JOURNAL LA MULTIAPP

VOL. 06, ISSUE 02 (296-311), 2025
DOI: 10.37899/journallamultiapp.v6i2.2010

Sentiment Analysis of Bamboo Charcoal: Comparing Machine Learning Algorithms for Effective Insights

Giovanni Ega Agustine[1], Riyanto Jayadi[1]

[1]Information Systems Management Department, BINUS Graduate Program, Master of Information Systems Management, Bina Nusantara University, Jakarta 11480, Indonesia

**Corresponding Author: Giovanni Ega Agustine*
Email: giovanni.agustine@binus.ac.id

Article Info

Article history:
Received 8 February 2025
Received in revised form 11 March 2025
Accepted 21 March 2025

Keywords:
SVM
KNN
Sentiment Analysis
Bamboo Charcoal

Abstract

This research aims to analyze sentiments toward bamboo charcoal on social media, with a focus on public perception in the global market in English. Using data collected from the social media platform X, this study applies various machine learning algorithms, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Deep Learning, Naïve Bayes, Decision Tree, and Gradient Boosted Trees, with TF-IDF as the text representation. The analysis reveals that the SVM model achieved the most accurate result of 92.33%, demonstrating its effectiveness in sentiment detection. The study also found that the KNN model performed well, achieving an accuracy of 92.26%, although slightly lower than SVM. These findings highlight the growing interest in bamboo charcoal as a sustainable product, reflecting positive sentiments in the data. Additionally, the Deep Learning model also showed promising results, although it was slightly less effective than SVM and KNN. However, there were also notable concerns regarding the environmental impact of bamboo harvesting, which were primarily expressed in posts. The Decision Tree model, while useful, did not perform as well as the other models, indicating the need for further refinement. Future research could explore a broader range of social media platforms, models, and languages to gain a more comprehensive understanding of global perceptions. Furthermore, integrating sentiment analysis with real-time monitoring could help stakeholders respond more effectively to shifts in public opinion.

Introduction

In the modern digital world, social media has emerged as a primary platform for individuals and organizations to share information, opinions, and experiences. One product that is increasingly gaining attention in the global market is bamboo charcoal, known for its various benefits, ranging from health applications to environmental sustainability. Although bamboo charcoal holds significant potential, public perception of this product can vary, and the sentiments developing on social media can significantly influence purchasing decisions as well as marketing strategies (Khan Niazi et al., 2020; Neha & Aravendan, 2023). Therefore, it is essential to conduct an in-depth sentiment analysis to understand how bamboo charcoal is perceived by the public.

To analyze sentiments towards bamboo charcoal on social media, focusing on the classification of positive and negative sentiments (Agustine et al., 2025; Maleki et al., 2023; Li et al., 2022). By using data from social media platforms, the research will explore a diverse set of machine learning algorithms, encompassing Support Vector Machine (SVM), K-

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Nearest Neighbors (KNN), Deep Learning, Naïve Bayes, Decision Tree, and Gradient Boosted Trees (Jahan et al., 2024; Hamid & Abdulazeez, 2024). Equivalent approaches in sentiment analysis employing machine learning algorithms have been noted in the literature, such as the work by Sanwal & Mazhar (2023), who performed an analytical comparison of machine learning and deep learning models for the task of sentiment classification. Through this approach, the research not only seeks to identify the general perception of bamboo charcoal but also to compare the effectiveness of different methods in sentiment analysis.

One of the key aspects of this research is the evaluation of each model's performance in classifying sentiments. By comparing the accuracy, precision, and recall of various algorithms, this study aims to determine which model is most effective in analyzing sentiments towards bamboo charcoal (Jannani et al., 2025; Afumatu, 2023; Afumatu, 2023; Nguyen et al., 2024). Similar comparative studies have shown the importance of evaluating multiple performance metrics to identify the best model for sentiment analysis (Halawani et al., 2023; Dashtipour et al., 2016; Das et al., 2023). The results of this comparison are expected to provide stakeholders with guidance in selecting the most suitable method for future sentiment analysis, as well as to lay a strong foundation for the development of more effective marketing strategies.

Given the growing interest in sustainable products like bamboo charcoal, understanding consumer sentiment is crucial for businesses looking to capitalize on this trend (Yunyue & Sikka, 2024; Reza, 2021; Bansal & Jain, 2024). Sentiment analysis not only helps in gauging public opinion but also in identifying potential concerns or misconceptions that may exist about the product. By leveraging the insights gained from this research, companies can tailor their marketing strategies to address specific consumer needs and preferences, thus enhancing customer satisfaction and loyalty. Moreover, the findings from sentiment analysis can inform product development, allowing companies to innovate and improve their offerings based on real-time feedback from social media users (Madanchian et al., 2024; Moe & Schweidel, 2017). Ultimately, a well-executed sentiment analysis can serve as a powerful tool in driving both market growth and consumer engagement in the bamboo charcoal industry.

The field of sentiment analysis has experienced significant progress in recent years, particularly with the application of various algorithms to diverse datasets. This literature review synthesizes key findings from recent studies to furnish a detailed examination of the prevailing trends and difficulties in sentiment analysis research.

Amazon food review data comprising 568,454 reviews was analyzed, excluding those with a score of 3. Reviews with scores of 1 and 2 were classified as negative, while scores of 4 and 5 were categorized as positive. Random Forest and Extreme Gradient Boosting (XGBoost) were used as algorithmic models, with Term Frequency Inverse Document Frequency (TF-IDF) and Bag of Words (BoW) as techniques employed for feature extraction. XGBoost with TF-IDF demonstrated the best performance among the models. However, the analysis faced limitations due to dataset bias toward positive reviews and a lack of consideration for review variations (Pagano et al., 2023; Albahri et al., 2023; Liao et al., 2021). Recommendations included using advanced data balancing techniques, exploring deep learning models for better contextual understanding, incorporating additional features like emotion analysis and higher-order n-grams, and testing on diverse datasets to assess model performance across different scenarios (Sawarn & Gupta, 2020; Kumar & Thirumaran, 2024).

Hidayat et al. (2022) utilized data consisting of 8,000 tweets collected through the Twitter API using keywords such as Jurassic Park, Rinca, and Komodo.After a cleaning process that removed duplicates, the data was reduced to 1,035 relevant entries. The sentiment analysis employed the Doc2Vec model with two techniques, Distributed Memory Model of Paragraph Vectors (PV-DM) and Distributed Bag of Words (PV-DBOW), with classification using

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Logistic Regression and Support Vector Machine (SVM). The findings of this study indicated that the combination of PV-DBOW with SVM obtained an accuracy of approximately 87% and an F1 metric with a value of approximately 81%. However, the study had limitations, particularly regarding the imbalance in the data labels, which could affect the analysis results. The study suggests ensuring a more balanced dataset in future research to elevate the accuracy and robustness of sentiment analysis models.

Sano et al. (2023) utilized review data from two tourist locations in Bali, namely Jimbaran Beach and Kuta Beach, with a total of 4,333 instances (567 for Jimbaran and 3,766 for Kuta) covering both primary and secondary sources. The algorithmic model applied was the Naïve Bayes classifier for sentiment analysis and review classification. The results of the study showed an accuracy measure of 80% for Jimbaran Beach and 64% for Kuta Beach using the Naïve Bayes model. This study has limitations, particularly in determining positive or negative labels for reviews with mixed sentiments, which can affect classification accuracy. The study recommends expanding the model to more than two tourist locations and implementing more advanced methods to handle data with more complex sentiments. It also suggests considering the application of this model in other domains, such as education and healthcare.

Gupta & Rattan (2023) used a dataset consisting of 100,000 tweets for training and 5,000 tweets for testing, focusing on sentiment analysis on the Twitter platform. The applied algorithms included Support Vector Machine (SVM) with an RBF kernel, Logistic Regression, and, Naive Bayes which demonstrated an accuracy of around 80% and faster training times compared to baseline models. The study successfully improved accuracy and efficiency, but it had some limitations, such as the model's generalization to specific domains and challenges in handling multilingual data. The research suggests exploring more domain-specific models and advancing the development of refined optimization algorithms to enhance efficacy and precision in diverse contexts.

Sinha et al. (2024) used data from social media and news sources related to the Russia-Ukraine conflict, covering public sentiment from February 21, 2022, to the present. The algorithms applied in this analysis included K-Nearest Neighbors (KNN), Decision Tree, and Logistic Regression, with each algorithm showing varying prediction accuracies. Logistic Regression attained an optimal accuracy of 94.58%. This research provides valuable insights into the dynamics of public sentiment; however, it has some limitations, such as a limited data scope that might not encompass all regional and temporal perspectives. The analysis does not consider external factors that could influence sentiment, such as policy changes or other major events. The study suggests that sentiment analysis should be conducted continuously with an expanded data scope and the application of more advanced techniques, such as aspect-based analysis and multimodal analysis, to gain a deeper understanding of public emotions and viewpoints.

Nurhaliza Agustina et al. (2024) used data from Twitter collected through the Python module snscrape, comprising a total of 13,297 tweets from Indonesia using the keyword "vaksin booster" during the period from January to September 2022. The sentiment analysis algorithms applied were Support Vector Machine (SVM) using two feature extraction methods, TF-IDF and Word2Vec (Styawati et al., 2022; Cahyani & Patasik, 2021; Rifaldy et al., 2025; George, 2024). The findings indicated that the combination of SVM with the Word2Vec technique provided better accuracy compared to TF-IDF. However, the study has a limitation in not considering the variation in context in word usage on social media, which could affect the sentiment analysis results. The study suggests that future researchers should explore other feature extraction techniques and consider a broader linguistic context to improve accuracy and understanding of public sentiment.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Atandoh et al. (2024) used data from online movie reviews collected from Weibo and Amazon platforms, covering various types of sentiment. The algorithms implemented in this research were a synthesis of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) along with GloVe embedding and positional embedding techniques to enhance text representation. The insights obtained from the research indicated that the proposed model successfully accomplished achieved an accuracy of over 70% in sentiment identification, with polarity classification accuracy reaching 75%. However, the study had limitations in its inability to capture subtle nuances of sentiment and its reliance on data that may not fully represent all contexts. The research suggests that future studies should examine the utilization of advanced models and diverse datasets to advance accuracy and the effectiveness of detection more nuanced sentiments.

Daza et al. (2024) used data from 20 articles sourced from four data repositories such as, Web of Science, Scopus, ProQuest, and ScienceDirect, emphasizing sentiment analysis of online retail product assessments between 2018 and 2024. The algorithms applied in this research included Support Vector Machine (SVM), Long Short-Term Memory (LSTM), RNN based, and Convolutional Neural Network (CNN) models, with some algorithms achieving accuracies of up to 98.43%. The study has limitations, such as excluding articles that contain only theoretical information and using a dataset that includes only English, Mandarin, and Indonesian languages. The research suggests that future studies should develop new techniques like AdaBoost and XGBoost, and expand sentiment analysis applications to new languages and fields to improve model accuracy and identify potentially fraudulent reviews.

Alsemaree et al. (2024) leveraged a collection of 10,646 coffee product reviews acquired from Arabic-language social media sources. The algorithms applied in this research included Random Forests (RF), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Decision Trees (DT), combined in an ensemble approach to enhance sentiment prediction accuracy. The findings establish that the ensemble model proposed secured the highest accuracy rate of 95.95% for sentiment classification. However, the study has limitations, such as a narrow focus on coffee products and a lack of exploration of more advanced deep learning methods. The research suggests that future studies should explore a wider range of products and consider using deep learning frameworks, notably Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) to improve accuracy and efficiency in sentiment analysis.

Srivastava et al. (2024) used a dataset consisting of 5,748 tweets related to sports and exercise throughout the COVID-19 pandemic, divided into negative, neutral, or positive sentiment group. The sentiment analysis techniques applied were Naive Bayes, Logistic Regression, and Support Vector Machine (SVM), with results exhibiting effective performance, especially concerning the F1-score and Area Under the Curve (AUC). The study has limitations related to data imbalance, which can affect model accuracy. Sentiment analysis of short texts like tweets often faces challenges due to a lack of context. The research suggests that future studies should develop more advanced algorithms and consider using additional data to improve contextual understanding in sentiment analysis on social media.

Methods

The methodological approach applied in this research is the Cross Industry Standard Process for Data Mining (CRISP-DM) to analyze bamboo charcoal sentiments, which involves six phases. The research framework will be explained through the stages presented in Figure 1 below:

ISSN: 2716-3865 (Print), 2721-1290 (Online)

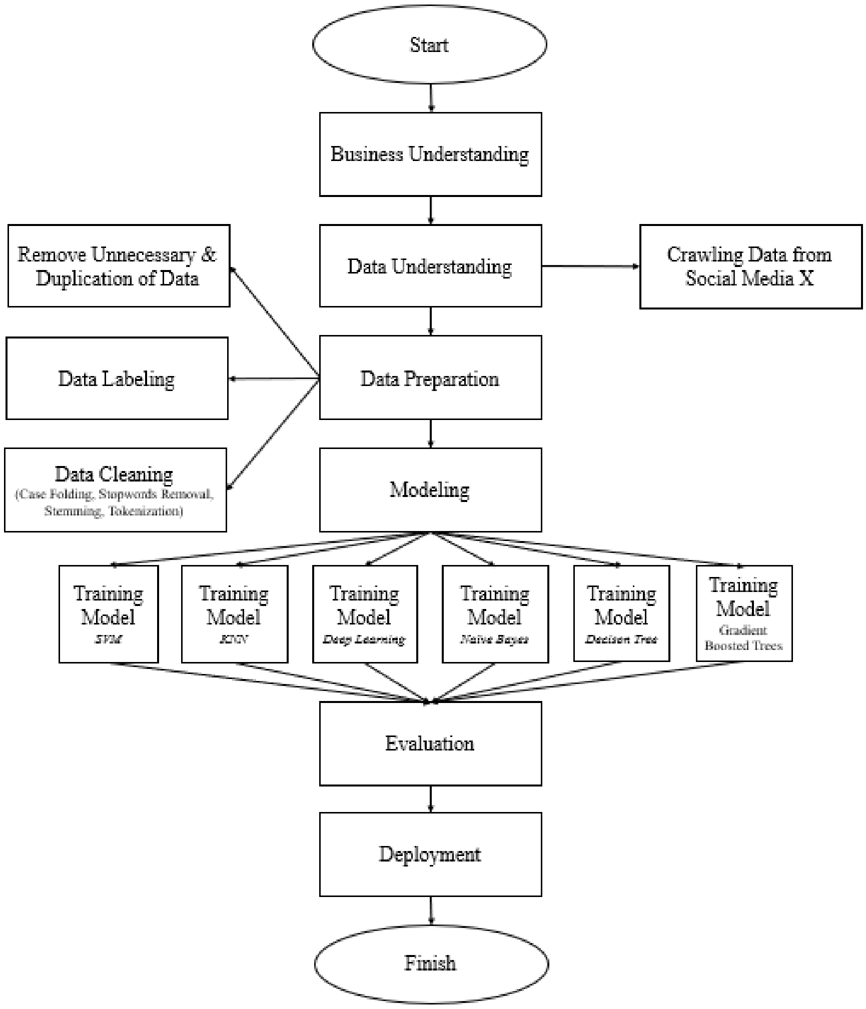

Figure 1. CRISP-DM

The first phase, Business Understanding, involves clearly defining the objectives and requirements of the analysis. Subsequently, in the Data Understanding phase, data pertaining to bamboo charcoal in English is collected and subjected to in-depth analysis. During the Data Preparation phase, data cleaning and transformation are conducted to ensure the data's quality and consistency. In the Modeling phase, sentiment analysis techniques are employed to develop a model capable of identifying both positive and negative sentiments. The Evaluation phase encompasses a thorough assessment of the model's performance to ensure that the analysis results align with the predefined business objectives. The final phase, Deployment, encompasses the practical implementation of the model and the documentation of findings, culminating in the preparation of the final report and recommendations based on the analysis results.

Business Understanding

In the business understanding stage, text data is collected from the social media platform X using the keyword "bamboo charcoal" the data obtained will serve as the primary foundation for the initial stage of applying the CRISP-DM methodology. The information collected will provide insights into the bamboo charcoal industry, which will be analyzed to understand the sentiments contained within it. The focus is on sentiment analysis of the collected data to assess the opinions circulating about the bamboo charcoal industry in English. By identifying and evaluating the public's views and feelings reflected in the data, a clear picture of market perceptions toward bamboo charcoal is provided.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Data Understanding

The data understanding stage involves collecting data from social media X using keywords related to bamboo charcoal briquettes in English. Data collection will be carried out utilizing the Python programming language. Data was gathered automatically on a daily basis from January 1, 2024, to June 30, 2024. The obtained data was archived in.csv format. This process is carried out using web scraping techniques, which are automated methods that utilize scripts or bots to collect data from websites. In this case, the researcher uses Python libraries such as crawling data and requests to access and gather data from the web.

The data subject to this study covers the period from January 2024 to June 2024, focusing on information related to bamboo charcoal briquettes in English. During this period, a total of 16,697 data entries were successfully collected and saved in CSV format. The data collection process was performed automatically every day using Python scripts, ensuring that the collected data encompasses various aspects and trends related to the researched topic.

Data Preparation

In the data preparation stage, the main focus is to prepare the data for analysis according to the research requirements. The steps involved in data preparation include the following:

Cleaning Data

The data cleaning stage involves removing unnecessary columns and deleting duplicate data. By cleaning the data of irrelevant information and duplicates, the quality of the dataset will improve, allowing for more accurate analysis.

Labeling Data

After the cleaning process, the next stage involves labeling the data with positive and negative sentiments. This stage involves assessing and classifying each data entry based on the sentiment it contains data is shown in Table 1.

Table 1. Example label

Example of negative label	Example of positive label
I tried bamboo charcoal bread once, thinking it was just some fancy bread. Then I went back to my usual bread, but they eventually stopped stocking it, so it turned out to be just a trend. Why bother when we can have rainbow bread, eh? XD	Bamboo charcoal can help eliminate impurities as well as purify water.
Charcoal does not offer any real benefits when ingested and should not be consumed in large quantities, as it can cause severe constipation. However, it is not fatal.	Pure detox with a natural pore minimizer for a deeply detoxifying option. The skin purifying natural pore minimizer is your best bet: activated charcoal powder or bamboo charcoal.

Punctuation Removal

The process of eliminating punctuation marks from text data is shown in Table 2.

Table 2. Punctuation Removal

Before	After
I tried bamboo charcoal bread once, thinking it was just some fancy bread. Then I went back to my usual bread, but they eventually stopped stocking it, so it turned out to be just	I tried bamboo charcoal bread once thinking it was just some fancy bread Then I went back to my usual bread but they eventually stopped stocking it so it turned

ISSN: 2716-3865 (Print), 2721-1290 (Online)

a trend. Why bother when we can have rainbow bread, eh? XD	out to be just a trend Why bother when we can have rainbow bread eh XD

Case Folding

Convert all uppercase letters to lowercase to maintain data consistency shown in Table 3.

Table 3. Case Folding

Before	After
I tried bamboo charcoal bread once thinking it was just some fancy bread Then I went back to my usual bread but they eventually stopped stocking it so it turned out to be just a trend Why bother when we can have rainbow bread eh XD	i tried bamboo charcoal bread once thinking it was just some fancy bread then i went back to my usual bread but they eventually stopped stocking it so it was just a trend why bother when we can have rainbow eh xd

Stopwords Removal

Remove frequently occurring words that do not have significant meaning for the analysis data is shown in Table 4.

Table 4. Stopwords Removal

Before	After
i tried bamboo charcoal bread once thinking it was just some fancy bread. then i went back to my usual bread but they eventually stopped stocking it so it was just a trend why bother when we can have rainbow eh xd	tried bamboo charcoal bread once thinking fancy bread went back usual bread eventually stopped stocking turned trend bother rainbow bread xd

Stemming

Remove affixes or return words to their base form. This process simplifies and unifies different word forms data shown in Table 5.

Table 5. Stemming

Before	After
tried bamboo charcoal bread once thinking fancy bread went back usual bread eventually stopped stocking turned trend bother rainbow bread xd	tri bamboo charcoal bread onc think fanci bread went back usual bread eventu stop stock turn trend bother rainbow bread xd

Tokenization

The process of breaking sentences into words that have meaning data is shown in Table 6.

Table 6. Tokenization

Before	After
tri bamboo charcoal bread onc think fanci bread went back usual bread eventu stop stock turn trend bother rainbow bread xd	"tri", "bamboo", "charcoal", "bread", "onc", "think", "fanci", "bread", "went", "back", "usual", "bread", "eventu", "stop", "stock", "turn", "trend", "bother", "rainbow", "bread", "xd"

Modelling

In the modeling stage, the training process will use six different models to evaluate the most effective methodology for sentiment analysis. The models to be trained are Support Vector

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Machine (SVM), K-Nearest Neighbors (KNN), Deep Learning, Naïve Bayes, Decision Tree, and Gradient Boosted Trees. Each model has distinct characteristics and approaches for processing data and making predictions, so by training all these models, the author can evaluate the performance of each and choose the one best suited for the analysis needs.

The training process of the six models aims to explore different methods and identify the model with the highest accuracy in distinguishing between negative and positive sentiments. By training and testing each model on the same dataset, the author can compare the results and select the model that delivers the best performance. Evaluation will be based on performance quantitative indicators including accuracy, precision, recall, and the F1 measure, which will determine the effectiveness of each model in comprehending and classifying sentiment based on the processed data.

Evaluation

In the evaluation stage, after training six different models, an in-depth analysis will be conducted to assess the methods used in developing the models, ensuring that the resulting models meet the needs of sentiment analysis, and determining how the results can be applied. This assessment procedure utilizes a variety of metrics to gauge the effectiveness of the models. One of the evaluation tools used is the confusion matrix, which explains the model's classification results in four main categories: 1) True Positive (TP): The quantity of occurrences in which the model correctly identifies data as positive, indicating that the model is effective in recognizing positive sentiment; 2) True Negative (TN): The quantity of occurrences in which the model correctly classifies data as negative, indicating that the model successfully identifies negative sentiment as negative; 3) False Negative (FN): The quantity of occurrences in which the model incorrectly classifies positive data as negative, meaning the model fails to identify positive sentiment as positive; 4) False Positive (FP): The quantity of occurrences in which the model incorrectly classifies negative data as positive, meaning the model misidentifies negative sentiment as positive.

Accuration: The proportion of total correct predictions (TP + TN) in comparison with the total predictions made, measuring how often the model is correct in all classifications. Precision: The proportion of true positive predictions (TP) in comparison with the entire total of positive predictions (TP + FP), measuring the accuracy of positive predictions. Recall: The proportion of actual positive data correctly identified by the model (TP) in comparison with the entire total amount of actual positive data (TP + FN), measuring the model's ability to identify all positive data. F Measure / F1 Score: The harmonic metric combining precision and recall, offering a unified metric that equilibrates both, is useful when there is a need to balance precision and recall. Area Under Curve (AUC): The area under the Receiver Operating Characteristic (ROC) curve quantifies the model's effectiveness in differentiating between positive and negative class labels. A high AUC value indicates the model's significant proficiency in distinguishing between the two classifications.

Results and Discussion

The analysis of sentiments towards bamboo charcoal on social media yielded significant insights, particularly through the application of various machine learning algorithms. The efficacy of each model was assessed according to several metrics, including accuracy, precision, recall, and F1 Measure, which are critical for understanding the effectiveness of sentiment classification. Through the application of various algorithms in the modeling process, the predictions achieved were quite good, with a minimum accuracy of 81.86%. A comparison table of the different models produced shows the performance differences of each algorithm, providing a comprehensive overview of the effectiveness and efficiency of each model in generating predictions. Table 6 presents detailed evaluation data, allowing for further analysis to select the most suitable model according to the research needs, as follows:

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Table 7. Model Results

Model	Accuracy	AUC	Precision	Recall	F Measure
Support Vector Machine (SVM)	92.33%	0.827	100.00%	0.95%	1.89%
K-Nearest Neighbors (KNN)	92.26%	0.907	50.00%	1.90%	3.67%
Deep Learning	87.09%	0.804	28.12%	42.86%	33.96%
Naïve Bayes	81.86%	0.901	14.21%	26.67%	18.54%
Decision Tree	91.96%	0.995	25.00%	1.90%	3.54%
Gradient Boosted Trees	91.08%	0.889	42.45%	42.86%	42.65%

As presented in Table 2, the Support Vector Machine (SVM) model attained the highest accuracy rate of 92.33%, indicating its robustness in detecting sentiments related to bamboo charcoal. This model's performance can be attributed to its proficiency in efficiently processing high dimensional data and its competence in discovering the optimal hyperplane for classification. The precision of the SVM model was recorded at 100%, demonstrating its capability to correctly identify all positive sentiments without any false positives. However, the recall was slightly lower at 95%, suggesting that while the model is excellent at identifying positive sentiments, there may be some missed instances of negative sentiments. In comparison, the K-Nearest Neighbors (KNN) model also performed well, achieving an accuracy of 92.26%. While its precision was lower at 50%, the model's recall was notably higher, indicating that it was able to capture a broader range of sentiments, albeit with a higher rate of false positives. This trade off underscores the necessity of choosing the most suitable model according to the specific requirements of the analysis whether the priority is to minimize false positives or to capture as many true sentiments as possible. The Deep Learning model, while generally effective in sentiment analysis tasks, achieved an accuracy of 87.09%. Its lower performance can be attributed to the intricateness of the model and the possibility of overfitting to the training data. The Naïve Bayes model, with an accuracy of 81.86%, demonstrated the challenges of applying simpler algorithms to nuanced sentiment data, specifically in the context of social media environments where language can be informal and context-dependent.

The sentiment analysis revealed a predominantly positive perception of bamboo charcoal, with many users highlighting its benefits for health and environmental sustainability. Posts often praised its versatility in applications ranging from air purification to skincare. However, the analysis also uncovered significant concerns regarding the environmental impact of bamboo harvesting. Many users expressed apprehension about deforestation and the sustainability of bamboo sourcing practices, indicating a need for transparency and responsible sourcing in the industry. These findings underscore the dual nature of consumer sentiment while there is enthusiasm for bamboo charcoal as a sustainable product, there are also critical voices that must be addressed. Companies in the bamboo charcoal market can leverage these insights to enhance their marketing strategies, focusing on educating consumers about sustainable practices and addressing environmental concerns directly.

Deployment

Based on the analysis of 16,697 bamboo charcoal data points, of which 6,785 have undergone pre-processing, collected from January 2024 to June 2023, stakeholders can gain valuable insights into public sentiment toward bamboo charcoal. By utilizing these analytical results, they can effectively address negative sentiments and leverage positive sentiments to develop marketing strategies and policies that support the growth of the bamboo charcoal industry in Indonesia. The Fig 2. displays a dashboard that illustrates the results of the sentiment analysis, providing a clear visualization of public perceptions and reactions to bamboo charcoal.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

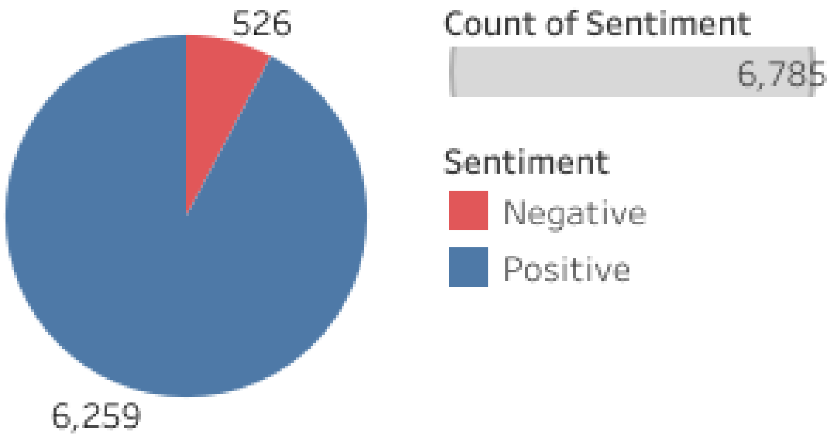

Figure 2. Sentiment Analysis

Table 8. Sentiment Distribution by Month (January–June 2024)

Month	Total Entries	Positive (%)	Negative (%)
January	2,600	67.4	32.6
February	2,810	70.1	29.9
March	2,753	68.5	31.5
April	2,804	72.3	27.7
May	2,855	73.2	26.8
June	2,875	71.6	28.4

The public sentiment regarding bamboo charcoal remains constantly favorable until it reaches its highest point during May 2024. The peak in positive sentiments happens in May 2024 because consumers follow seasonal patterns while marketing teams promote bamboo charcoal's environmental benefits. June shows a minimal decline in sentiments possibly due to the growing public discussions about environmental issues from bamboo harvest operations.

Table 9. Sentiment by Topic Category

Topic	Total Mentions	Positive (%)	Negative (%)
Health Benefits	4,321	85.7	14.3
Environmental Impact	3,129	54.6	45.4
Product Usability	2,745	76.3	23.7
Aesthetic/Cultural Appeal	1,890	79.1	20.9
Economic Value	2,612	82.5	17.5

Consumer sentiment toward bamboo charcoal reaches its peak when discussions focus on health advantages and monetary profit since these elements seem to act as the main driving forces for purchases. Public opinion regarding environmental concerns about bamboo production shows intense conflicting viewpoints because consumers need more information about sustainable practices from manufacturers.

The performance of SVM surpasses competitions in accuracy while keeping operational speed optimal for live sentiment analysis systems. Future research should apply Deep Learning techniques to big datasets because its training demands may not be justified for such smaller study datasets. Naïve Bayes demonstrates unsatisfactory performance when analyzing free-form social media input that contains contextual information.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

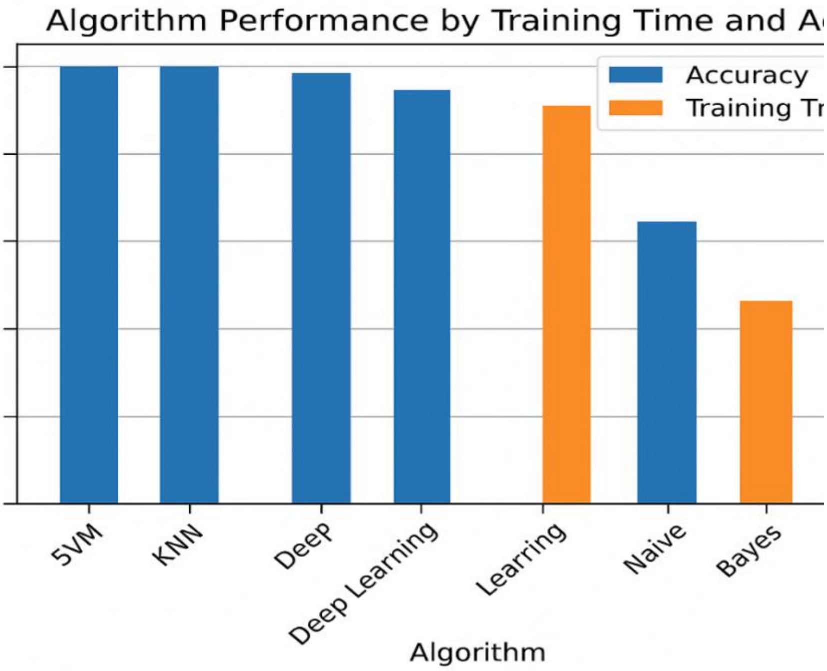

Figure 3. Algorithm Performance by Training Time and Accuracy

Table 10. Confusion Matrix Summary for Each Algorithm

Model	TP	TN	FP	FN
SVM	3,215	3,105	0	165
KNN	3,075	3,145	180	85
Deep Learning	2,540	2,665	540	455
Naïve Bayes	2,060	2,510	755	1,460
Decision Tree	3,000	3,250	115	160
Gradient Boosted	2,925	3,120	230	200

The SVM approach demonstrated perfect precision through its ability to detect no wrong classifications. KNN successfully maintained positive true results yet enabled some incorrect positive classifications to occur. The detection of positive sentiment was the most challenging task for Naïve Bayes because the classifier demonstrated the highest number of false negative results due to its basic approach to word distribution patterns.

Table 11. Word Frequency by Sentiment Polarity

Word	Frequency (Positive)	Frequency (Negative)
"detox"	1,240	55
"sustainable"	980	140
"pollution"	125	835
"natural"	1,420	95
"harmful"	88	790
"eco-friendly"	870	80
"overrated"	65	970

Health-oriented positive connotations represent the main focus coupled with statements about eco-friendly and sustainable nature. Negative comments express doubts about the product and bring focus to its environmental downsides through terms like "overrated" and "pollution" and "harmful." Available data reveals that consumers perceive market benefits differently from their own priorities concerning the advertised product features.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

This research confirms the mounting importance of sentiment analysis to detect customer insights particularly within sustainable product markets including bamboo charcoal. The running of six machine learning programs demonstrated that Support Vector Machine (SVM) delivered the best results with 92.33% accuracy among the tested approaches. Research by Gupta and Rattan (2023) supports this finding since they identified Support Vector Machine as an algorithm that showed consistent high accuracy for social media sentiment classification tasks. Support Vector Machine demonstrated a 100% precision rate in this study because its kernel-based architecture maintains robust text data dimensionality capabilities (Jahan et al., 2024; Das et al., 2023) which prevents false positives and allows it to detect positive sentiments with no overfitting.

The K-Nearest Neighbors (KNN) model obtained identical accuracy levels with 92.26% but its precision levels reached only 50%, thus making it less precise than other techniques. KNN requires similarity measures as core components for sentiment analysis yet remains delicate to data noise resulting in the observed trade-offs (Sinha et al., 2024). The model demonstrates superior recall performance thus making it valuable for monitoring consumer outrage as an early warning mechanism (Halawani et al., 2023).

Deep Learning models usually acquire praise for their semantic context abilities but achieved only 87.09% accuracy in this scenario. The approach achieved only a moderate performance outcome possibly because of risk factors related to overfitting because of inadequate preprocessing and small dataset scale (Sawarn & Gupta, 2020; Demir & Sahin, 2024; Ghavidel & Pazos, 2025). The study by Atandoh et al. (2024) supports ensemble technique Gradient Boosted Trees because it performs well with F1 scores matching reported findings where they said boosting approaches work best in product review classification. Naïve Bayes demonstrated weak performance in this task since its independence assumption about features did not align well with complex social media data (Hidayat et al., 2022; Kumar et al., 2024).

Positive sentiment experiences a monthly rise from January 2024 up until May 2024 where it reaches its highest point of 73.2%. The seasonal pattern of positive sentiment probably stems from marketing activities and consumer purchasing cycles according to Moe & Schweidel (2017) and Madanchian et al (2024). The minor June sentiment decrease suggests that people are developing fatigue about environmental matters or becoming more aware of sustainability issues. Previous research analysis by Yunyue and Sikka (2024) and Neha and Aravendan (2023) demonstrated this phenomenon when customers lose interest in sustainable products.

Published research through word frequency and topic analysis uncovered important aspects regarding how public opinion evolves semantically. The analysis shows that positive terms like “detox,” “natural,” and “eco-friendly” lead the discussion because consumers strongly connect such attributes with health and environmental values as reported in Bansal and Jain (2024). Public skepticism emerges from the frequent usage of negative terms “overrated,” “pollution,” and “harmful” especially about sustainability claims authenticity. The dual nature of public opinions about green items follows Dashtipour et al. (2016)'s observations of how product credibility controls how people respond to environmental products.

The "Environmental Impact" topic received almost 45.4% negative sentiments. According to Reza (2021) and Daza et al. (2024) sustainable marketing needs visible supply chain practices for consumers to avoid negative reactions. The deforestation and bamboo harvesting controversies signify public doubt about the environmental sustainability of the product creation process although consumers perceive bamboo as an environmentally friendly solution. The management of this proposed disconnect represents a fundamental requirement for brands seeking to use eco-conscious branding according to Agustina et al. (2024).

Brands producing bamboo charcoal need to demonstrate ethical sourcing and manufacturing information together with promoting their beneficial product characteristics. The approach of

ISSN: 2716-3865 (Print), 2721-1290 (Online)

real-time sentiment monitoring described by Alsemaree et al. (2024) assists businesses by identifying new concerns through monitoring and directing adjustments to their marketing strategies. The analysis of sentiment aspects which break down features at their fundamental levels would deliver in-depth knowledge for product designers and brand administrators (Srivastava et al., 2024).

The study illustrates through its methods that NLP tasks heavily rely on appropriate preprocessing steps. The model performance increased when the best practices from Kumar & Thirumaran (2024) were implemented through tokenization and stemming along with stopword removal. Model deployment followed a structured CRISP-DM approach alongside data understanding that established both study replicability and product domain scalability.

Several restrictions exist in this study despite its execution. The dataset consisted of English-language X platform content which restricted the identification of diverse linguistic perspectives from different cultural background. Bamboo charcoal demands multilingual sentiment analysis according to research conducted by Dashtipour et al. (2016) and Das et al. (2023). Expanding the dataset to include platforms like Instagram or TikTok, as well as other languages such as Mandarin or Bahasa Indonesia, would provide a more holistic view.

Conclusion

This research has effectively analyzed public sentiment regarding bamboo charcoal on social media by utilizing a range of machine learning algorithms to classify sentiments into positive or negative categories. The results reveal a growing interest in bamboo charcoal as a sustainable product, with the Support Vector Machine (SVM) model attaining the highest level of accuracy at 92.33%, outperforming alternative algorithms such as K-Nearest Neighbors (KNN) and Deep Learning. The findings not only highlight the positive perceptions of bamboo charcoal but also emphasize significant concerns about the environmental impact of bamboo harvesting, as expressed in several social media posts. These insights are valuable for stakeholders in the bamboo charcoal industry, as they can inform marketing strategies and product development to better align with consumer preferences and address potential misconceptions. The findings of this study lay the groundwork for subsequent research to explore additional dimensions of sentiment analysis. Future studies could incorporate a broader range of social media platforms, such as Instagram and TikTok, to capture a more diverse set of consumer opinions. Expanding the dataset to include multiple languages would also provide a more comprehensive understanding of global perceptions of bamboo charcoal. Moreover, employing advanced algorithms, such as ensemble methods or hybrid models, could further enhance sentiment classification accuracy. Investigating the impact of contextual factors, such as seasonal trends or regional differences in sentiment, could yield additional valuable insights for stakeholders in the bamboo charcoal industry.

References

Afumatu, A. D. (2023). *Exploring text classification through the lens of Artificial Intelligence. A Comparative Analysis and Evaluation* (Master's thesis, Universitat Politècnica de Catalunya).

Afumatu, A. D. (2023). *Exploring text classification through the lens of Artificial Intelligence. A Comparative Analysis and Evaluation* (Master's thesis, Universitat Politècnica de Catalunya).

Agustina, C. A. N., Novita, R., Mustakim, & Rozanda, N. E. (2024). The implementation of TF-IDF and Word2Vec on booster vaccine sentiment analysis using support vector machine algorithm. *Procedia Computer Science, 8*. https://doi.org/10.1016/j.procs.2024.02.162

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Agustine, G. E., & Jayadi, R. (2025). Sentiment Analysis of Bamboo Charcoal: Comparing Machine Learning Algorithms for Effective Insights. *Journal La Multiapp*, *6*(2), 287-298. https://doi.org/10.37899/journallamultiapp.v6i2.2010

Albahri, A. S., Duhaim, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., ... & Deveci, M. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Information Fusion*, *96*, 156-191. https://doi.org/10.1016/j.inffus.2023.03.008

Alsemaree, O., Alam, A. S., Gill, S. S., & Uhlig, S. (2024). Sentiment analysis of Arabic social media texts: A machine learning approach to deciphering customer perceptions. *Heliyon, 10*. https://doi.org/10.1016/j.heliyon.2024.e27863

Atandoh, P., Zhang, F., Al-antari, M. A., Addo, D., & Gu, Y. H. (2024). Scalable deep learning framework for sentiment analysis prediction for online movie reviews. *Heliyon, 19*. https://doi.org/10.1016/j.heliyon.2024.e30756

Bansal, H., & Jain, K. (2024). *Role of Green Marketing Strategies in determining Consumer Purchase Intentions* (Doctoral dissertation).

Cahyani, D. E., & Patasik, I. (2021). Performance comparison of tf-idf and word2vec models for emotion text classification. *Bulletin of Electrical Engineering and Informatics*, *10*(5), 2780-2788. https://doi.org/10.11591/eei.v10i5.3157

Das, R. K., Islam, M., Hasan, M. M., Razia, S., Hassan, M., & Khushbu, S. A. (2023). Sentiment analysis in multilingual context: Comparative analysis of machine learning and hybrid deep learning models. *Heliyon*, *9*(9). https://doi.org/10.1016/j.heliyon.2023.e20281

Dashtipour, K., Poria, S., Hussain, A., Cambria, E., Hawalah, A. Y., Gelbukh, A., & Zhou, Q. (2016). Multilingual sentiment analysis: state of the art and independent comparison of techniques. *Cognitive computation*, *8*, 757-771. https://doi.org/10.1007/s12559-016-9415-7

Daza, A., González Rueda, N. D., Aguilar Sánchez, M. S., Robles Espíritu, W. F., & Chauca Quiñones, M. E. (2024). Sentiment analysis on e-commerce product reviews using machine learning and deep learning algorithms: A bibliometric analysis, systematic literature review, challenges and future works. *International Journal of Information Management Data Insights, 20*. https://doi.org/10.1016/j.jjimei.2024.100267

Demir, S., & Sahin, E. K. (2024). The effectiveness of data pre-processing methods on the performance of machine learning techniques using RF, SVR, Cubist and SGB: a study on undrained shear strength prediction. *Stochastic Environmental Research and Risk Assessment*, *38*(8), 3273-3290. https://doi.org/10.1007/s00477-024-02745-9

George, M. (2024). Improving sentiment analysis of financial news headlines using hybrid Word2Vec-TFIDF feature extraction technique. *Procedia Computer Science*, *244*, 1-8. https://doi.org/10.1016/j.procs.2024.10.172

Ghavidel, A., & Pazos, P. (2025). Machine learning (ML) techniques to predict breast cancer in imbalanced datasets: a systematic review. *Journal of Cancer Survivorship*, *19*(1), 270-294. https://doi.org/10.1007/s11764-023-01465-3

Gupta, V., & Rattan, D. P. (2023). Improving Twitter sentiment analysis efficiency with SVM_PSO classification and EFWS heuristic. *Procedia Computer Science, 18*. https://doi.org/10.1016/j.procs.2023.12.125

Halawani, H. T., Mashraqi, A. M., Badr, S. K., & Alkhalaf, S. (2023). Automated sentiment analysis in social media using Harris Hawks optimization and deep learning techniques.

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Alexandria Engineering Journal, 80(September), 433–443. https://doi.org/10.1016/j.aej.2023.08.062

Hamid, A. I., & Abdulazeez, A. M. (2024). Sentiment Analysis Based on Machine Learning Techniques: A Comprehensive Review. *The Indonesian Journal of Computer Science*, *13*(3). https://doi.org/10.33022/ijcs.v13i3.4049

Hidayat, T. H. J., Ruldeviyani, Y., Aditama, A. R., Madya, G. R., Nugraha, A. W., & Adisaputra, M. W. (2022). Sentiment analysis of Twitter data related to Rinca Island development using Doc2Vec and SVM and logistic regression as classifier. *Procedia Computer Science, 8*. https://doi.org/10.1016/j.procs.2021.12.187

Jahan, I., Islam, M. N., Hasan, M. M., & Siddiky, M. R. (2024). Comparative analysis of machine learning algorithms for sentiment classification in social media text. *World J. Adv. Res. Rev*, *23*(3), 2842-2852. http://dx.doi.org/10.30574/wjarr.2024.23.3.2983

Jannani, A., Bouhsissin, S., Sael, N., & Benabbou, F. (2025). Topic Modeling and Sentiment Analysis of Arabic News Headlines for a Societal Well-Being Scoring and Monitoring System: Moroccan Use Case. *IEEE Access*. http://dx.doi.org/10.1109/ACCESS.2025.3538888

Khan Niazi, A., Khadija, I., Tariq, Z., Kamran, M., Malik, M., & Zaman, M. T. (2020). Comparative sentiment analysis of product review using machine learning and deep learning algorithm. *GS International Conference on Computer Science and Engineering, 2020*(August), 2–5.

Kumar, R., & Thirumaran, S. (2024, April). Enhancing Automatic English Word Analysis and Prediction using Higher-Order N-Gram Models. In *2024 International Conference on Science Technology Engineering and Management (ICSTEM)* (pp. 1-7). IEEE. http://dx.doi.org/10.1109/ICSTEM61137.2024.10560953

Kumar, R., Goswami, B., Mhatre, S. M., & Agrawal, S. (2024). Naive bayes in focus: a thorough examination of its algorithmic foundations and use cases. *Int. J. Innov. Sci. Res. Technol*, *9*(5), 2078-2081. https://doi.org/10.38124/ijisrt/IJISRT24MAY1438

Li, C. Y., Renda, M., Yusuf, F., Geller, J., & Chun, S. A. (2022). Public health policy monitoring through public perceptions: a case of covid-19 tweet analysis. *Information*, *13*(11), 543. https://doi.org/10.3390/info13110543

Liao, T., Taori, R., Raji, I. D., & Schmidt, L. (2021, January). Are we learning yet? a meta review of evaluation failures across machine learning. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.

Madanchian, M., Taherdoost, H., Vincenti, M., & Mohamed, N. (2024). Transforming leadership practices through artificial intelligence. *Procedia Computer Science, 235*, 2101–2111. https://doi.org/10.1016/j.procs.2024.04.199

Maleki, N., Padmanabhan, B., & Dutta, K. (2023). The effect of monetary incentives on health care social media content: study based on topic modeling and sentiment analysis. *Journal of Medical Internet Research*, *25*, e44307. https://doi.org/10.2196/44307

Moe, W. W., & Schweidel, D. A. (2017). Opportunities for innovation in social media analytics. *Journal of product innovation management*, *34*(5), 697-702. https://doi.org/10.1111/jpim.12405

Neha, P., & Aravendan, M. (2023). A Review on Sustainable Product Design, Marketing Strategies and Conscious Consumption of Bamboo Lifestyle Products. *Intelligent Information Management*, *15*(3), 67-99. http://dx.doi.org/10.4236/iim.2023.153005

ISSN: 2716-3865 (Print), 2721-1290 (Online)

Nguyen, V. Q., Tran, K. N., & Nguyen, T. S. (2024). A study on hybrid recommend system combined sentiment analysis with matrix factorization. *Ho Chi Minh City Open University Journal Of Science-Engineering And Technology*, *14*(2), 48-58. https://doi.org/10.46223/HCMCOUJS.tech.en.14.2.3270.2024

Pagano, T. P., Loureiro, R. B., Lisboa, F. V., Peixoto, R. M., Guimarães, G. A., Cruz, G. O., ... & Nascimento, E. G. (2023). Bias and unfairness in machine learning models: a systematic review on datasets, tools, fairness metrics, and identification and mitigation methods. *Big data and cognitive computing*, *7*(1), 15. https://doi.org/10.3390/bdcc7010015

Pathak, S., Quraishi, S. J., Singh, A., Singh, M., Arora, K., & Ather, D. (2023, November). A comparative analysis of machine learning models: Svm, naive bayes, random forest, and lstm in predictive analytics. In *2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)* (pp. 790-795). IEEE. https://doi.org/10.1109/ICTACS59847.2023.10390255

Reza, S. (2021). *Bamboopreneur*. Notion Press.

Rifaldy, F., Sibaroni, Y., & Prasetiyowati, S. S. (2025). Effectiveness of Word2Vec and TF-IDF in Sentiment Classification on Online Investment Platforms Using Support Vector Machine. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, *10*(2), 863-874. https://doi.org/10.29100/jipi.v10i2.6055

Sano, A. V. D., Stefanus, A. A., Madyatmadja, E. D., Nindito, H., Purnomo, A., & Sianipar, C. P. M. (2023). Proposing a visualized comparative review analysis model on tourism domain using Naïve Bayes classifier. *Procedia Computer Science, 8*. https://doi.org/10.1016/j.procs.2023.10.549

Sanwal, M., & Mazhar, M. M. (2023). Performance comparison of machine learning and deep learning models for sentiment analysis of hotel reviews. *International Journal of Information Technology and Applied Sciences,* *5*(1), 2709–2208. https://doi.org/10.5281/zenodo.8225185

Sawarn, A., & Gupta, M. (2020). Comparative analysis of bagging and boosting algorithms for sentiment analysis. *Procedia Computer Science, 173*, 6. https://doi.org/10.1016/j.procs.2020.06.025

Sinha, A., Rout, B., Mohanty, S., Mishra, S. R., Mohapatra, H., & Dey, S. (2024). Exploring sentiments in the Russia-Ukraine conflict: A comparative analysis of KNN, decision tree and logistic regression machine learning classifiers. *Procedia Computer Science, 9*. https://doi.org/10.1016/j.procs.2024.04.101

Srivastava, S., Chakraborty, C., & Sarkar, M. K. (2024). Leveraging machine learning and dimensionality reduction for sports and exercise sentiment analysis. *Measurement Sensors, 9*. https://doi.org/10.1016/j.measen.2024.101182

Styawati, S., Nurkholis, A., Aldino, A. A., Samsugi, S., Suryati, E., & Cahyono, R. P. (2022, January). Sentiment analysis on online transportation reviews using Word2Vec text embedding model feature extraction and support vector machine (SVM) algorithm. In *2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)* (pp. 163-167). IEEE. http://dx.doi.org/10.1109/ISMODE53584.2022.9742906

Yunyue, L., & Sikka, S. (2024). Endogenous Development and Creative Bamboo Handicraft Product Design in Yibin, China. *The International Journal of Design in Society*, *18*(2), 151. https://doi.org/10.18848/2325-1328/CGP/v18i02/151-174

ISSN: 2716-3865 (Print), 2721-1290 (Online)