

Exploring Trends and Techniques in Sentiment Analysis for Online Product Ratings: A Comprehensive Review

Mary Rose Columbres

Information Technology and Data Science Department, Bulacan State University

DOI: <https://doi.org/10.51244/IJRSI.2025.12020027>

Received: 21 January 2025; Revised: 29 January 2025; Accepted: 31 January 2025; Published: 04 March 2025

ABSTRACT

The COVID-19 epidemic has led to a notable rise in the dependence of product purchases on social media and online shopping. Before the pandemic, customers preferred to shop in person to evaluate the quality of the products personally. However, the pandemic forced people to buy things online, which forced companies to use social media and e-commerce sites to conduct business. Due to this change, client testimonials, remarks, and evaluations are now vital for customers and companies. Consumers rely on these reviews to establish confidence, while companies examine them to improve their products and competitive tactics. This systematic review, which focuses on data mining techniques and semantic analysis, attempts to identify and assess the different approaches utilized in sentiment analysis of online product evaluations. Thirty published research papers and journals were content-analyzed as part of a qualitative research strategy.

Keywords: Sentiment Analysis, Semantic Analysis, Data Mining, e-Commerce Site, Social Media Site

INTRODUCTION

Technology is now everywhere. Most people use the internet to view and order what products they want or need. Most people rely on comments, feedback, and the rate of the product to avoid trust issues. A vast amount of information from all the platforms that have comments and feedback was collected and used by all the companies or organizations to improve their service, offer products, and implement new strategies that help them compete with other competitors.

According to **Bernard Marr of Enterprise Tech**, “In 2019, there are 2.5 quintillion bytes of data created each day at our current pace, but that pace is only accelerating with the growth of the Internet of Things (IoT). Over the last two years alone, 90 percent of the data in the world was generated. More than 3.7 billion humans use the internet (a growth rate of 7.5 percent over 2016). On average, Google now processes more than 40,000 searches EVERY second (3.5 billion searches per day). While 77% of searches are conducted on Google, not remembering other search engines also contribute to our daily data generation would be inconsistent. Worldwide there are 5 billion searches a day.”

According to **Simon Kemp of DigitalReportal**, “76.01 million internet users in the Philippines in January 2022. The Philippines' internet penetration rate stood at 68.0 percent of the total population in 2022. Kepios analysis indicates that internet users in the Philippines increased by 2.1 million (+2.8 percent) between 2021 and 2022.” according to Simon Kemp, in January 2022, 92.05 million Filipinos were using social media. As of the beginning of 2022, 82.4 percent of Filipinos were active on social media. However, it's crucial to remember that social media users cannot all be distinct individuals.

Online reviews are now an essential part of the buying process because of the shift in customer behavior brought about by the COVID-19 epidemic and the growth of e-commerce. This research aims to close the

knowledge gap in the field by examining data mining methods for sentiment analysis in online product reviews. It also seeks to draw attention to the best methods for resolving the issues in this field.

Table 1. Common social media used for a product review with their statistics in the Philippines in early 2022.

Social Media	Number of Users
Facebook	83.85 million
Instagram	18.65 million
Snapchat	10.60 million
Twitter	10.50 million

According to worldpopulationreview.com, the population in the Philippines is 112,233,339. According to digitalreportal.com, 82.44% of the people of the Philippines are social media users. Still, we must consider that these social media users may not represent unique account owners.

Table 2. Common e-commerce Sites with their statistics in early 2022.

e-Commerce Site	Monthly Traffic Estimate
Lazada	43.38M
Shopee	74.91M
Metrodeal	770K
eBay	277.65K
FB Marketplace	800M

The table shown above is from magenest.com; the table shows that most Filipinos have been transacting online. Due to the COVID-19 Pandemic, the rate of online transactions has increased. With this, bulk data is now available for companies to collect comments and suggestions, analyze, and develop business strategies and data-driven decisions, particularly regarding product reviews.

According to getthematic.com, *Sentiment* analysis helps to pinpoint the feelings expressed in a text. It is frequently used to examine product reviews, survey results, and consumer feedback. Sentiment analysis has applications in customer experience, reputation management, and social media monitoring, to name a few. Furthermore, sentiment analysis is used to analyze thousands of product reviews coming from different social media and e-commerce platforms. It can generate helpful feedback about your product or service pricing or forecast future product development. This analysis identifies the given text and whether it contains negative, positive, or neutral emotions. It is a text analysis that uses Natural Language Processing (NLP) and machine learning. The key aspect of sentiment analysis is the Polarity Classification; this polarity can be expressed in a numerical value known as Sentiment Scoring. This is the overall sentiment delivered by a particular text, phrase, or word.

The author aims to find the gap between different research regarding data mining sentiment analysis for an online product review, and second, to highlight the best approach for data mining sentiment analysis for an online product review to address challenges.

METHODOLOGY

The author used qualitative research using the content analysis method. This paper will conduct comprehensive research reviews, including journals, articles, books, and published research papers from different sources and online databases. Significant procedures were followed to ensure a high-quality review of the literature and provide a clear picture of the state of knowledge in data mining sentiment analysis. A qualitative research methodology was employed, utilizing content analysis to review literature from various sources, including peer-reviewed journals, articles, and books. The following databases were consulted: IEEE.org, Semantic Scholar, Google Scholar, Google Books, Research Gate, Academia.edu, and doaj.org. A

total of thirty relevant studies were analyzed to provide a comprehensive overview of the state of knowledge in sentiment analysis.

First, the author conducted thorough research from different data sources, such as published papers and articles, regarding data mining using semantic analysis of product reviews. Seven data sources were used, including IEEE.org, Semantic Scholar, Google Scholar, Google Books, Research Gate, Academia.edu, and doaj.org. Second, the author searched for online journals to add thorough reviews of semantic analysis of product reviews, including The International Journal of Computer Applications, International Journal of Soft Computing, Eurasia Journal of Mathematics, Science, and Technology Education, Didactics and Technology in Mathematical Education, Innovations in Computer Science and Engineering, GRD Journals, Journal of Big Data, International Conference on Artificial Intelligence and Big Data (ICAIBD), Journal of Physics: Conference Series, International Journal Of Scientific & Technology Research, International Journal for Research in Engineering Application & Management, International Research Journal of Engineering and Technology (IRJET), Journal of Emerging Technologies and Innovative Research (JETIR), International Journal Of Engineering Research & Technology (IJERT), and International Journal of Innovations in Engineering and Science. This systematic review includes 30 journals, articles, and research papers.

Data Mining Techniques

The process of sifting through massive data sets to find links and patterns that may be used to address business problems through data analysis is known as data mining. Businesses are able to forecast future trends and make better-informed business decisions by utilizing data mining techniques and technologies. (Stedman, 2021). The following algorithms/techniques/methods were commonly used in all the research, journals, and articles that the author included:

1. Lexicon-based approach - The Lexicon-based approach aggregates the sentiment ratings of each word in a document using a pre-made sentiment lexicon to score the document. There should be a term and matching sentiment score in the pre-prepared sentiment lexicon. Vocabulary words in negation should be added to the lexicon separately and be given more weight than their equivalent non-negation counterparts [31].
2. Support Vector Machine - For two-group classification problems, supervised machine learning models such as support vector machines (SVMs) employ classification techniques. An SVM model can classify new text after being given sets of labeled training data for each category [32].
3. Ordinal classification - One of the most important classification issues in machine learning is ordinal classification, which is proposed to predict ordinal target values [33].
4. Naïve Bayesian - is a classification method that relies on the independence of predictors and is based on the Bayes Theorem. To put it simply, a Naive Bayes classifier makes the assumption that a given feature in a class is independent of the existence of any other feature [34].
5. Random Forest – a popular supervised machine learning algorithm for tasks involving regression and classification. Using various samples, it constructs decision trees and uses the majority vote for categorization and the average vote for regression [35].
6. Logistic Regression - Regression analysis should be performed appropriately when the dependent variable is dichotomous (binary). used to define data and clarify how one binary dependent variable and one or more nominal, ordinal, interval, or ratio-level independent variables relate to one other [36].
7. Bi-LSTM - The technique of creating a neural network to have sequence information in both directions—forward (past to future) or backward (future to past)—is known as bidirectional long-short-term memory, or bi-lsTM. The bidirectional nature of our input distinguishes a bidirectional LSTM from a conventional LSTM [37].
8. Sentiment lexicon – It is a crucial tool for determining the textual and lexical sentiment polarity. The creation of sentiment lexicons automatically has emerged as a research area in opinion mining and sentiment analysis [38].

9. Deep learning technology – It is a method of machine learning that teaches machines to learn by doing what humans do naturally: following examples. One of the main technologies underlying driverless cars is deep learning, which allows them to tell a pedestrian from a lamppost or recognize a stop sign. In consumer electronics like phones, tablets, TVs, and hands-free speakers, it is the secret to voice control. Deep learning is getting lots of attention lately, and for a good reason. It's achieving results that were not possible before [39].
10. Feature-based vector - It is a vector that has several details about an object in it. Feature space can be created by combining object feature vectors. The characteristics could collectively represent a full image or just one single pixel. The level of detail relies on what the user is attempting to understand or convey about the thing. A feature vector that indicates a three-dimensional shape's height, width, depth, etc., could be used to characterize it [40].
11. Novel weighting algorithm - Machine learning models employ attribute weighting modifications to enhance their efficacy. This work presents a new mutual information-based attribute weighting technique and applies it to four traditional machine learning models for classification [41].
12. SentiWordNet lexical resource - A lexical resource that assigns three numerical scores, Obj(s), Pos(s), and Neg(s), to each synset of WORD-NET (version 2.0), indicating how Objective, Positive, and Negative the terms in the synset are [42].
13. Cross-Category Test - The tactic of shifting between categories (as opposed to inside one) to accommodate financial restraints while sating cravings [43].
14. Weighted k-Nearest Neighbor (Weighted k-NN) Classifier - Based on two or more numerical predictor variables, the weighted k-nearest neighbors (k-NN) classification algorithm is a rather straightforward method for predicting an item's class. For instance, you may like to forecast a person's political party (independent, democratic, or republican) based on factors like age, gender, years of schooling, and so forth. This post describes how to use Python to create the weighted k-nearest neighbor method [44].
15. K-means cluster – It is a straightforward unsupervised learning approach that addresses clustering issues. Using a fixed letter "k," it classifies a given data set into many clusters using a straightforward process. After that, the clusters are arranged as points, and each observation or data point is linked, computed, and adjusted in relation to the closest cluster. After that, the procedure begins to overuse the new modifications until the intended outcome is obtained [45].
16. Multi-layer perceptron – It is an artificial neural network that learns to produce a set of outputs from inputs through feedforward. Multiple layers of input nodes connected as a directed graph between the input and output layers define an MLP [46].
17. Sentiment Classification methods - Sentiment classification is the automatic method of recognizing viewpoints inside a text and classifying them as neutral, positive, or negative depending on the feelings that users convey [47].
18. Decision Tree - A decision tree is a model that resembles a tree that is used as a tool for decision support. It shows decisions along with their costs, consequences, and outcomes. It is simple to assess and compare the "branches" in order to choose the best courses of action [50].
19. Stanford NLP parser - For many languages, the Stanford Parser can produce constituency and dependency parses of sentences. PCFG, Shift Reduce, and Neural Dependency parsers are included in the package. Make sure to download the model jar for the particular language you are interested in in order to fully utilize the parser [48].
20. Convolutional Neural Network - A type of artificial neural network that is gaining popularity in radiology is drawing interest from a variety of fields, including computer vision applications where they have grown dominating. CNN uses a variety of building pieces, including convolution layers, pooling layers, and fully connected layers, to automatically and adaptively learn spatial hierarchies of information through backpropagation [49].
21. Shallow Neural Network - Shallow neural networks have one or two hidden layers, at most. Knowing an external neural network allows us to gain insight into a deep neural network's internal workings [50].

22. BERT - Bidirectional Encoder Representations from Transformers is referred to as BERT. It is intended to jointly train on both left and right contexts in order to pre-train deep bidirectional representations from the unlabeled text. Therefore, state-of-the-art models for a variety of NLP applications can be created by fine-tuning the pre-trained BERT model with just one extra output layer [51].

Application of Data Mining in Product Review Using Sentiment Analysis

Sentiment Analysis focuses on predicting the emotion of a given word, phrase, or text. With the use of Natural Language Processing, it can identify three polarities such as “Negative,” “Positive,” and “Neutral.” There are many research and articles that include different approaches to enhance the effectiveness and efficiency of sentiment analysis. Their dataset mostly comes from Amazon.com, a filtered dataset; Twitter and Facebook, which provides structured and unstructured; and e-commerce Sites, which include structured and unstructured datasets. Data mining approaches that are common and most effective in giving a result for Online Product reviews based on all research and articles stated in Table 3.

Support Vector Machine. Performs classification by finding the hyper-plane that differentiates the classes we plotted in n-dimensional space [53]. It applies as the analysis model to improve its efficiency and effectiveness. Some of the research applies it to a recommendation system.

Naïve Bayesian. It is a classifier commonly used to enhance the performance of the semantic analysis. This classifier is widely used for a large amount of data. Hence, it is a simple yet fastest classifier that helps improve the analysis of all product reviews from different data sources. Furthermore, a structured and unstructured dataset might affect the result.

Random Forest. It is a type of supervised machine learning algorithm which is widely used in Classification and regression problems. Regression problems involve output variables that are real or continuous values. Getting datasets from online pages deals with real and continuous values. Hence, Random Forest is very effective in Sentiment Analysis. As a result, it can cover structured and unstructured datasets.

Logistic Regression. It is a classification that solves binary classification problems. It is implemented in the semantic analysis since it includes polarities such as Positive, Negative, and Neutral. However, Logistic Regression has only two possible results, 1 and 0. This classification might result from the same effectivity of SVM and the same execution time with Naïve Bayesian.

Multi-layer Perception. Input, output, and one or more hidden layers with many neurons stacked on top of each other comprise a multilayer perceptron. Additionally, neurons in a Multilayer Perceptron can utilize any arbitrary activation function, but in a Perceptron, a neuron must have an activation function that enforces a threshold, such as ReLU or sigmoid [54]. This algorithm enhances the accuracy of semantic analysis, and based on the research cited, MPL does better than SVM, which can reach more than 90% classification correctly.

Product Review using Sentiment Analysis Process

Collection of Dataset. Collecting a dataset is choosing one or more data sources that can essentially contribute to the study. In the sentiment analysis, the common dataset comes from social media sites and e-commerce sites like Facebook, Amazon, Twitter, YouTube, Tiktok, Lazada, and Shopee. Datasets can be structured and unstructured.

Preprocessing. The input data preprocessing removes unnecessary elements in the data, such as symbols, numbers, spacing, etc. Preprocessing is used to clean the inputted data.

Extraction. The input data can be transformed into a reduced set of data. So that, from here, the part of speech will be identified. It includes nouns, verbs, adverbs, adjectives, pronouns, conjunctions, and prepositions.

Polarity. The polarity of the specified part of speech will check if it is negative, positive, or neutral. This process classifies all input datasets and returns clustered data.

Evaluation. The evaluation identifies different aspects of the analysis: First is accuracy, time, satisfaction, and correctness. Second, determine the highest value of the polarities that will help the organization enhance its strategy and develop an effective data-driven decision.

Comparative Metrics for Evaluating Sentiment Analysis Methods

1. **Accuracy:** This is the primary metric which measure the percentage of correct prediction made by the model, it indicates how well each method can correctly classify sentiment.
2. **Precision and Recall:**
 - a. Precision measures the percentage of true positive identifications of all positive predictions.
 - b. Recall measures the proportion of true positives out of all actual positives.
3. **Execution Time:** The amount of time it takes a method to process a dataset. This is vital for applications requiring real-time sentiment analysis.
4. **Scalability:** Ability to maintain performance as dataset size increases. This metric assesses effectiveness with large volumes of structured and unstructured data.
5. **Robustness:** Assessed by the method's performance stability across different datasets and settings, including varying levels of noise.
6. **Contextual Sensitivity:** The capability of the method to identify sentiment based on context, such as sarcasm, metaphor, or industry-specific terminology that may alter meaning.
7. **Cultural Adaptability:** The method's ability to accurately analyze sentiment in different cultural contexts. This includes understanding how language, idioms, and expressions affect sentiment interpretation.
8. **Demographic insights:** The degree to which the method incorporates and reflects the sentiments of different demographic groups (age, gender, and language).

RESULTS AND DISCUSSION

This literature review explains how data mining contributes to product reviews using semantic analysis. The collaboration of various techniques and concepts produces a more effective algorithm that could be used with all continuously changing datasets and an algorithm that could be more effective with big datasets coming from different data sources, and even if the datasets are structured or unstructured. Research and articles cited here in this literature review include social media and e-commerce platforms. Furthermore, this review includes 30 research articles from different data sources.

Table 3. Papers According to Applied Technique

No.	Technique	Research Title	References
1	Lexicon Based Approach	Sentilyzer: Aspect-Oriented Sentiment Analysis of Product Reviews	Wladislav, S., Johannes, Z., Christian, W., Andre, K., Madjid, F. (2018)
2	Support Vector Machine and Random Forest	Sentiment Analysis for Product Recommendation Using Random Forest	Gayatri, K., Prof. Deepali, V. (2018)
3	Naïve Bayesian, Random Forest, and Support Vector Machine	Sentiment analysis using product review data	Xing, F., Justin, Z. (2015)
4	Logistic Regression, Naïve Bayes, Random Forest, and Bi-LSTM	Product Sentiment Analysis for Amazon Reviews	Arwa S. M. A. (2021)

5	Support Vector Machine Algorithm-Based Particle Swarm Optimization.	Sentiment Analysis of Smartphone Product Review Using Support Vector Machine Algorithm-Based Particle Swarm Optimization	Mochamad W., Dinar Ajeng K. (2016)
	Evaluation: 10-Fold Cross-Validation		
	Accuracy: Confusion Matrix and ROC curve		
6	Sentiment lexicon and deep learning technology	Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning	Li Y., Ying L., Jin W., Simon S., (2020)
7	Python and Tableau	Sentimental Visualization: Semantic Analysis of Online Product Reviews Using Python and Tableau	Hanan A. (2020)
8	Support Vector Machine	A feature based approach for sentiment analysis using SVM and coreference resolution	M. Hari K., K. R., Ali A. (2017)
9	Ensemble the classifier with the random forest technique	Sentiment Analysis Using Random Forest Ensemble for Mobile Product Reviews in Kannada	Yashaswini H., S.K. P. (2017)
10	Feature-based vector model and a novel weighting algorithm	A novel feature-based method for sentiment analysis of Chinese product reviews	Liu L., Song W., Wang H., Li C., Lu J. (2014)
11	SentiWordNet lexical resource	Sentiment analysis from product reviews using SentiWordNet as lexical resource	Alexandra C., Valentin S., Bogdan M. (2015)
12	Fusion Semantic, Fusion All, and Cross-Category Test	Semantic Analysis and Helpfulness Prediction of Text for Online Product Reviews	Yinfei Y., Yaowei Y., Minghui Q., Forrest Sheng B. (2015).
13	Weighted k-Nearest Neighbor (Weighted k-NN) Classifier	Supervised Semantic Analysis of Product Reviews Using Weighted k-NN Classifier	Ankita S., M.P. S., Prabhat K. (2014)
14	Naive Bayes, Logistic Regression, and Support Vector Machines	Sentiment analysis of Twitter data: A machine learning approach to analyse demonetization tweets	Brinda H., Nagashree H., Madhura P. (2018)
15	K-means cluster	Sentiment Analysis on Online Product Review	Raheesa S., K.R.S., T.S.Shri S., E.A.V. (2017)
16	Machine learning and Natural Language Processing	Sentiment Analysis: On Product Review	Ugandhara N., Priti N., Bhagyashree G. (2016)
17	Support Vector Machine, Naive Bayes algorithm, and multi-layer perceptron.	Product Review Sentiment Analysis – A Survey	Uma D., Vallinayagi V. (2019)
18	SENTIWORDNET	Sentiment Analysis of Product Reviews and Evaluation of Trustworthiness	Vivek P., Zaineb P., Sneha P., Rhea S., Prof. Reena M. (2017)

19	Content analysis, and Sentiment classification	Sentiment analysis of product review	Krutika W., Pranali R., Rushabh B., Nadim B., Bhuvneshwar K. (2018)
20	Naïve Bayes Classifier and Support Vector Machine	Application of Sentiment Analysis on Product Review Ecommerce	Yuniarta B., Harris S., Sinta Ida Patona S., Jen Presly S. (2019)
21	Lexicon-based approach, Deep learning techniques, and Sentiment Classification methods	Sentiment Analysis of Product Reviews – A Survey	Dishi J., Bitra Harsha V., Saravanakumar K. (2019)
22	Opinion Mining	A Survey on Sentiment Analysis of (Product) Reviews	Nisha Jebaseeli A., Kirubakaran E. (2012)
23	Naïve Bayes, Logistic Regression, Linear Support Vector Classifier (SVC), and Decision Tree	Sentiment Analysis for Product Review	Najma S., Pintu K., Monika Rani P., Sourabh C., S.K. Safikul A. (2019)
24	Naïve Bayes, Support Vector Machine, Decision Tree, and Random Forest. Extraction Techniques: s Bag of words and TF-IDF	Sentiment Analysis Using Machine Learning Approach	Andreea-Maria C. (2021)
25	Regression Analysis and Supervised Machine Learning	Predicting Supervise Machine Learning Performances for Sentiment Analysis Using Contextual-Based Approaches	Azwa Abdul Az., Andrew S. (2019)
26	Random Forest method and 10-fold cross-validation	Sentiment Analysis on Tokopedia Product Online Reviews Using Random Forest Method	Stephanie, Budi W., Alan P. (2020)
27	Multivariate filter-based approach, and Stanford NLP parser	A supervised scheme for aspect extraction in sentiment analysis using the hybrid feature set of word dependency relations and lemmas	Bhavana R. B., Jeyanthi P. (2021)
28	Convolutional Neural Network, Shallow Neural Network, Support Vector Machines, K-Nearest Neighbor (KNN), Naive Bayes, and Random Forest	A Domain-Independent Classification Model for Sentiment Analysis Using Neural Models	Nour J., Fadi Al M., ORCID and Wolfgang K., (2020)
29	Machine Learning Algorithms, I.E., Logistic Regression, Random Forest, Naïve Bayes, Bidirectional Long-Short Term Memory, and Bert	Product Sentiment Analysis for Amazon Reviews	Arwa S. M. A. (2021)
30	Support vector machine (SVM), and Naïve Bayes	Sentiment analysis of product reviews: A review	T. K. S., Jyothi S. (2017)

Table 3 shows Online Product Reviews using Semantic Analysis, Random Forest, Naïve Bayesian, Support Vector Machine, and logistic regression are the standard, and effective techniques applied.

Table 4. Summary of Algorithms

No.	Reference	Problems/Objectives	Algorithm/Method/Technique	Key Findings
1	Wladislav, S., Johannes, Z., Christian, W., Andre, K., Madjid, F. (2018)	Aspect-oriented sentiment analysis of product reviews	Sentilyzer	Effective in identifying product features influencing sentiment.
2	Gayatri, K., Prof. Deepali, V. (2018)	Product recommendation	Random Forest	Improved accuracy in sentiment classification for product recommendations.
3	Xing, F., Justin, Z. (2015)	Sentiment analysis using product review data	Various techniques	Emphasized the importance of data preprocessing for effective sentiment analysis.
4	Arwa S. M. A. (2021)	Sentiment analysis for Amazon reviews	Support Vector Machine	Achieved high accuracy in classifying sentiments of Amazon product reviews.
5	Mochamad W., Dinar Ajeng K. (2016)	Sentiment analysis of smartphone product reviews	Support Vector Machine with Particle Swarm Optimization	Enhanced performance in sentiment classification of smartphone reviews.
6	Li Y., Ying L., Jin W., Simon S. (2020)	Sentiment analysis for e-commerce product reviews in Chinese	Sentiment Lexicon and Deep Learning	Demonstrated effectiveness in analyzing sentiments in Chinese e-commerce reviews.
7	Hanan A.	Feature-based sentiment analysis	Naïve Bayes, Support Vector Machine, Decision Tree, Random Forest	Developed a classifier predicting consumer happiness with high accuracy.

Table 4 summarizes the different algorithms and techniques applied to sentiment analysis of product reviews. The most commonly used methods include Random Forest, Naïve Bayes, Support Vector Machine, and Logistic Regression. Each technique's effectiveness varies based on the dataset's structure and size.

Table 5. Practical Case Studies

Practical Case Studies	Methodology	Evaluate Metrics to Evaluate	Cultural and Demographic Focus
E-commerce Product Reviews	Naïve Bayes and SVM for product reviews analysis	Accuracy, precision, recall, execution time	Analysis based on reviews from diverse cultural demographics.
Social Media Sentiment During Events	Deep learning techniques (e.g., LSTM) for Twitter data during significant events	Execution time, contextual sensitivity	Analyzing posts by demographic characteristics (age, location).
Customer Service Feedback Analysis	Hybrid models (lexicon-based + Random Forest) for analyzing customer feedback	Accuracy, robustness, user comprehensibility	Evaluation of linguistic variations and their impact on sentiment across regions (e.g., US vs. UK English).

The table 5 presents the practical case studies that shows a structured overview of various research studies in sentiment analysis, focusing on methodology, evaluation metrics, and cultural or demographic context.

Findings and Challenges Identified

The study identified three (3) findings:

1. Variety of Approaches: many methodologies produce different effectiveness based on the dataset and context.
2. Dataset Characteristics: the difference between structured and unstructured data significantly impacts analysis outcomes.
3. Contextual Factors: Incorporating demographic insights enhances the accuracy and relevance of sentiment analysis.

Challenges Identified:

1. Gaps in Current Research: There are limited integration of contextual factors such as user demographics and product categories. Moreover, it needs for hybrid models that combine multiple techniques for better categories.
2. Scalability Issues: There is also challenges in efficiently processing and analyzing large volumes of online reviews.

CONCLUSION

The landscape of data mining sentiment analysis approaches applied to online product evaluations from 2012 to 2021 has been critically explored in this systematic research. According to the type of dataset and the particular context of the analysis, the results show a wide range of approaches, including lexicon-based approaches, deep learning techniques, and machine learning algorithms like Support Vector Machines (SVM), Random Forest, and others. Each approach demonstrates varying degrees of effectiveness.

The review emphasizes how crucial it is to use methods that are suited to the particulars of the data, especially when it comes to differentiating between structured and unstructured datasets. Moreover, it emphasizes how important it is to take reviewers' demographic and contextual information into account in order to improve the precision and dependability of sentiment analysis results. Researchers can improve the interpretability of sentiment analysis results by better understanding customer behavior and preferences by grouping reviewers into separate categories.

Future studies should concentrate on creating hybrid models that combine several strategies to take advantage of their advantages while minimizing their disadvantages. Furthermore, investigating more sophisticated techniques for natural language processing (NLP), including transformer-based models, may improve the semantic comprehension of product reviews even more. The demand for reliable, scalable, and context-aware sentiment analysis frameworks is rising in tandem with the volume of online reviews. Thus, the following are the possible future studies directions:

1. Hybrid Models: future studies should focus on developing models that integrate strengths from various techniques.
2. Advanced NLP Techniques: Exploring the use of transformer-based models to improve semantic understanding and sentiment classifications.
3. Consumer Behavior Insights: Investigating how demographic factors influence sentiment can provide deeper insights into consumer preferences.

To sum up, this review not only points out gaps in the literature that currently exist, but it also offers a path forward for further research focused on improving sentiment analysis techniques. Researchers can help

develop more advanced systems that can accurately analyze customer attitudes by filling up these gaps, which would ultimately benefit both consumers and businesses.

REFERENCES

1. Wladislav, S., Johannes, Z., Christian, W., Andre, K., & Madjid, F. (2018). Sentlyzer: Aspect-oriented sentiment analysis of product reviews. In *2018 International Conference on Computational Science and Computational Intelligence*.
2. Gayatri, K., & Prof. Deepali, V. (2018). Sentiment analysis for product recommendation using random forest. *International Journal of Engineering & Technology*, 7(3.3), 87-89.
3. Xing, F., & Justin, Z. (2015). Sentiment analysis using product review data. *Journal of Big Data*, 2, Article 5.
4. Arwa, S. M. A. (2021). Product sentiment analysis for Amazon reviews. *International Journal of Computer Science & Information Technology (IJCSIT)*, 13(3), 15. <https://doi.org/10.5121/ijcsit.2021.13302>
5. Mochamad, W., & Dinar Ajeng, K. (2016). Sentiment analysis of smartphone product review using support vector machine algorithm-based particle swarm optimization. *Journal of Theoretical and Applied Information Technology*, 91(1).
6. Li, Y., Ying, L., Jin, W., & Simon, S. (2020). Sentiment analysis for e-commerce product reviews in Chinese based on sentiment lexicon and deep learning. <https://doi.org/10.1109/ACCESS.2020.2969854>
7. Hanan, A. (2020). Sentimental visualization: Semantic analysis of online product reviews using Python and Tableau. In *2020 IEEE International Conference on Big Data (Big Data)*. <https://doi.org/10.1109/BigData50022.2020.9391769>
8. Hari, K. M., K. R., & Ali, A. (2017). A feature-based approach for sentiment analysis using SVM and coreference resolution. In *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*.
9. Yashaswini, H., & S.K. P. (2017). Sentiment analysis using random forest ensemble for mobile product reviews in Kannada. In *2017 IEEE 7th International Advance Computing Conference (IACC)*. <https://doi.org/10.1109/IACC.2017.0160>
10. Liu, L., Song, W., Wang, H., Li, C., & Lu, J. (2014). A novel feature-based method for sentiment analysis of Chinese product reviews. *China Communications*, 11(3).
11. Alexandra, C., Valentin, S., & Bogdan, M. (2015). Sentiment analysis from product reviews using SentiWordNet as lexical resource. In *2015 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*.
12. Yinfei, Y., Yaowei, Y., Minghui, Q., & Forrest Sheng, B. (2015). Semantic analysis and helpfulness prediction of text for online product reviews. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Short Papers)*(pp. 38–44).
13. Ankita, S., M.P. S., & Prabhat, K. (2014). Supervised semantic analysis of product reviews using weighted k-NN classifier. In *2014 11th International Conference on Information Technology: New Generations*.
14. Brinda, H., Nagashree, H., & Madhura, P. (2018). Sentiment analysis of Twitter data: A machine learning approach to analyze demonetization tweets. *International Research Journal of Engineering and Technology (IRJET)*, 5.
15. Raheesa, S., K.R.S., T.S. Shri, S., & E.A.V. (2017). Sentiment analysis on online product review. *International Research Journal of Engineering and Technology (IRJET)*, 4.
16. Ugandhara, N., Priti, N., & Bhagyashree, G. (2016). Sentiment analysis: On product review. *International Journal for Research in Engineering Application & Management (IJREAM)*, 2(2), 1-5. ISSN: 2494-915

17. Uma, D., & Vallinayagi, V. (2019). Product review sentiment analysis – A survey. *Journal of Emerging Technologies and Innovative Research (JETIR)*, 6(2), 1-5.
18. Vivek, P., Zaineb, P., Sneha, P., Rhea, S., & Reena, M. (2017). Sentiment analysis of product reviews and evaluation of trustworthiness. *International Journal of Engineering Research & Technology*, 5(1), 1-5.
19. Krutika, W., Pranali, R., Rushabh, B., Nadim, B., & Bhuvneshwar, K. (2018). Sentiment analysis of product review. *International Journal of Innovations in Engineering and Science*, 3(5), 1-5.
20. Yuniarta, B., Harris, S., Sinta Ida Patona, S., & Jen Presly, S. (2019). Application of sentiment analysis on product review e-commerce. *Journal of Physics: Conference Series*, 1st International Conference on Advance and Scientific Innovation (ICASI), 1-5.
21. Dishy, J., Bitra Harsha, V., & Saravanakumar, K. (2019). Sentiment analysis of product reviews – A survey. *International Journal of Scientific & Technology Research*, 8(12), 1-5.
22. Nisha Jebaseeli, A., & Kirubakaran, E. (2012). A survey on sentiment analysis of (product) reviews. *International Journal of Computer Applications*, 47(11), 1-5.
23. Najma, S., Pintu, K., Monika Rani, P., Sourabh, C., & Safikul, A. S. (2019). Sentiment analysis for product review. *ICTACT Journal on Soft Computing*, 9(3), 1-5.
24. Andreea-Maria, C. (2021). Sentiment analysis using machine learning approach. *Ovidius University Annals, Economic Sciences Series*, 21(1), 1-5.
25. Azwa Abdul, A., & Andrew, S. (2019). Predicting supervised machine learning performances for sentiment analysis using contextual-based approaches. *IEEE Access*, 8, 1-5. <https://doi.org/10.1109/ACCESS.2019.2958702>
26. Stephenie, B., Budi, W., & Alan, P. (2020). Sentiment analysis on Tokopedia product online reviews using random forest method. *E3S Web of Conferences*, 202, 16006.
27. Bhavana, R. B., & Jeyanthi, P. (2021). A supervised scheme for aspect extraction in sentiment analysis using the hybrid feature set of word dependency relations and lemmas. *PeerJ Computer Science*, 7, e347.
28. Nour, J., Fadi Al, M., & Wolfgang, K. (2020). A domain-independent classification model for sentiment analysis using neural models. *Applied Sciences*, 10(18), 6221.
29. Arwa, S. M. A. (2021). Product sentiment analysis for Amazon reviews. *International Journal of Computer Science & Information Technology (IJCSIT)*, 13(3), 1-5.
30. K. S., & Jyothi, S. (2017). Sentiment analysis of product reviews: A review. In 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT) (pp. 1-5). <https://doi.org/10.1109/ICICCT.2017.7975207>
31. Kannan, S., & Kejariwal, A. (2016). Big data analytics for social media. *Big Data*.
32. Stecanella, B. (2017). Support vector machines (SVM) algorithm explained. Retrieved from <https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm/>
33. Yildirim, P., Birant, U. K., & Birant, D. (2019). EBOC: Ensemble-based ordinal classification in transportation. *Journal of Advanced Transportation*, 2019, Article ID 7482138. <https://doi.org/10.1155/2019/7482138>
34. Sunil. (2017). 6 easy steps to learn Naive Bayes algorithm with codes in Python and R. Retrieved from <https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/>
35. Sruthi, E. R. (2021). Understanding random forest. Retrieved from <https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/#:~:text=Random%20forest%20is%20a%20Supervised,average%20in%20case%20of%20Oregression.>
36. Statistics Solutions. (n.d.). What is logistic regression? Retrieved from <https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/what-is-logistic-regression/>
37. Yugesh, V. (2021). Complete guide to bidirectional LSTM (with Python codes). Developer Corner. Retrieved from [https://analyticsindiamag.com/complete-guide-to-bidirectional-lstm-with-python-codes/#:~:text=Bidirectional%20long%2Dshort%20term%20memory\(bi%2Dlstm\)%20is,different%20from%20the%20regular%20LSTM.](https://analyticsindiamag.com/complete-guide-to-bidirectional-lstm-with-python-codes/#:~:text=Bidirectional%20long%2Dshort%20term%20memory(bi%2Dlstm)%20is,different%20from%20the%20regular%20LSTM.)

38. Wang, L., & Xu, R. (2017). Sentiment lexicon construction with representation learning based on hierarchical sentiment supervision. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. <https://aclanthology.org/D17-1052>
39. MathWorks. (n.d.). What is deep learning? 3 things you need to know. Retrieved from <https://www.mathworks.com/discovery/deep-learning.html#:~:text=Deep%20learning%20is%20a%20machine,a%20pedestrian%20from%20a%20lamppost>.
40. Brilliant.org. org. (2022). Feature vector. Retrieved from <https://brilliant.org/wiki/feature-vector/#:~:text=A%20feature%20vector%20is%20a,pixel%20or%20an%20entire%20image>.
41. Subhash, B., Wang, T., & B., S. (2021). A novel weighting attribute method for binary classification. This work is licensed under the Creative Commons Attribution International License (CC BY). Retrieved from <http://creativecommons.org/licenses/by/4.0/>
42. Esuli, A., & Sebastiani, F. (2006). SENTIWORDNET: A publicly available lexical resource for opinion mining. Retrieved from <https://www.esuli.it/publications/LREC2006.pdf>
43. M, T., S., C., B., J., & M., S. (2016). Cross-category indulgence: Why do some premium brands grow during recession? *Journal of Brand Management*, 114-129.
44. James, M. (2019). Weighted k-NN classification using Python. Retrieved from <https://visualstudiomagazine.com/articles/2019/04/01/weighted-k-nn-classification.aspx#:~:text=The%20weighted%20k%2Dnearest%20neighbors,or%20more%20numeric%20predictor%20variables>.
45. Techopedia. (n.d.). K-means clustering: What does K-means cluster? Retrieved from <https://www.techopedia.com/definition/32057/k-means-clustering#:~:text=K%2Dmeans%20clustering%20is%20a%20method%20used%20for%20clustering%20analysis,the%20data%20into%20Voronoi%20cells>.
46. Make sure to adjust any specific details (like author names or titles) if they were misinterpreted or if additional information is available.
47. Techopedia. (n.d.). Multilayer perceptron (MLP). [https://www.techopedia.com/definition/20879/multilayer-perceptron-mlp#:~:text=A%20multilayer%20perceptron%20\(MLP\)%20is,the%20input%20and%20output%20layers](https://www.techopedia.com/definition/20879/multilayer-perceptron-mlp#:~:text=A%20multilayer%20perceptron%20(MLP)%20is,the%20input%20and%20output%20layers).
48. Danial, D. (2020). Sentiment classification – The low-down. <https://monkeylearn.com/blog/sentiment-classification/#:~:text=Sentiment%20classification%20is%20the%20automated,emotions%20customers%20express%20within%20them>.
49. Heavy. AI. (n.d.). What is decision tree analysis? <https://www.heavy.ai/technical-glossary/decision-tree-analysis#:~:text=A%20decision%20tree%20is%20a,the%20best%20courses%20of%20action>.
50. Core NLP (2020). Parser. <https://stanfordnlp.github.io/CoreNLP/parser-standalone.html#:~:text=The%20Stanford%20Parser%20can%20be,language%20you%20are%20interested%20in>.
51. Rikiya, Y., Mizuho, N., Kinh Gian, R., & Kaori, T. (2018). Convolutional neural networks: An overview and application in radiology. In *IEEE International Symposium on Biomedical Imaging* (pp. 611–629).
52. Rochak, A. (2019). Shallow neural networks. Towards Data Science. <https://towardsdatascience.com/shallow-neural-networks-23594aa97a5>
53. Mohdsana, D. (2019). Demystifying BERT: A comprehensive guide to the groundbreaking NLP framework .<https://www.analyticsvidhya.com/blog/2019/09/demystifying-bert-groundbreaking-nlp-framework/#:~:text=%E2%80%9CBERT%20stands%20for%20Bidirectional%20Encoder,both%20left%20and%20right%20context>.
54. Vasista, R. (2018). Sentiment analysis using SVM. <https://medium.com/@vasista/sentiment-analysis-using-svm-338d418e3ff1#:~:text=Sentiment%20Analysis%20is%20the%20NLP%20technique%20that%20performs%20on%20the,positive%2C%20negative%2C%20or%20neutral>.
55. Carolina, B. (2021). Multilayer Perceptron Explained With A Real-Life Example And Python Code: Sentiment Analysis. Towards Data Science. <https://towardsdatascience.com/multilayer-perceptron-explained-with-a-real-life-example-and-python-code-sentiment-analysis-cb408ee93141>