despite being initially developed for numeric prediction, LR has been successfully used in sentiment analysis. A statistical approach that models a relationship between one or more independent variables and the logit function of the dependent variable to predict the probability of a binary event [20]. LR can classify text data into positive or negative attitudes by converting a linear combination of input attributes into likelihood. One such change that enhances its text data processing efficiency is regularization [21], [22]. Furthermore, since LR is interpretable, understanding how each aspect influences the prediction is most likely straightforward [23]. SVM is a practical machine-learning approach to regression and classification problems. Another vital assumption to the LR analysis is that there should be no, or minimal, multicollinearity or linear solid relationship between the predictor variables to avoid issues when estimating coefficients [24], [25]. Similarly, higher complexity models may perform better than LR at handling such cases.

2.4.3. Support vector machine

The SVM is the prominent method of machine learning, typically used for classification and regression analysis. SVM assesses the information to identify the pattern or boundaries that point towards choices made within a dataset. To separate the different class borders, SVM generates hyperplanes in a multidimensional space, the number of which is referred to as the feature vector of the dataset. Further, SVM can handle very challenging classification problems and works excellently in high-dimensional spaces. Support vectors are the data points lying closest to the plane and what this algorithm uses to determine the positioning and orientation [26]. SVM comes in a variety of forms. Linear SVM is used for linearly separable data, whereas non-linear SVM maps data into higher-dimensional spaces where a linear hyperplane can separate the data [27].

2.5. Evaluation and interpretation

These methods were investigated using the Python programming language and libraries such as scikit-learn for machine learning. Implementation steps included splitting the data into training, validation, and testing sets of 80%, 10%, and 10%, respectively. To find the most appropriate hyperparameters of each classification model during the model building, we used a greedy-based cross-validation algorithm based on GridSearchCV. Data balancing was applied before and after model training to compare performance. The testing set was used with evaluation metrics such as training time, accuracy, precision, recall, and F1-score to measure each model's performance. Accuracy, precision, recall, and F1-score calculation are expressed in [28].

3. RESULTS AND DISCUSSION

Using an investigation framework, we made the necessary adjustments to support and initiate our study. The earlier investigation found that NB and LR were adequate for the classification. Thus, in this study, SMOTE and SVM were investigated to improve the performance of the earlier research. The obtained results must first undertake data preprocessing steps, features, and models such as the Naive-Bayes, LR, SVM algorithms, data balancing, and model validation. Naive-Bayes, LR, and SVM algorithms were applied for the sentiment analysis of Ben & Jerry's ice cream products without and with data balancing.

Making the results of this study reproducible and replicable, the materials with the accessible source code are available at [29].

3.1. Data preparation, multicollinearity investigation, and balancing

Using labels from the Ben & Jerry's Ice Cream Dataset based on the star rating and a combination of product name and review text, sentiment analysis of the company ice cream product reviews was performed using this model. We converted the star rating into sentiment by excluding the neutral ones. Next, we performed text preprocessing on the only predictor based on product names and reviews. Thus, no further steps were taken to handle the multicollinearity as no other feature was involved. Regarding class imbalance, to improve the results, the SMOTE technique was employed to address the issue of class disparity. Unlike methods that oversample the minority class by randomly re-sampling the minority class data, SMOTE creates synthetic samples for the minority class [12].

3.2. Performance evaluation and interpretation: comparison of sentiment analysis algorithms

The analysis included sentiment analysis with NB, LR, and SVM instruments. Based on uneven samples, the data is categorized through the NB classifier with the help of conditional probability theory. In statistics, LR is a valuable tool that allows a logistic function to assess probabilities and predict this or that binary event. On the other hand, the SVM classification approach is based on determining the hyperplane that best separates data points with different classifications with the most significant margin.

As reported in Table 1, the NB algorithm had a sentiment distribution of 96% positive and 4% negative, the LR algorithm was 89% positive and 11% negative, and the SVM algorithm had a distribution of 86% positive and 14% negative. After the data balancing process with SMOTE, the NB, and LR algorithms had a sentiment distribution of 81% positive and 19% negative. The SVM algorithm had an 84% positive and 16% negative distribution. With ADASYN, NB reached 80% positive and 20% negative, while LR and SVM have similar figures at 84% positive and 16% negative. Overall, the data balancing process using SMOTE and ADASYN reduced the imbalance in sentiment distribution.

Table 1. The performance comparison of sentiment distributions without data balancing (without), using SMOTE and ADASYN data balancing between positive reviews (pos) and negative ones (neg)

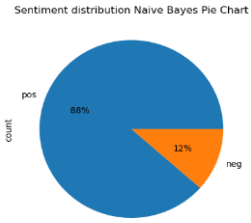
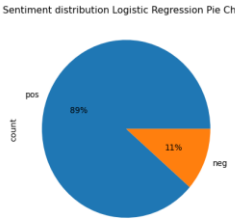
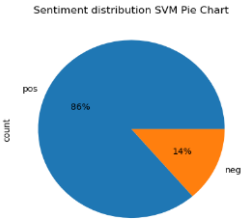
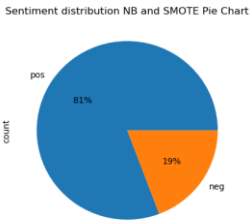
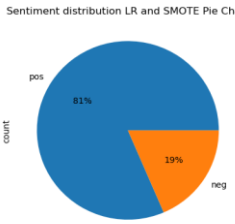
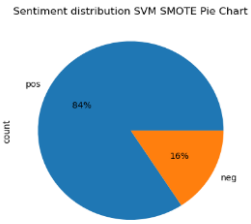
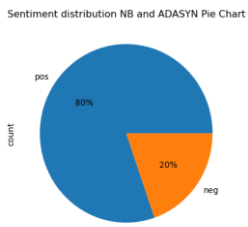
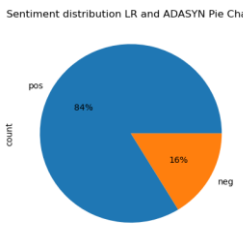
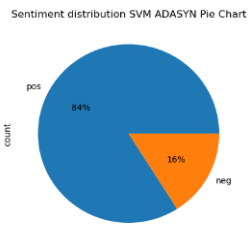
Data balancing	NB	LR	SVM
Without			
SMOTE			
ADASYN			

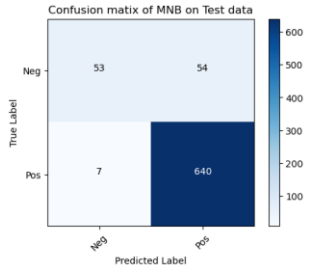
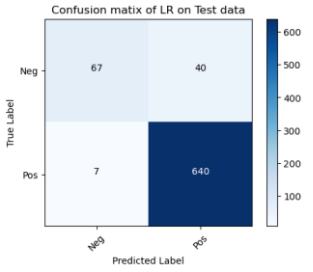
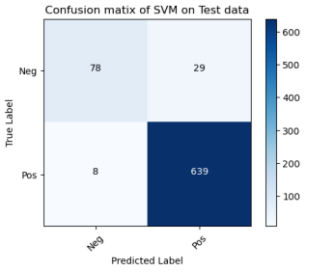
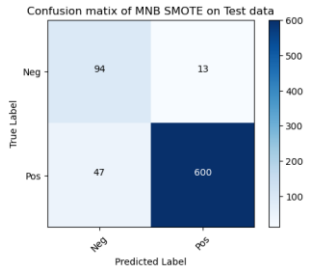
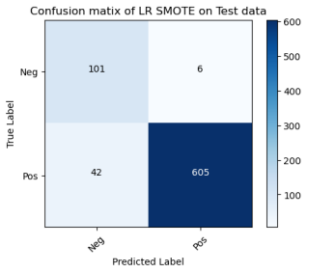
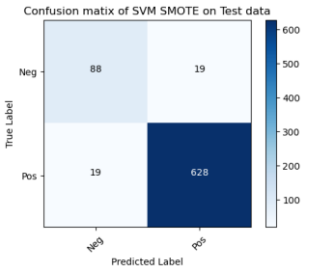
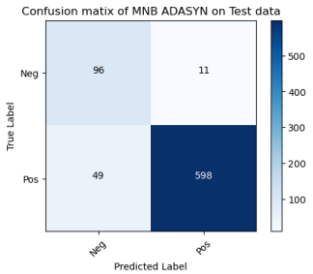
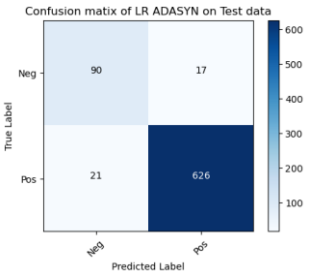
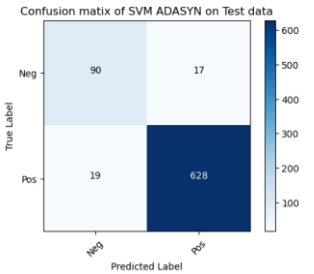
Table 2 shows the model performance evaluation results. Without data balancing, the SVM algorithm showed the best sentiment analysis performance with an accuracy of 95.09%, a training time of

0.14 seconds, a precision of 95.0%, a recall of 95.1%, and an F1-score of 94.9%. After integrating data balancing using SMOTE, the SVM algorithm showed a slight decrease in performance with an accuracy of 94.96%, a training time of 0.21 seconds, a stable precision of 95.0%, a recall of 95.0%, and an improved F1-score of 95.0%. Similarly, LR suffered a performance decrease in accuracy from 93.77% to 93.63% after data balancing using SMOTE. However, the precision score was raised to 95.0%, which is comparable to SVM's. Either NB or LR has the fastest training time and is nearly instant, making both methods promising algorithms for real-time sentiment analysis. Applying SMOTE and ADASYN generally increased the processing time across all three algorithms. Based on the confusion matrices in Table 3, applying SMOTE and ADASYN also typically develops the detection of the minority class (negative samples) across all three algorithms. Contrarily, the development of the minority class may have affected the true negative (TN) and false positive (FP) rates. Overall, SVM shows relatively stable performance with minor updates on both the TP and TN, making it still a robust choice for imbalanced datasets in sentiment analysis.

Table 2. Comparison of the algorithm performance in training time in second (s), accuracy, precision, recall, and F1-score

Algorithm	Data balancing	Training time (s)	Testing results			
			Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
NB	Without	0.00	91.90	91.7	91.9	90.9
	SMOTE	0.00	92.04	93.4	92.0	92.5
	ADASYN	0.00	92.04	93.7	92.0	92.5
LR	Without	0.00	93.77	93.6	93.8	93.3
	SMOTE	0.00	93.63	95.0	93.6	94.0
	ADASYN	0.00	94.96	95.0	95.0	95.0
SVM	Without	0.14	95.09	95.0	95.1	94.9
	SMOTE	0.21	94.96	95.0	95.0	95.0
	ADASYN	0.22	**95.23**	**95.3**	**95.2**	**95.2**

Table 3. Comparison of the confusion matrices displaying the numbers of true positive (TP), TN, FP, and false negative (FN)

Data balancing	NB	LR	SVM
Without			
SMOTE			
ADASYN			

3.3. Discussion

This study investigated the effects of data balancing techniques on the performance metrics of sentiment analysis utilizing NB, LR, and SVM algorithms. While prior research has explored the impact of data balancing, the current investigation addresses its influence when the data exhibits considerable skewness, as evaluated through accuracy, precision, recall, and F1-score metrics in the context of sentiment analysis. The findings indicate that the SVM algorithm achieved the highest accuracy without any data balancing intervention, while the application of ADASYN improved the overall detection of the minority class. Although the implementation of SMOTE marginally reduced the SVM's overall accuracy to 94.96% and increased the processing time, it remained the best-performing algorithm for sentiment analysis tasks. In contrast, the LR model also experienced a slight decrease in accuracy after the data balancing process. Nonetheless, the NB and LR algorithms offered the fastest training times among the investigated models.

The current study suggests that higher accuracy does not inherently equate to poorer performance in representing the minority class. The proposed approaches may gain advantages from data balancing without significantly compromising the overall accuracy. This finding aligns with existing scholarly literature [30], which indicates a trade-off between accuracy and minority class representation in sentiment analysis. The present investigation examined a range of algorithms and balancing techniques. However, further in-depth examinations may be necessary to confirm the generalizability of these results, particularly regarding the influence of varying levels of imbalance and the use of advanced neural network models.

Our research reveals that data balancing methods are more robust than focusing solely on the performance of the majority class. Future investigations may explore the connection between sentiment and specific product characteristics, identifying viable approaches to generating accurate and unbiased sentiment analysis. Recent observations suggest that the challenges posed by imbalanced datasets in sentiment analysis can lead to biased results and misinformed business decisions [31]. Our findings demonstrate that this phenomenon is associated with changes in performance rather than solely attributable to increased false positives or negatives.

4. CONCLUSION

Based on the results, we provided collaborated investigations covering the metrics for the three algorithms without and with the data balancing process. With balanced data, all algorithms demonstrated substantial performance improvements. NB exhibited the most significant improvements across all metrics, with accuracy increasing from 91.91% to 92.04%, precision from 91.7% to 93.7%, recall rising from 91.9% to 92.0%, and the F1-score substantially improving from 90.9% to 92.5%. LR showed performance enhancements, slightly increasing the accuracy from 93.77% to 94.96%. Its precision increased marginally from 93.6% to 95.0%, and the F1-Score improved from 93.3% to 95.0%. SVM remained the best-performing algorithm for sentiment analysis without and with data balancing, achieving the highest scores across all metrics. The LR and SVM performance showed relatively negative trends using SMOTE. However, integrating ADASYN into SVM improved the overall performance, with accuracy marginally increasing from 95.09% to 95.23%, precision relatively remaining stable at 95.3%, recall growing somewhat from 95.1% to 95.2%, and the F1-Score improving from 94.9% to 95.2%. These performance improvements with SMOTE and ADASYN, particularly in F1-score for NB, LR, and SVM, indicate that the data balancing process has slightly enhanced the understanding and prediction of minority classes.

Based on the performance metrics on ADASYN, the models have become more reliable in sentiment analysis, with a more balanced prediction distribution across sentiment classes. The improvements across all metrics demonstrate that the balanced sampling of all classes during the learning process has led to more precise and dependable sentiment reading from the provided data, benefiting in a superior sentiment analysis evaluation only for NB. Investigating neutral reviews would be of interest, as we hypothesize that this direction will present a broader view of sentiment analysis on skewed product reviews.

ACKNOWLEDGMENTS

The author acknowledges the support and academic resources provided by the Faculty of Information Technology and Data Science, Universitas Sebelas Maret, which facilitated the completion of this research.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

The research project involved multiple contributors with varied responsibilities. Nabilla Nurulita Dewi was extensively involved across most aspects, contributing to conceptualization, methodology, software, formal analysis, investigation, resources, data curation, original drafting, editing, and visualization. Sekar Gesti Amalia Utami focused on conceptualization, methodology, and software development. Shalsabila Aura Adiar had a significant role, participating in conceptualization, methodology, software, formal analysis, investigation, resources, data curation, original drafting, editing, and visualization. Hasan Dwi Cahyono contributed to methodology, validation, editing, project administration, and funding acquisition, suggesting a supervisory or administrative role in the research team, with particular emphasis on securing financial resources for the project.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Nabilla Nurulita Dewi	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓			
Sekar Gesti Amalia Utami	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓			
Shalsabila Aura Adiar	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓			
Hasan Dwi Cahyono		✓	✓	✓	✓	✓				✓		✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest

DATA AVAILABILITY

- The dataset used in this study is publicly available on Kaggle at https://www.kaggle.com/datasets/tysonpo/ice-cream-dataset.
- The source code and preprocessing scripts that support the findings of this study are available at https://github.com/nabilland/ice-cream-product-sentiment-analysis.

REFERENCES

[1] V. Shankar, D. Grewal, S. Sunder, B. Fossen, K. Peters, and A. Agarwal, "Digital marketing communication in global marketplaces: a review of extant research, future directions, and potential approaches," *International Journal of Research in Marketing*, vol. 39, no. 2, pp. 541–565, Jun. 2022, doi: 10.1016/j.ijresmar.2021.09.005.

[2] G. Ma, J. Ma, H. Li, Y. Wang, Z. Wang, and B. Zhang, "Customer behavior in purchasing energy-saving products: big data analytics from online reviews of e-commerce," *Energy Policy*, vol. 165, p. 112960, Jun. 2022, doi: 10.1016/j.enpol.2022.112960.

[3] G. S. Budhi, R. Chiong, I. Pranata, and Z. Hu, "Using machine learning to predict the sentiment of online reviews: a new framework for comparative analysis," *Archives of Computational Methods in Engineering*, vol. 28, no. 4, pp. 2543–2566, Jan. 2021, doi: 10.1007/s11831-020-09464-8.

[4] B. J. Ali *et al.*, "Hotel service quality: the impact of service quality on customer satisfaction in hospitality," *International Journal of Engineering, Business and Management*, vol. 5, no. 3, pp. 14–28, 2021, doi: 10.22161/ijebm.5.3.2.

[5] J. Schilling, B. Schyns, and D. May, "When your leader just does not make any sense: conceptualizing inconsistent leadership," *Journal of Business Ethics*, vol. 185, no. 1, pp. 209–221, Apr. 2023, doi: 10.1007/s10551-022-05119-9.

[6] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artificial Intelligence Review*, vol. 55, no. 7, pp. 5731–5780, Feb. 2022, doi: 10.1007/s10462-022-10144-1.

[7] S. A. H. Bahtiar, C. K. Dewa, and A. Luthfi, "Comparison of Naïve Bayes and logistic regression in sentiment analysis on marketplace reviews using rating-based labeling," *Journal of Information Systems and Informatics*, vol. 5, no. 3, pp. 915–927, Aug. 2023, doi: 10.51519/journalisi.v5i3.539.

[8] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, Sep. 2020, doi: 10.1016/j.neucom.2019.10.118.

[9] P. Ray, S. S. Reddy, and T. Banerjee, "Various dimension reduction techniques for high dimensional data analysis: a review," *Artificial Intelligence Review*, vol. 54, no. 5, pp. 3473–3515, Jan. 2021, doi: 10.1007/s10462-020-09928-0.

[10] S. F. Hussain, "A novel robust kernel for classifying high-dimensional data using support vector machines," *Expert Systems with Applications*, vol. 131, pp. 116–131, Oct. 2019, doi: 10.1016/j.eswa.2019.04.037.

[11] F. Nie, W. Zhu, and X. Li, "Decision tree SVM: an extension of linear SVM for non-linear classification," *Neurocomputing*, vol. 401, pp. 153–159, Aug. 2020, doi: 10.1016/j.neucom.2019.10.051.

[12] D. Elreedy, A. F. Atiya, and F. Kamalov, "A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning," *Machine Learning*, vol. 113, no. 7, pp. 4903–4923, Jan. 2024, doi: 10.1007/s10994-022-06296-4.

[13] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: adaptive synthetic sampling approach for imbalanced learning," in *Proceedings of the International Joint Conference on Neural Networks*, Jun. 2008, pp. 1322–1328, doi: 10.1109/IJCNN.2008.4633969.
[14] H. Gao, X. Zeng, and C. Yao, "Application of improved distributed naive Bayesian algorithms in text classification," *Journal of Supercomputing*, vol. 75, no. 9, pp. 5831–5847, Apr. 2019, doi: 10.1007/s11227-019-02862-1.
[15] J. Hartmann, J. Huppertz, C. Schamp, and M. Heitmann, "Comparing automated text classification methods," *International Journal of Research in Marketing*, vol. 36, no. 1, pp. 20–38, Mar. 2019, doi: 10.1016/j.ijresmar.2018.09.009.
[16] K. Maswadi, N. A. Ghani, S. Hamid, and M. B. Rasheed, "Human activity classification using decision tree and Naïve Bayes classifiers," *Multimedia Tools and Applications*, vol. 80, no. 14, pp. 21709–21726, Mar. 2021, doi: 10.1007/s11042-020-10447-x.
[17] B. Ghaddar and J. Naoum-Sawaya, "High dimensional data classification and feature selection using support vector machines," *European Journal of Operational Research*, vol. 265, no. 3, pp. 993–1004, Mar. 2018, doi: 10.1016/j.ejor.2017.08.040.
[18] S. Ghosh, A. Dasgupta, and A. Swetapadma, "A study on support vector machine based linear and non-linear pattern classification," in *Proceedings of the International Conference on Intelligent Sustainable Systems, ICISS 2019*, Feb. 2019, pp. 24–28, doi: 10.1109/ISS1.2019.8908018.
[19] E. Y. Boateng and D. A. Abaye, "A review of the logistic regression model with emphasis on medical research," *Journal of Data Analysis and Information Processing*, vol. 07, no. 04, pp. 190–207, 2019, doi: 10.4236/jdaip.2019.74012.
[20] E. W. Steyerberg, "Statistical models for prediction," in *Clinical Prediction Models*, Springer International Publishing, 2019, pp. 59–93.
[21] K. Shah, H. Patel, D. Sanghvi, and M. Shah, "A comparative analysis of logistic regression, random forest and KNN models for the text classification," *Augmented Human Research*, vol. 5, no. 1, Mar. 2020, doi: 10.1007/s41133-020-00032-0.
[22] L. Khairunnahar, M. A. Hasib, R. H. Bin Rezanur, M. R. Islam, and M. K. Hosain, "Classification of malignant and benign tissue with logistic regression," *Informatics in Medicine Unlocked*, vol. 16, p. 100189, 2019, doi: 10.1016/j.imu.2019.100189.
[23] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: a survey on methods and metrics," *Electronics (Switzerland)*, vol. 8, no. 8, p. 832, Jul. 2019, doi: 10.3390/electronics8080832.
[24] N. Shrestha, "Detecting multicollinearity in regression analysis," *American Journal of Applied Mathematics and Statistics*, vol. 8, no. 2, pp. 39–42, Jun. 2020, doi: 10.12691/ajams-8-2-1.
[25] J. Y. Le Chan *et al.*, "Mitigating the multicollinearity problem and its machine learning approach: a review," *Mathematics*, vol. 10, no. 8, p. 1283, Apr. 2022, doi: 10.3390/math10081283.
[26] W. S. Noble, "What is a support vector machine?," *Nature Biotechnology*, vol. 24, no. 12, pp. 1565–1567, Dec. 2006, doi: 10.1038/nbt1206-1565.
[27] R. Dixit, R. Kushwah, and S. Pashine, "Handwritten digit recognition using machine and deep learning algorithms," *International Journal of Computer Applications*, vol. 176, no. 42, pp. 27–33, Jul. 2020, doi: 10.5120/ijca2020920550.
[28] S. Lee, H. Kim, H. Cho, and H. J. Jo, "FIDS: filtering-based intrusion detection system for in-vehicle CAN," *Intelligent Automation and Soft Computing*, vol. 37, no. 3, pp. 2941–2954, 2023, doi: 10.32604/iasc.2023.039992.
[29] "nabilland/ice-cream-product-sentiment-analysis," *Github*, 2025. https://github.com/nabilland/ice-cream-product-sentiment-analysis (accessed Mar. 23, 2025).
[30] M. Kamruzzaman and G. Kim, "Efficient sentiment analysis: a resource-aware evaluation of feature extraction techniques, ensembling, and deep learning models," in *Proceedings of the 11th International Workshop on Natural Language Processing for Social Media*, 2024, pp. 9–20, doi: 10.18653/v1/2023.socialnlp-1.2.
[31] R. Mubarak, T. Alsboui, O. Alshaikh, I. Inuwa-Dutse, S. Khan, and S. Parkinson, "A survey on the detection and impacts of deepfakes in visual, audio, and textual formats," *IEEE Access*, vol. 11, pp. 144497–144529, 2023, doi: 10.1109/ACCESS.2023.3344653.

BIOGRAPHIES OF AUTHORS

Nabilla Nurulita Dewi is a student of the Faculty of Information Technology and Data Science at Universitas Sebelas Maret (UNS) in Surakarta, Central Java, Indonesia. She has interests in data science, software development, cloud computing, and computer vision. She can be contacted at email: nabillanurulita17@gmail.com.

Sekar Gesti Amalia Utami is a student of the faculty of information technology and data science at Universitas Sebelas Maret (UNS) in Surakarta, Central Java, Indonesia. She has interests in programming, data science, cloud computing, software development, and web development. She can be contacted at email: sekargestiau@gmail.com.

Shalsabila Aura Adiar is a student of the faculty of information technology and data science at Universitas Sebelas Maret (UNS) in Surakarta, Central Java, Indonesia. She has interests in programming, data science, cloud computing, software development, and web development. She can be contacted at email: shalsabilaauraadiar@gmail.com.

Hasan Dwi Cahyono is a member of the faculty of information technology and data science at Universitas Sebelas Maret (UNS) in Surakarta, Central Java, Indonesia. His interests are centered across software engineering, generative modelling, normalizing flow, a crucial aspect that underpins distribution transformations of generative modelling. Although he has a tropical origin, he has strong fascination for snow. He is reachable via his email: hasan.dwi.cahyono@gmail.com.