Jurnal Sistem Informasi Bisnis 03(2025)
 p-ISSN: 2502-2377, e-ISSN: 2088-3587
On-*line*: http://ejournal.undip.ac.id/index.php/jsinbis/article/view/60543

Comparison of Sentiment Analysis Models Using Machine Learning Methods for Customer Response Evaluation (Case Study: Bosca Living)

Ryan Pangestu, Tio Alwi Priantama, Virman Agustian, Sulis Sandiwarno*.

Faculty of Computer Science, Information Systems, Mercu Buana University, Jakarta, Indonesia

Submitted: December 12th, 2023; Revised: January 8th, 2025; Accepted: June 17th, 2025; Available Online: June 17th, 2025
DOI: 10.21456/vol15iss3pp359-366

Abstract

Bosca Living, a star seller on Shopee and Tokopedia, faces challenges in customer sentiment analysis. This research evaluates models and methods to strengthen responses to customer feedback. In previous studies, feature extraction techniques such as Term Frequency-Inverse Document Frequency (TF-IDF), Word2Vec, FastText, and Global Vectors for Word Representation (GloVe) have been tested. Machine learning models like K-Nearest Neighbors (KNN), Random Forest, Support Vector Classifier (SVC), XGBoost, Logistic Regression, and Decision Tree have been used, but a more in-depth comparison is needed as per Bosca Living's assessment. This research proposes a model comparison through pre-processing, feature extraction, and parameter determination using GridSearchCV. Machine learning models such as KNN, Random Forest, SVC, XGBoost, Logistic Regression, and Decision Tree are evaluated with StratifiedKFold to reduce the risk of overfitting. The goal is to provide in-depth insights, enabling Bosca Living to enhance responses to customer reviews. The results of this research are expected to optimize business strategies, support continuous improvement, and be responsive to market dynamics and evolving customer needs.

Keywords: *Bosca Living; Customer Sentiment Analysis; Machine Learning Models; Feature Extraction*

1. Introduction

In the rapidly evolving digital era, e-commerce has emerged as a vital aspect of consumer life worldwide. Swift technological advancements have revolutionized internet accessibility, transforming shopping habits. With thousands of products and services at their fingertips, e-commerce customers rely heavily on reviews and ratings from previous purchases to inform buying decisions. These reviews provide invaluable insights into product quality and offer perspectives on shopping experiences, guiding potential buyers in selecting reliable e-commerce platforms.

In the increasingly competitive e-commerce landscape, understanding customer sentiment is crucial for maintaining customer satisfaction, enhancing brand reputation, and improving product quality. This study analyzes customer sentiment in Bosca Living online store reviews, aiming to provide valuable insights into online shopping experiences and customer satisfaction. A primary challenge is the complexity of interpreting customer sentiment across 1-5 rating scales. This varied range requires meticulous analytical approaches. Ratings reflect diverse sentiment levels, from extremely negative (1) to extremely positive (5). Mid-range ratings (e.g., 3) often convey ambiguous meanings, making categorization difficult. To address this, the study focuses on comparing feature extraction techniques (TF-IDF, Word2Vec, FastText, GloVe) and machine learning models (KNN, RF, SVC, LR, XB, DT) to effectively capture subtle sentiment nuances and explicit differences. (Demidova & Gorchakov, 2021; Domashova & Gultiaev, 2021; Larose, 2019; Taufiq et al., 2020).

Previous studies compared various machine learning algorithms, but lacked consideration for Bosca Living's specific needs. This research conducts an in-depth comparison, evaluating the unique characteristics of Bosca Living customer reviews, including varying text lengths, imbalanced sentiment distributions, and linguistic patterns. The evaluation encompasses a comprehensive analysis of optimal feature extraction methods (TF-IDF, Word2Vec, FastText, GloVe) and classification models (KNN, RF, SVC, LR, XB, DT), utilizing GridSearchCV for parameter optimization.

To address sentiment analysis challenges in Bosca Living product reviews, this study compares the performance of multiclass classification models. Multiclass classifiers are machine learning algorithms categorizing data into more than two classes (Sandiwarno, 2020; Sandiwarno et al., 2023). This research focuses on evaluating and comparing classification models' effectiveness in handling datasets with broad rating scales (1-5). By applying classifiers to these ratings, the study aims to gain deeper insights into their accuracy in categorizing diverse customer reviews (Nugroho & Eliyani, 2021).

Previous studies (Muktafin et al., 2020) tdemonstrated the effectiveness of KNN and XGBoost for sentiment analysis. However, most research neglected data imbalance challenges and parameter optimization using techniques like GridSearchCV.

*)Corresponding author: sulis.sandiwarno@mercubuana.ac.id

This study contributes by evaluating optimal feature and algorithm combinations for customer review data, specifically within Indonesia's e-commerce context. The classifier performance analysis provides valuable insights into algorithms' capabilities in identifying and categorizing sentiments based on Bosca Living product review ratings. This research aims to offer practical insights for Bosca Living's business strategy development.

Sentiment analysis conducted not only aims to better understand customer reviews but also provides strategic insights for Bosca Living to identify relevant sentiment patterns aligned with market trends. Consequently, the research findings can support new product development, enhance customer experience, and inform more effective data-driven marketing strategies.

In a study by Muktafin et al. (Muktafin et al., 2020), researchers analyzed customer sentiment reviews of instant hijab products on Shopee marketplace. Utilizing KNN algorithm, Term Frequency-Inverse Document Frequency (TF-IDF) weighting, and Natural Language Processing (NLP) approach, the study demonstrated significant improvement in sentiment classification accuracy following NLP implementation in data preprocessing.

The study by Tyagi and Sharma (Tyagi & Sharma, 2018) also focused on sentiment analysis, using logistic regression and heuristic effective word scores to classify tweets as subjective or objective. While finding that machine learning methods are effective in sentiment classification, this study highlights the challenges associated with requiring large training datasets.

In Bahrawi's research (Bahrawi, 2019), sentiment analysis was conducted on data collected from various sources using the Naïve Bayes, Random Forest, and SVM algorithms. These methods achieved the highest accuracy in sentiment analysis of Twitter data.

The study by Song and Liu (Song & Liu, 2020) explored the use of the XGBoost algorithm to predict consumer purchasing behavior on e-commerce platforms. The results demonstrated that XGBoost improved predictive performance compared to traditional Random Forest algorithms.

Benarafa et al. (Benarafa et al., 2023) proposed an SVM model enhanced with external knowledge from WordNet for identifying implicit aspects in sentiment analysis. The findings indicated improvements in SVM classification, particularly in addressing overfitting and underfitting issues.

Research by Kiedrowsky and Andrianingsih (Kiedrowsky & Andrianingsih, 2023) focused on consumer satisfaction analysis of payment systems across various platforms. Using Naïve Bayes and Decision Tree methods, the study provided valuable insights into consumer satisfaction levels regarding payment systems, although Decision Tree exhibited lower accuracy compared to Naïve Bayes.

Referring to the findings of previous studies and the review presented above, this research focuses on comparing feature extraction methods, namely: TF-IDF (Wang et al., 2020), Word2Vec (Onan, 2021), FastText (Onan, 2020), dan GloVe (Krishna Das et al., 2022), as well as machine learning models, such as: KNN (Demidova & Gorchakov, 2021), RF (Iyer et al., 2019), SVC (Bi et al., 2019), LR (Domashova & Gultiaev, 2021), XB (Song & Liu, 2020), dan DT (Almaiah et al., 2021).

The aim of this research is to provide deeper insights into consumers' experiences and perspectives regarding products on the Bosca Living e-commerce platform, as well as to offer valuable recommendations for store owners to improve the quality of their services and products.

2. Theoretical Framework

2.1. K-Nearest Neighbor

K-Nearest Neighbors (KNN) is a classification algorithm that predicts the label of an instance based on the majority label among its kk-nearest neighbors. KNN uses the concept of distance to determine the nearest neighbors. The purpose of this method is to identify kk instances in the training data that have the most similar spatial patterns to the sample. The Euclidean distance formula between two instances A and B in an n-dimensional feature space is shown in the equation (1).

$$\text{Distance } A, B = \sqrt{\sum_{i=1}^{n}(A_i - B_i)^2} \quad (1)$$

With A_i and B_i represent the feature values of instances A and B at the i-th dimension, while n indicates the entire number of dimensions or features used in the analysis.

The implementation of KNN usually requires selecting the appropriate kk parameter and choosing the distance metric suitable for the data. In addition, data preprocessing, such as feature normalization, can also influence the performance of KNN.

2.2. Decision Tree

TF-Decision Tree is a highly effective and well-known method for classification and prediction. Decision Trees also help individuals explore data and uncover hidden relationships between various input factors and the desired outcomes. There are several stages in creating a decision tree: preparing the data for training, calculating the root of the tree, determining the root by calculating the entropy value, and then calculating the gain for each attribute. The attribute with the highest gain becomes the first root of the decision tree, optimal for reducing uncertainty in data classification or prediction. The calculation of entropy values can be seen in the equation (2).

$$\text{Entropy (S)} = \sum_{i=1}^{n} -pi * \log_2 pi \quad (2)$$

With S as the set of cases, n as the number of partitions of S, and pi as the proportion of S_i to S.

Calculating the Information Gain can be seen in the equation (3).

$$Gain\ (S,A) = \text{Entropy (S)} \sum_{i=1}^{n} \frac{|S_i|}{|S|} x\ Entropy(S_i) \qquad (3)$$

With S as the set of cases, A as an attribute, n as the number of attribute partitions of A, $|S_i|$ as the cardinality of partition i, and $|S|$ as the total cardinality of S.

2.3. Support Vectore Machine

Classification is performed using the SVM method to generate system output. SVM is a classification method focused on finding the hyperplane with the largest margin. The hyperplane is the dividing line between classes, while the margin is the distance between the hyperplane and the closest data in each class. Data closest to the hyperplane are called support vectors. The hyperplane calculation equation is (4).

$$f(x) = sign\left(\sum_{i=1}^{N} \alpha_i y_i K(X_i, X) + b\right) \qquad (4)$$

With $f(x)$ as the decision function generating class predictions for observation x, α_i as the weight assigned to each support vector X_i, and y_i as the class label of observation X_i, the classification process is determined based on these parameters. Additionally $K(X_i, X)$ represents the kernel function that measures similarity between observations X_i and X in the feature space using the kernel trick, while b serves as the threshold value in the decision function.

2.4. Logistic Regression

Logistic Regression is a method of modeling the relationship between an independent variable and a dependent variable. An example is determining whether a certain text belongs to a positive or negative sentence. Linear Regression has several types, namely Binary Logistic Regression if it only has 2 outputs, Multinomial Logistic Regression if it has 2 or more outputs, and Ordinal Logistic Regression if it has 2 or more outputs while considering the order. The calculation of this algorithm can be seen in equation (5).

$$P(Y=1) = \frac{1}{1 + e^{(b_0 + b_1 X_1 + b2_{X2} \ldots + b_k X_k)}} \qquad (5)$$

$$P(Y=0) = 1 - P(Y=1)$$

With $P(Y=1)$ representing the probability of an observation belonging to class 1 and $P(Y=0)$ representing the probability of an observation belonging to class 0, the classification model estimates these probabilities using a logistic function. The parameter e denotes the base of the natural logarithm (Euler's number), while $b_0 + b_1 + b_2 \ldots + b_k$ are model coefficients that must be estimated from the training data. Additionally, $X_1 + X_1 \ldots + X_k$ represent the input variables or features used in the model.

2.5. Random Forest

Random Forest is a technique in ensemble learning that builds a large number of decision trees during the training process and combines the prediction results of each tree to produce a final prediction. This algorithm can be applied in the context of both classification and regression tasks. There are several stages in performing random forest, namely dividing the sample values (X) and labels (Y), creating a random forest using the data, which can be seen in the equation (6).

$$RF = \{T_1, T_2, \ldots, T_b\} \qquad (6)$$

With RF representing Random Forest, T_b refers to the data in the ensemble that learns to predict the class based on the label.

Classification for classifying new samples (X_{new}):

$$\hat{y}new = mode\big(T_1(X_{new}), T_2(X_{new}), \ldots, T_B(X_{new})\big)$$

The information for the classification process includes $X_{new,}$ which represents a numerically vectorized sample, ŷnew, which denotes the predicted class for the new sample, and $T_B(X_{new})$, which refers to the prediction made by the b-th tree for X_{new}.

2.6. XGBoost

XGBoost, or Extreme Gradient Boosting, is a highly efficient and popular ensemble algorithm for regression, classification, and ranking tasks. The formula for this algorithm can be seen in the equation (7).

$$\hat{y_i} = \sum_{k=1}^{K} f_k(X_i), f_k \in \mathcal{F} \qquad (7)$$

With $\hat{y_i}$ representing the prediction for the i-th observation, K denoting the number of decision trees (weak learners) in the ensemble, and f_k as the decision function of the k-th tree, the classification process relies on multiple trees to enhance prediction accuracy. Additionally, X_i represents the feature vector of the i-th observation, which serves as the input for the model.

3. Method

3.1 Problem Identification

The initial stage of the research is to identify customer sentiment towards a company's products or services. By analyzing customer reviews and feedback on e-commerce platforms, this research aims to achieve several strategic goals in the context of sentiment analysis.

In this research, the data used is review data on Shopee and Tokopedia obtained by accessing the Application Programming Interface (API) to retrieve JavaScript Object Notation (JSON) data on e-commerce. The research data is included in textual and qualitative data because the research data is customer reviews in the form of text and the purpose of this research is to conduct sentiment analysis of customer reviews.

3.2 Data Pre-processing

The data pre-processing stage is carried out for each use of the TF-IDF, Word2Vec, GloVe, and

FastText methods. Preprocessing is an essential step in Information Retrieval, Text mining, and NLP with the aim of determining which sentences should be taken in a document to meet user information needs. In the context of Text Mining, data pre-processing is used to extract significant knowledge from unstructured text. The unstructured format of the data and the diversity in writing become challenges in text analysis because words can appear in different formats. In addition, additional problems arise with the presence of irrelevant words or noise in the text data, which, if not removed, can make interpretation difficult and obscure the results of the analysis.

The data pre-processing stage in this research involves several essential steps to ensure that the data is ready to be used in sentiment analysis and machine learning models. This process includes:

3.2.1 Cleansing

Cleansing is removing empty or null data and irrelevant elements such as symbols and numbers.

3.2.2 Regular Expression (Regex)

Regular Expression (Regex) is filtering data to leave only alphabetical characters (a-z) and removing other characters.

3.2.3 Case Folding

Case Folding is converting all text to lowercase to ensure consistency.

3.2.4 Tokenizing

Tokenizing is breaking sentences into individual words or tokens for easier further analysis.

3.2.5 Stemming

Stemming is reducing words to their root form by removing affixes, using Python libraries such as Sastrawi.

3.2.6 Filtering

Filtering is removing irrelevant connecting words (stopwords) to improve analysis efficiency.

3.2.7 Merged Word

Merged word is combining words like "tidak" with the following word to ensure that the meaning is not distorted (for example, "tidak_rusak").

After the pre-processing stage is complete, the clean and structured data is then extracted into numerical features using methods such as TF-IDF, Word2Vec, FastText, or GloVe. This stage ensures that the data can be processed by machine learning models to produce accurate and relevant sentiment analysis.

3.3 Data Extraction

In text processing, it is necessary to extract words into numerical form because computers, in principle, cannot process data other than numerical data. Feature extraction aims to uncover potential information and represent words as feature vectors. These vectors are then used as input for classification methods in the next stage. Some techniques in feature extraction are TF-IDF, Word2Vec, FastText, and GloVe.

4. Results and Discussion

4.1 Experiment Setup

This research utilizes several libraries to implement multi-class classification in machine learning.

4.2 Dataset

This research uses a dataset of e-commerce reviews from Tokopedia and Shopee for the Bosca Living store. The dataset contains four attributes: Review, ProductName, Category, and Rating, as shown in Table 1.

Table 1. Research Dataset

No	Category	Rating (%)					Total (%)
		1	2	3	4	5	
1	Decoration and Accessories	7,42	4,45	12,62	21,34	54,17	100
2	Services and Additional Products	45,29	3,77	15,09	15,09	20,76	100
3	Equipment and Storage	9,65	5,54	9,18	21,99	53,64	100

4.3 Machine Learning Classifer Evaluation Step

In this study, the performance of each classifier was evaluated using the 10-fold cross-validation method. Additionally, performance metrics were measured through average precision (Pre), recall (Rec), F1-scores (F1), and accuracy (Accuracy). The evaluation was conducted after the data training process for each algorithm applied to binary classification, including Random Forest, SVM, KNN, Logistic Regression, Decision Tree, and XGBoost. Once the models were trained using the training data, evaluations were performed to measure prediction accuracy and the models' ability to generalize to new data. This evaluation process is essential for assessing model performance, identifying potential weaknesses, making improvements, and gaining a better understanding of the models.

The evaluation of models in this study utilized a confusion matrix to calculate the ratio of correct and incorrect predictions. According to Sulis Sandiwarno et al. (Sandiwarno et al., 2023), the evaluation was conducted after the data training process for each algorithm applied to binary classification, including Random Forest, SVM, KNN, Logistic Regression, Decision Tree, and XGBoost. Once the models were trained using the training data, evaluations were performed to measure prediction accuracy and the models' ability to generalize to new data. This evaluation process is useful for assessing model performance, identifying potential weaknesses, making improvements, and enhancing understanding of the model. The model evaluation in this study utilized a confusion matrix to calculate the ratio of correct and incorrect predictions, as stated by Sulis Sandiwarno et al. (Sandiwarno et al., 2023).

4.4 Model Implementation

In this study, the use of Stratified K-Fold (SKF) with 10 folds (10-fold) (n_splits=10) was applied to ensure balanced class distribution in each fold. The shuffle=True setting helped randomize the dataset before splitting, while random_state=0 ensured reproducibility of the shuffling process. The amount of training data in each iteration was Fold 1: Training Data – 1,101, Testing Data – 123.

To address the issue of class imbalance in the dataset, this study employed the Synthetic Minority Over-sampling Technique (SMOTE) to increase the number of samples in the minority class, thereby achieving a more balanced class distribution. The SMOTE object was created with the parameter configuration sampling_strategy='auto'. The 'auto' setting ensures automatic adjustment of synthetic sample numbers to match the majority class. The oversampling process was performed using the fit_resample method on the feature matrix (X) and the encoded target vector (y).

In this study, GridSearchCV was used to select the optimal values for hyperparameters. GridSearchCV is a technique in machine learning that helps fine-tune a model's hyperparameters comprehensively. By exploring combinations of hyperparameter values within a predefined "grid," GridSearchCV ensures the best configuration for the model is identified.

By utilizing cross-validation techniques, this tool not only systematically tests each combination but also evaluates the model's performance for every configuration, helping to minimize overfitting and optimize the model's ability to handle new data, Ghulab Nabi Ahmad et al. (Ahmad et al., 2022).

4.5 Research Results

In this section, we present the results of the research that has been conducted. The results are presented in Table 2, which shows the performance of several machine learning models as well as the various feature extraction techniques used.

Tabel 2. Research Results

Feature	Model	Metric			Feature	Model	Metric		
		Pre	Rec	F1			Pre	Rec	F1
TF-IDF	KNN	74.23	74.40	72.20	Word2Vec	KNN	70.81	70.62	68.23
	RF	77.80	77.85	77.26		RF	71.08	72.53	71.72
	SVC	76.83	76.12	76.75		SVC	63.36	75.02	70.30
	XGBoost	**83.29**	**83.36**	**83.73**		XGBoost	**75.20**	**76.86**	**76.64**
	LR	73.34	73.11	73.08		LR	62.32	66.54	63.32
	DT	72.22	72.39	72.28		DT	61.83	62.78	60.01
fastText	KNN	51.81	67.62	67.34	GloVe	KNN	58.44	74.58	74.23
	RF	67.08	68.53	67.72		RF	71.49	75.60	73.62
	SVC	58.36	70.02	64.30		SVC	66.44	74.33	72.13
	XGBoost	**71.20**	**72.86**	**72.64**		XGBoost	**75.63**	**78.20**	**77.60**
	LR	59.32	63.54	60.32		LR	68.29	70.19	69.15
	DT	58.83	58.78	56.01		DT	63.91	64.39	63.53

Through the classification evaluation results, it is evident that the XGBoost model with TF-IDF feature extraction outperforms other machine learning models. This model delivers balanced and consistent performance across various categories, achieving precision, recall, and F1 score of 83.29%, 83.36%, and 83.73%, respectively. The model's advantage becomes even more apparent when tested in scenarios that require deep understanding and high sensitivity in classifying data. Additionally, it is worth noting that the XGBoost model with GloVe feature extraction also provides positive contributions, particularly in handling complex contexts, with precision, recall, and F1 score values of 75.63%, 78.20%, and 77.60%, respectively.

4.5.1 Statistical Analysis

The Mean (average) value, StDev (standard deviation), and 95% Confidence Interval (CI) for each multiclass classification model are provided. These results offer an overview of the distribution of F1 scores for each model, as well as information on the confidence interval, which encompasses the actual range of the mean value.

The observation results show the average (Mean) F1-score and standard deviation (StDev) for each machine learning model. The XGBoost model has an average F1-score of 77.65%, with a standard deviation of approximately 1.647. The 95% Confidence Interval (95% CI) indicates the range within which the mean F1-score is likely to fall. The Pooled StDev is the combined standard deviation that reflects the overall variability of the data. With this information, one can understand the distribution of the model's performance and make assessments regarding the reliability of the evaluation results.

4.5.2 Discussion

The researcher specifically focuses on the emotional or sentiment aspect contained in the reviews or ratings provided by users for a particular product, service, or experience. To achieve this goal, the first step taken is transforming the numerical rating information into more descriptive sentiment labels. This transformation process involves dividing the rating values into two main sentiment categories: negative and positive. Ratings of 1, 2, and 3 are categorized as negative sentiment, while ratings of 4 and 5 are categorized as positive sentiment. This is based on the assumption that low ratings tend to reflect dissatisfaction or less satisfying experiences, while high ratings generally reflect satisfaction or positive experiences. Table 3 describes the results of the research conducted.

Table 3. Data Distribution Results

No	Category	Rating					Sentiment	
		1	2	3	4	5	Negatif	Positif
1	Decorations and Accessories	7.42	4.45	12.62	21.34	54.17	24.49	75.51
2	Services and Additional Products	45.29	3.77	15.09	15.09	20.76	64.15	35.85
3	Equipment and Storage	9.65	5.54	9.18	21.99	53.64	24.37	75.63

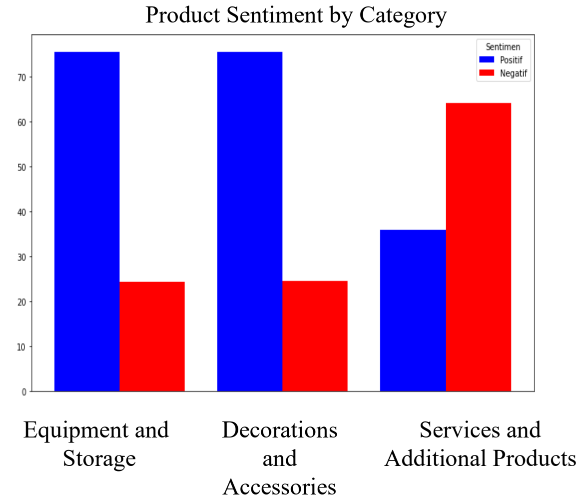

Figure 1. Product Sentiment Percentage by Category

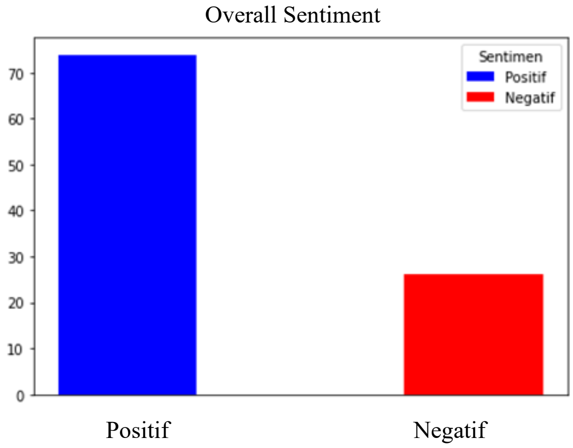

Figure 2. Overall Category Sentiment Percentage

Based on the analysis of customer rating data in three main categories, there are important nuances to understand, which can be seen in Figures 1 and 2. In the "Decorations and Accessories" category, it is observed that positive sentiment peaks at rating 5 (54.17%), indicating significant customer satisfaction with the decoration and accessories aspect. Although there is some negative sentiment in ratings 1-3, the overall trend shows dominant satisfaction. In the "Services and Additional Products" category, negative sentiment dominates at rating 1 (45.29%), which may be an area of focus for improving service quality. However, a positive shift is seen at ratings 4 and 5 (35.85%), suggesting potential to improve customer perception of services and additional products. Meanwhile, in the "Equipment and Storage" category, positive sentiment reaches its peak at rating 5 (53.64%), reflecting high satisfaction with the equipment and storage aspects. While there is slight negative sentiment in ratings 1-3, the high positive percentage indicates success in meeting customer expectations in this category.

From the sentiment analysis results, it is evident that approximately 73.86% of the overall customer rating data indicate positive sentiment towards the evaluated products or services. This reflects the majority of customers who provided favorable or satisfied feedback regarding their experience. However, around 26.14% also show negative sentiment, indicating that there are still certain aspects that may require improvement or further attention.

The evaluation results show that the XGBoost model with TF-IDF features is the most optimal in the sentiment analysis of Bosca Living customer reviews. This model delivers the best performance with evaluation metrics of 83.29% precision, 83.36% recall, and 83.73% F1-score. This advantage makes it the primary choice for understanding customer sentiment and guiding business strategies more accurately. The XGBoost model allows for the high-accuracy identification of negative reviews, which can be used to formulate corrective actions, such as improving service or product quality. Conversely, positive reviews can be leveraged for marketing strategies, highlighting the elements most appreciated by customers, such as the quality of decorations and accessories.

To further understand the underlying causes behind negative ratings, in-depth text analysis can be conducted. For example, reviews with ratings of 1-3 in the "Services and Additional Products" category often mention delays in delivery or inaccurate product descriptions. This suggests that improving logistics and product information transparency could be key strategies for enhancing customer satisfaction. On the other hand, for positive reviews (ratings 4-5), further analysis can help identify the elements most valued by customers. In the "Decorations and Accessories" category, for instance, customers often praise the product quality and ease of installation. This information can be used to highlight these aspects in marketing strategies. By gaining a better understanding of customer sentiment and linking it to specific actions, Bosca Living can improve its response to customer reviews and strengthen their loyalty.

5. Conclusion and Recommendations

Based on the research conducted on e-commerce reviews from the "Bosca Living" store, using Stratified K-Fold Cross Validation with 10-fold cross-validation and testing several models such as K-Nearest Neighbors, Logistic Regression, Random Forest, Support Vector Machine, eXtreme Gradient Boosting, and Decision Tree, with feature extraction using TF-IDF, Word2Vec, FastText, and GloVe, the evaluation results show that the TF-IDF feature with the XGBoost model achieved the best model evaluation with Precision, Recall, and F1-Score of 83.29%, 83.36%, and 83.73%, respectively. However, the dataset and feature extraction used were relatively small, so adding more data and exploring additional features should be done to avoid overfitting. In conclusion, the study recommends using a larger dataset, a more varied train-test data split, and a broader exploration of feature extraction methods to improve model performance and classification accuracy in future research.

References

Ahmad, G. N., Fatima, H., Shafiullah, Salah Saidi, A., & Imdadullah. (2022). Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques with and Without GridSearchCV. *IEEE Access*, *10*. https://doi.org/10.1109/ACCESS.2022.3165792

Almaiah, M. A., Al-Lozi, E. M., Al-Khasawneh, A., Shishakly, R., & Nachouki, M. (2021). Factors affecting students' acceptance of mobile learning application in higher education during covid-19 using ann-sem modelling technique. *Electronics (Switzerland)*, *10*(24). https://doi.org/10.3390/electronics10243121

Bahrawi, N. (2019). Sentiment Analysis Using Random Forest Algorithm-Online Social Media Based. *Journal of Information Technology and Its Utilization*, *2*(2). https://doi.org/10.30818/jitu.2.2.2695

Benarafa, H., Benkhalifa, M., & Akhloufi, M. (2023). An Enhanced SVM Model for Implicit Aspect Identification in Sentiment Analysis. *International Journal of Advanced Computer Science and Applications*, *14*(5). https://doi.org/10.14569/IJACSA.2023.0140505

Bi, J. W., Liu, Y., Fan, Z. P., & Cambria, E. (2019). Modelling customer satisfaction from online reviews using ensemble neural network and effect-based Kano model. *International Journal of Production Research*. https://doi.org/10.1080/00207543.2019.1574989

Demidova, L. A., & Gorchakov, A. V. (2021). Application of chaotic Fish School Search optimization algorithm with exponential step decay in neural network loss function optimization. *Procedia Computer Science*, *186*. https://doi.org/10.1016/j.procs.2021.04.156

Domashova, J. V., & Gultiaev, A. A. (2021). Predicting the revocation of a bank license using machine learning algorithms. *Procedia Computer Science*, *190*. https://doi.org/10.1016/j.procs.2021.06.021

Iyer, K., Ewoorkar, G., Patil, A., Gummani, H., & Parkavi, A. (2019). Efficient Model for Searching and Detecting Semantically Similar Question in Discussion Forums of e-learning Platforms. *2019 4th IEEE International Conference on Recent Trends on Electronics, Information, Communication and Technology, RTEICT 2019 - Proceedings*. https://doi.org/10.1109/RTEICT46194.2019.9016786

Kiedrowsky, F. F., & Andrianingsih. (2023). Sentiment Analysis Marketplaces Digital menggunakan Machine Learning. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, *7*(3). https://doi.org/10.35870/jtik.v7i3.1002

Krishna Das, J., Das, A., & Rosak-Szyrocka, J. (2022). A Hybrid Deep Learning Technique for Sentiment Analysis in E-Learning Platform with Natural Language Processing. *2022 30th International Conference on Software, Telecommunications and Computer Networks, SoftCOM 2022*. https://doi.org/10.23919/SoftCOM55329.2022.9911232

Larose, D. T. (2019). Discovering Knowledge in Data: An Introduction to Data Mining. In *Discovering Knowledge in Data: An Introduction to Data Mining*. https://doi.org/10.1002/0471687545

Muktafin, E. H., Kusrini, K., & Luthfi, E. T. (2020). Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing. *Jurnal Eksplora Informatika*, *10*(1). https://doi.org/10.30864/eksplora.v10i1.390

Nugroho, A., & Eliyani. (2021). Framework to Sell Item in Online Marketplace in Indonesia. *International Journal of Open Information Technologies*, *9*(1).

Onan, A. (2020). Mining opinions from instructor evaluation reviews: A deep learning approach.

Computer Applications in Engineering Education, *28*(1), 117–138. https://doi.org/10.1002/cae.22179

Onan, A. (2021). Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. *Concurrency and Computation: Practice and Experience*, *33*(23). https://doi.org/10.1002/cpe.5909

Sandiwarno, S. (2020). Users' Emotions Analysis based on Hybrid Feature Extraction Techniques. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*. https://doi.org/10.32628/cseit206658

Sandiwarno, S., Niu, Z., & Selemani Nyamawe, A. (2023). A novel hybrid machine learning model for analyzing e-learning users' satisfaction. *International Journal of Human-Computer Interaction*. https://doi.org/10.1080/10447318.2023.2209986

Song, P., & Liu, Y. (2020). An xgboost algorithm for predicting purchasing behaviour on e-commerce platforms. *Tehnicki Vjesnik*, *27*(5). https://doi.org/10.17559/TV-20200808113807

Taufiq, R., Septarini, R. S., Hambali, A., & Yulianti, Y. (2020). Analysis and Design of Decision Support System for Employee Performance Appraisal with Simple Additive Weighting (SAW) Method. *Jurnal Informatika Universitas Pamulang*, *5*(3). https://doi.org/10.32493/informatika.v5i3.6777

Tyagi, A., & Sharma, N. (2018). Sentiment Analysis using logistic regression and effective word score heuristic. *International Journal of Engineering and Technology(UAE)*, *7*(2). https://doi.org/10.14419/ijet.v7i2.24.11991

Wang, T., Lu, K., Chow, K. P., & Zhu, Q. (2020). COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.3012595

Waruwu, M. (2023). Pendekatan Penelitian Pendidikan: Metode Penelitian Kualitatif, Metode Penelitian Kuantitatif dan Metode Penelitian Kombinasi (Mixed Method). *Jurnal Pendidikan Tambusai* , *7*(1).