Journal of Theoretical and Applied Information Technology
31st October 2023. Vol.101. No 20

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

FERVENT ANT COLONY OPTIMIZATION-BASED DECISION TREES (FACO-DT) FOR ENHANCING CLASSIFICATION ACCURACY IN SENTIMENT ANALYSIS

G.M.BALAJI[1], K.VADIVAZHAGAN[2]

[1] Research Scholar & Assistant Professor,
Department of Computer and Information Science,
Annamalai University, Chidambaram, Tamilnadu, India.
[2] Assistant Professor
Department of Computer and Information Science,
Annamalai University, Chidambaram, Tamilnadu, India.
E-mail: [1]gmbalajimca@gmail.com, [2]vadivazhagan.k@gmail.com

ABSTRACT

Online shopping has transformed how consumers purchase products, offering convenience and a wide range of choices. As customers increasingly rely on online platforms, the role of product reviews has grown significantly. These reviews, submitted by customers who have purchased products online, provide valuable insights to potential buyers about product quality, functionality, and overall satisfaction. However, classifying sentiments accurately from these diverse reviews is a challenging task. Reviews can encompass a broad spectrum of sentiments, often nuanced and context-dependent. Traditional sentiment classification techniques might struggle to handle complex linguistic structures and domain-specific expressions in online product reviews. The proposed work addresses these challenges by introducing the "Fervent Ant Colony Optimization-based Decision Trees (FACO-DT)" approach. FACO-DT harnesses Ant Colony Optimization techniques with Decision Trees to enhance sentiment classification accuracy. Ant Colony Optimization optimizes feature selection, while Decision Trees provide a structured sentiment classification framework. This synergistic approach enables FACO-DT to effectively capture intricate sentiment patterns in reviews, accommodating domain-specific expressions and linguistic complexities. To evaluate the proposed approach, an Amazon product review dataset is utilized. The dataset comprises reviews from Electronics, Industrial, Scientific, and Software domains. Results show that FACO-DT consistently outperforms traditional classification accuracy, recall, and F-measure methods. The approach's ability to adapt to different domains and enhanced sentiment classification accuracy underscores its potential for real-world sentiment analysis applications. The FACO-DT approach offers a promising solution by fusing optimization and decision-making techniques to handle the complexities of sentiment analysis in diverse online product reviews, as evidenced by its robust performance on the Amazon dataset.

Keywords: *Sentiment, Amazon, Ant Colony, Decision Tree, Optimization*

1. INTRODUCTION

The digital revolution has reshaped the shopping landscape, making online purchases routine for millions. Central to this transformation is the influences of customer reviews, which have become integral to the decision-making process of modern consumers. Reviews wield immense power in the online shopping ecosystem. They provide a collective voice that helps shoppers assess product quality, performance, and suitability. Positive reviews instill confidence and validate purchasing choices, while negative reviews serve as red flags, steering buyers away from potential disappointments[1].For businesses, the significance of reviews goes beyond feedback. Positive reviews can translate into increased sales and loyalty, while negative reviews offer insights for product refinement and customer-centric improvements. Engaging with reviews fosters a sense of authenticity and transparency, ultimately building strong customer relationships. In a world where online shopping is more prevalent than ever, reviews serve as digital testimonials. They bridge the gap between consumers and products, offering a communal perspective that shapes the e-commerce landscape and drives continual enhancements[2].

Sentiment analysis, or opinion mining, is a potent tool for transforming online shopping. This technique, rooted in natural language processing (NLP), involves gauging the sentiments expressed in customer reviews and feedback[3]. By categorizing these sentiments as positive, negative, or neutral, businesses gain a profound understanding of customer opinions and can fine-tune their strategies accordingly. For online retailers, sentiment analysis offers valuable insights into product reception. Positive reviews highlight strengths, aiding marketing efforts, while negative feedback pinpoints areas for improvement. The technique enables businesses to identify trends, track the popularity of products, and even anticipate potential issues. Effective sentiment analysis involves preprocessing reviews, extracting key features, and employing machine learning algorithms for sentiment classification. The challenges, such as context ambiguity, sarcasm detection, and cultural nuances, persist, necessitating ongoing research and innovation. As e-commerce grows, sentiment analysis remains a crucial ally for online retailers seeking to enhance customer experiences, optimize product offerings, and maintain a competitive edge in a dynamic market [4].

Bioinspired optimization techniques, drawing inspiration from nature's intricate systems, reshape sentiment classification in the ever-changing online shopping landscape. As customer reviews wield considerable influence, harnessing these techniques is transformative for accurate sentiment analysis[4]. Emulating concepts like genetic evolution and swarm behavior, bioinspired optimization methods fine-tune sentiment classification models. These models decode overt expressions and subtle emotional nuances in online shopping reviews, enabling businesses to gain deeper insights into customer sentiments. Much like organisms adapt to survive in shifting environments, sentiment classification models optimized with bioinspired techniques adjust to evolving language trends and cultural shifts. This adaptability ensures models remain attuned to the intricacies of modern online communication. Bioinspired optimization provides a competitive edge for businesses navigating the digital marketplace's intricacies[5]. Enhanced sentiment analysis leads to improved customer insights, personalized experiences, and more effective marketing strategies. As the virtual shopping landscape continues to evolve, the fusion of nature-inspired optimization with technological advancements promises a refined approach to understanding and leveraging online shopping sentiment[6]. Bio-inspired Optimization [7]–[20][21][22][23][24] has several potentials to solve various research issues.

1.1. Problem Statement

Effective sentiment analysis in online shopping reviews is a formidable challenge due to the intricate complexity of human expression. As customers articulate their opinions through diverse linguistic patterns, idiomatic phrases, and contextual subtleties, traditional sentiment analysis methods face a daunting task in consistently and accurately categorizing sentiments. This challenge is further intensified by the dynamic nature of the e-commerce landscape, where new products, trends, and terminologies regularly emerge, leading to continuous shifts in sentiment nuances. The current sentiment analysis techniques often struggle to adapt to this multifaceted linguistic landscape, resulting in frequent misclassification of sentiments and compromising the overall analysis's accuracy. The consequences of such inaccuracies are significant. E-commerce platforms heavily rely on customer feedback to refine their offerings, tailor their marketing strategies, and enhance user experiences. Inaccurate sentiment analysis can lead to flawed decision-making, skewed perceptions of customer sentiment, and misguided strategic directions. Addressing this challenge is crucial for the credibility of sentiment analysis and for enabling businesses to respond effectively to customer feedback. A nuanced and accurate sentiment analysis approach specifically designed for the intricate world of online shopping is needed to ensure that the insights extracted from customer reviews align with the actual sentiments expressed. This would empower e-commerce platforms to make informed decisions, create meaningful improvements, and foster stronger customer relationships.

1.2. Motivation

In evolving modern commerce, online shopping has become a pivotal aspect of consumer behaviour. With this surge, the analysis of customer sentiment has gained immense significance. Deciphering the genuine attitudes and emotions embedded in online reviews and feedback is crucial for businesses striving to tailor their offerings to customer preferences. Sentiment analysis in the context of online shopping holds the promise of uncovering nuanced insights that extend beyond surface-level observations. By delving into the inherent challenges posed by diverse expressions and contexts, this research seeks to enhance sentiment analysis accuracy. This endeavour

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

contributes to businesses' ability to refine their products and services and empowers consumers with more informed purchasing decisions. The study thus bridges the gap between consumer intent and business strategies, rendering sentiment analysis a pivotal tool in the contemporary e-commerce landscape.

1.3. Objective

This research aims to heighten the precision and accuracy of sentiment analysis within online shopping. The objective of employing advanced methodologies that account for the intricate nuances embedded in customer feedback is to develop a sentiment analysis framework that transcends surface-level sentiment categorization. By effectively tackling the challenges posed by diverse linguistic expressions and contextual fluctuations, the study strives to equip businesses with a tool capable of decoding authentic customer sentiment precisely. This empowers businesses to make well-informed decisions, fine-tune product offerings, and strategically adapt marketing approaches according to customer preferences. Additionally, the research endeavours to bridge the gap between consumer intent and business strategies, ultimately contributing to every customer's individualized and gratifying online shopping experience.

2. LITERATURE REVIEW

"Multilingual Universal Sentence Encoder" [25] presents a novel approach for SA on Arabic texts by combining the recurrent gated unit with a multilingual universal sentence encoder. This combination is beneficial as it allows the model to understand better the context and sentiment expressed in Arabic texts. The article also highlights the importance of aspect-based SA for Arabic, as it helps better understand the sentiment expressed towards specific aspects. "Polarity-Aware Attention Network" [26] presents a new model for SA in images using a polarity-aware attention network. This model is designed to recognize and understand the sentiment expressed in images by focusing on the most relevant parts of the image. The article highlights the effectiveness of this model compared to previous approaches, demonstrating improved performance in SA tasks. "Covid-19 Vaccine Hesitancy" [27] presents a study on the sentiment towards COVID-19 vaccinations expressed in Twitter data. The study employs text mining, sentiment analysis, and machine learning techniques to analyze and understand public sentiment toward the COVID-19 vaccine. This research provides valuable insights into public perception and attitudes toward the vaccine, which can inform decision-making and communication strategies for promoting vaccine acceptance.

"Flexible Feature Extraction" [28] introduces a new feature extraction algorithm for SA on Spanish tweets. The algorithm considers the context of words in a tweet, allowing for a more nuanced understanding of the sentiment expressed. This flexible approach can be adapted to different SA tasks, making it a valuable contribution to the field. "Interactive Capsule Network" [29] proposes a new model for implicit SA using interactive capsule networks. The model is designed to effectively capture and understand the text's implicit sentiment. The article demonstrates the effectiveness of this model in comparison to previous approaches, highlighting its ability to accurately identify implicit sentiment in texts. "Marathi SentiWordNet" [30] introduces a new lexical resource, Marathi SentiWordNet, for SA in the Marathi language. The authors show that this resource can improve the accuracy of SA models and make them more accessible to the Marathi-speaking community.

"Hybrid Graph Convolutional Networks" [31] presents a novel aspect-level SA approach using a hybrid graph convolutional network. The authors show that this model outperforms existing models on benchmark datasets and can effectively capture the relationships between words and aspects. "SA with Covid-19 Vaccination Data" is a study [32] that uses SA to understand the public's perception of the COVID-19 booster vaccination program as a requirement for homecoming during Eid Fitr in Indonesia. The authors show that most tweets about the program are neutral or positive, indicating a general acceptance of the program. However, they also highlight the need for further research to understand the reasons behind any negative sentiment and to address any concerns raised by the public.

"Disentangled Linguistic Graph Model (DLGM)" [33] presents a novel approach to aspect-based SA using a disentangled linguistic graph model. The article highlights the importance of transparency in SA and the limitations of existing models. The proposed model uses a graph structure to represent the relationships between words and aspects, making it easier to understand the reasoning behind the sentiment predictions. "Effective Syntactic Dependency Model (ESDM)" [34] combines two state-of-the-art techniques,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Conditional Random Fields (CRF) and Graph Convolutional Networks (GCN), to improve aspect-level SA. The authors show that the combination of CRF and GCN outperforms existing models on benchmark datasets, demonstrating the effectiveness of the proposed approach.

3. FERVENT ANT COLONY OPTIMIZATION BASED DECISION TREES (FACO-DT)

3.1. Decision Trees

Decision Trees (DT) are robust tools extensively employed in sentiment classification tasks, serving as interpretable models that excel in capturing intricate patterns within textual data. DT recursively partitioning the feature space based on attribute values creates a hierarchical binary decision structure that culminates in classification outcomes[35]. DT traverses text's linguistic nuances in sentiment analysis, scrutinizing syntactic and semantic features to discern sentiment-bearing cues. This mechanism enables them to detect subtle sentiment shifts driven by context, idiomatic expressions, and modifiers.

DT's adaptability to non-linear relationships and feature interactions equips them to handle complex linguistic contexts in sentiment-bearing language. Constructing decision paths rooted in feature importance, these trees effectively learn the discriminative attributes contributing to sentiment variation. This, in turn, facilitates the identification of sentiment-laden phrases, expressions, and semantic constructions. Consequently, DT empowers sentiment analysis to discern polarised sentiments and capture the spectrum of nuanced emotional states[36].

The interpretability of DT is another asset, as their nodes and branches form interpretable decision rules that offer insights into sentiment determinants[37]. Moreover, DT's inherent capability to handle multi-class scenarios enables sentiment classification models to extend beyond binary sentiment detection, accommodating more nuanced sentiment categories. Algorithm 1 describes the step-by-step process involved in DT.

Algorithm 1: Decision Trees	
	//Decision Tree Construction
Step 1:	If the stopping criteria are met, or all samples have the same sentiment, create a leaf node with the sentiment label.
Step 2:	Find the feature that maximizes information gain or minimizes the Gini index.
Step 3:	Create a decision node using the chosen feature.
Step 4:	For each unique value of the feature: a. Split the dataset into subsets based on the value. b. Recursively construct the tree for each subset.
	//Decision Tree Classification
Step 5:	Begin at the root of the decision tree.
Step 6:	Traverse down the tree, following the path determined by the sample's feature values.
Step 7:	Keep moving through nodes based on feature conditions until reaching a leaf node.
	//Assigning Sentiment
Step 8:	Upon reaching a leaf node, assign the sentiment label associated with that node as the classification.
	//Handling Unknowns
Step 9:	If the input's feature value was not observed during training: a). If the tree lacks a corresponding branch, assign the default sentiment at an appropriate leaf node.

3.2. Fervent Ant Colony Optimization

Fervent Ant Colony Optimization (FACO) is a powerful optimization method that the collective behaviour of ants has inspired. Ants are primitive organisms with limited individual abilities, but they can work together to solve complex problems through emergent intelligence. FACO is derived from ACO. How ants collaborate and communicate through pheromone trails effectively solves optimization problems[38]. FACO mimics the behaviour of ants by using virtual agents to explore a graph representing the space of possible solutions to a given problem. Each agent lays down pheromone trails as it moves through the graph, attracting other agents to follow the same path. Over time, the pheromone trails are reinforced as more agents follow them, and this process leads to the emergence of a high-quality solution.

The principles of stigmergy, or the emergence of intelligent behaviour from a group of individuals working together towards a common goal, are at the heart of FACO. The self-organizing nature of FACO allows it to quickly and efficiently find high-quality solutions to complex problems. FACO has been applied to widespread real-time problems, from vehicle routing to protein structure prediction. Its ability to find high-quality solutions

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

quickly and efficiently has made it a popular optimization method among researchers and practitioners[39]. The principles of emergent intelligence and self-organization that underlie FACO could also inspire new methods and techniques for various fields beyond optimization. FACO offers a creative and innovative approach to optimization inspired by ants'behaviour. Its ability to find high-quality solutions quickly and efficiently has made it a valuable tool for researchers and practitioners in various fields. By taking natural cues, FACO could inspire new methods and techniques for solving complex problems.

3.2.1. Node Transition Rule

An ant in node s uses the node transition rule to decide which node to transition to next. The probability of transitioning to a node w is given by Eq.(1):

$$w = \begin{cases} arg\ max_{o\omega W_s^a}\{[\beta_{so}(f)].[\alpha_{so}]^{\gamma}\} & if\ x \le x_k \\ W & if\ x > x_k \end{cases}, \quad (1)$$

wherein x is a uniformly distributed random variable over $[0,1]$, $x_k$is an adjustable parameter $(0 \le x_k \le 1)$, and $W\ \omega\ W_s^a$ is the studied node selected at random with probability:

$$M_{sw}^a(f) = \begin{cases} \dfrac{\beta_{so}(f).[\alpha_{sw}]^{\gamma}}{\sum_{z\omega W_s^a}[\beta_{sz}(f)].[\alpha_{sz}]^{\gamma}} \end{cases} \quad (2)$$

Eq.(2) is close to the transition probability present in FACO. The transition rule between FAC states is thus similar to that of the Ant System when $x > x_k$ and different when $x \le x_k$, as is readily apparent. To be more specific, $x \le x_k$ relates to making use of previously acquired information about the problem, such as heuristics regarding the distances between node's pheromone trails, while $x > x_k$ encourages further investigation.

After completing their travel, all ants in the FACO are permitted to leave a pheromone trail. In contrast, just the ant that has found a better solution since the trail's inception is permitted to globally modify the pheromone concentrations on the branches. Eq.(3) indicates the rule of updation followed in FAC.

$$\Delta_{sw}(f+t) = (1-\mu).\beta_{sw}(f) + \mu.\Delta_{sw}(f, f+t), \quad (3)$$

wherein (s, w)represents the trail's best edge since the trail's inception, F^+, and μ represents the decay rate of pheromones.

$$\Delta\beta_{sw}(f, f+t) = \frac{1}{Z^+}, \quad (4)$$

For the f, substitute Z^+ for its distance.

Local updating is carried out as follows: when ant a is an analyzed node in s and chooses node $w\ \omega\ W_s^a$ to travel to during visits t different nodes. The pheromone content on the edges(s, w) is updated using Eq.(5).

$$\beta_{sw}(f+1) = (1-\varphi).\beta_{sw}(f) + \varphi.\beta_k. \quad (5)$$

3.3. FACO with Machine Learning

The FACO metaheuristic algorithm has been applied in diverse machine-learning contexts. The algorithm draws inspiration from the collective behaviour of actual ant colonies, wherein individual ants utilize pheromones to communicate with one another and tackle intricate tasks, such as determining the most efficient route between two given points. FACO has been identified as a viable technique for optimizing the hyperparameters of machine learning models, including but not limited to support vector machines, deep neural networks, DT, and within the domain of machine learning. Hyperparameter optimization aims to identify the optimal configuration of hyperparameters that maximizes the model's performance on a specific dataset. The FACO methodology generates a group of synthetic ants corresponding to various hyperparameter configurations. These ants work together to navigate the hyperparameter space cooperatively.

FACO has also been used for feature selection, identifying the most informative features from a given dataset. Feature selection is crucial in machine learning as it helps to reduce the dimensionality of the data, thus improving the performance and interpretability of the models. FACO achieves feature selection by assigning weights to each feature and allowing the artificial ants to explore different subsets of features to find the best subset that maximizes the classification accuracy. FACO is a versatile algorithm in machine learning applications, such as hyperparameter optimization and feature selection. In real-world issues, where input data is typically noisy or partial, its ability to manage uncertainty and imprecision in data is beneficial. Algorithm 2 describes the working process involved in FACO.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Algorithm 2: FACO

Step 1: Initialize the parameters, such as ants count, the number of iterations, the pheromone decay rate, and the probability parameters.

Step 2: Initialize the threshold level of pheromone levels to be maintained at the graph's edges.

Step 3: For each iteration, do the following:

- **a.** For each ant, start at a random node in the graph.
- **b.** Construct a solution by moving from node to node, guided by the heuristic information and pheromone levels to be maintained according to the problem.
- **c.** Update the pheromone levels of the edges visited by each ant based on the quality of the solution.
- **d.** Keep track of the so-far identified the best solution.

Step 4: If a stopping criterion is met, stop the algorithm and return the best solution found. Otherwise, go back to step 3.

3.4. FACO-based decision tree

FACO-based DT (FACO-DT) is a decision tree algorithm that uses the FACO metaheuristic to select attributes for splitting nodes. This approach was proposed to address traditional DT's limitations, such as noise sensitivity, highdimensionality, and overfitting. In FACO-DT, ants represent candidate attributes and deposit pheromone trails on the edges between nodes. The pheromone trail and heuristic information guide the ants' movement, leading to the selection of attributes that maximize the fitness function.

3.4.1. Initialization

The first step in FACO-DTis to initialize the algorithm. This involves setting up the τ(i.e., pheromone matrix) and the η (i.e., heuristic information matrix), which guide the ants in searching for an optimal solution. The τ is initialized using Eq.(6):

$$\tau_{ij} = \frac{1}{n} \tag{6}$$

where $n$indicates the nodes in the decision tree.

3.4.2. Ants Movement

The ants are then released and allowed to move through the decision tree. The ants use a probabilistic rule at each decision node to decide which branch to take. This rule is based on the pheromone trail and the heuristic information at each node. The probability of an ant taking the jth branch from the ith node is given byEq.(7):

$$P_{ij} = \tau_{ij}{}^{\alpha} \times \left\{\eta_{ij}{}^{\beta} \div \sum \tau_{ij}{}^{\alpha}\right\} \times \eta_{ij}{}^{\beta} \tag{7}$$

where α and βrepresent the parameters that control the trail of pheromone influence and the heuristic information, respectively.

3.4.3. Solution Construction

As the ants move through the decision tree, they construct a solution by selecting the branches to follow. At each decision node, the ant chooses the branch with the highest probability until it reaches a leaf node.

3.4.4. Solution Evaluation

Once a solution has been constructed, it is evaluated using a fitness function. In FAC-based DT, the fitness function is typically based on the categorization precision of the tree on a testing set E and is given byEq.(8):

$$fitness = d(F, E) = \sum_{u=1}^{A} \frac{FM_u}{|E|} \tag{8}$$

whereFM_u -the percentage of class items that are right, $|E|$- the total quantity of items in the testing set, u - class count of choices.

3.4.5. Pheromone Update

Following the construction and evaluation of their solutions, the ants update the pheromone matrix following the fitness of the solutions. The update rule can be expressed in Eq.(9):

$$\tau_{ij} = (1 - \rho) \times \tau_{ij} + \rho \times \Delta\tau_{ij} \tag{9}$$

where ρ represents the rate of pheromone evaporation, and $\Delta\tau_{ij}$ indicates the quantity of pheromone deposits by the ant that constructs the best solution.

3.4.6. Heuristic Information Update

The η can be updated based on the best solution constructed by the ants. The update rule is given in Eq.(10):

$$\eta_{ij} = (1 - \rho) \times \eta_{ij} + \rho \times \Delta\eta_{ij} \tag{10}$$

where $\Delta\eta_{ij}$ indicates the quantity of heuristic information deposited on the jth branch from the ith node by the ant, which constructs the enriched solution.

3.4.7. Termination

The termination of the algorithm occurs upon fulfilment of a stopping condition, which may include reaching a predetermined fitness value or a specified number of iterations.

The FACO-DT algorithm combines ant colony optimization with decision tree construction to create an approach capable of discovering high-quality DT. Using the pheromone matrix and heuristic information matrix to guide the ants' search, the algorithm can efficiently explore the space of possible DT and find optimal solutions.Algorithm 3 describes the overall process involved in FACO-DT.

Algorithm 3: FACO-DT

Input:

- Training set T
- Testing set E
- Parameters: α, β, ρ, $max_{iter}, desired_{fitness}$

Output:

- The best decision tree that maximizes the fitness function
- Utilizing the value of $1/n$, initialize τ and η (i.e., pheromone and heuristic information matrices), where n represents the count of decision nodes in the tree.

Procedure:

Step 1: Iterate the following steps till the stopping criterion is met.

Step 2: Release m ants into the decision tree.

i. Allow each ant to move through the decision tree, choosing the branch with the highest probability at each decision node.

ii. Construct a solution for each ant by selecting the branches it followed, and evaluate the solution's fitness using the fitness function.

iii. Update τ (i.e., matrix of pheromone) and η (i.e., matrix of heuristic information) using the best solution constructed by the ants.

iv. If the fitness of the best solution meets the expected fitness or the highest number of iterations, then the algorithm is terminated.

Step 3: Select the best solution among all the constructed solutions and return the corresponding decision tree.

4. ABOUT THE DATASET

The Amazon product review dataset is a vast collection of customer reviews and ratings that provides valuable insights into consumer behaviour and preferences. With millions of reviews across various product categories, it is an essential resource for businesses and researchers looking to understand consumer trends and sentiment. One of the most significant benefits of the Amazon product review dataset is the ability to perform sentiment analysis. The reviews' text provides qualitative data that can be analyzed to identify positive or negative sentiments towards a product, which can be used to inform marketing strategies and product development. Another essential feature of the dataset is the inclusion of helpful and total votes. These metrics measure how valuable other customers found a detailed review, making it easier to identify high-quality reviews that provide valuable insights into the product.

The metadata included in the Amazon product review dataset is also crucial for more in-depth analysis. For example, the product category can be used to compare sentiment across different types of products, while the review date can be used to identify seasonal trends in consumer sentiment. The Amazon product review dataset is valuable for machine learning and data science applications. The vast size and diversity of the dataset make it suitable for training models for recommendation systems, natural language processing, and other applications. The Amazon product review dataset is essential for businesses and researchers looking to understand consumer behaviour and sentiment. With its vast size, diversity, and range of features, the dataset offers a wealth of information that can be used to inform marketing and product development strategies and improve customer satisfaction.

Table 1. Dataset Details

Dataset Name	Total Reviews
Electronics	2,09,94,353
Industrial and Scientific	7,58,333
Software	4,59,436

5. PERFORMANCE METRICS

Performance metrics are quantifiable measures used to evaluate the efficiency of the classifiers. They provide objective insights for assessing progress, identifying areas for improvement, and making informed decisions in various domains. This research used the below-

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

mentioned metrics for evaluating the proposed classifier against the state-of-the-art classifiers.

- **Accuracy:** In machine learning and data analysis, accuracy refers to measuring how well a model or algorithm correctly classifies or predicts outcomes. It is calculated as the ratio of correct predictions to the total number of predictions made.
- **F-Measure:** F-measure is a metric used to evaluate the performance of a binary classification model, which considers both the precision and recall of the model. It is the harmonic mean of precision and recall and is often used when both precision and recall are essential in the given problem.
- **Precision:** Precision is a metric used to evaluate the accuracy of a binary classification model. It is calculated as the ratio of true positives to the number of positive predictions made.
- **Recall:** It is a metric used to evaluate the completeness of a binary classification model. It is calculated as the ratio of true positives to the total number of actual positives.

The four significant variables used in the metrics mentioned above are described below.

- **TP (True Positive):** Model correctly predicts positive sentiment when sentiment is positive.
- **TN (True Negative):** Model correctly predicts negative sentiment when sentiment is negative.
- **FP (False Positive):** Model predicts positive sentiment but negative sentiment.
- **FN (False Negative):** Model predicts negative sentiment, but the sentiment is positive.

6. RESULTS AND DISCUSSIONS

6.1. Precision Analysis

Precision is a vital metric in sentiment analysis, measuring the accuracy of positive sentiment predictions by algorithms. Figure 1 offers a snapshot of precision results from three sentiment analysis methods: DLGM, ESDM, and FACO-DT. These approaches were evaluated on Amazon product review datasets categorized as Electronics, Industrial, Scientific, and Software domains. The presented precision values are extracted from Table 2, revealing exciting trends.

DLGM employs a linguistic graph structure to uncover intricate relationships in text, yielding precision scores of 53.098%, 52.435%, and 57.175% across Electronics, Industrial and Scientific, and Software domains. This indicates DLGM's adaptability to diverse subjects but hints at potential challenges with domain-specific subtleties. ESDM classifies syntactic dependencies within sentence structures, achieving precision scores of 72.652%, 65.722%, and 65.998%. Its strength in sentiment identification, particularly in the Electronics domain, stems from its proficiency in understanding technical nuances. However, relatively consistent scores suggest difficulty capturing domain-specific context and vocabulary variations. FACO-DT combines Ant Colony Optimization with Decision Trees for sentiment classification, achieving precision scores of 81.218%, 80.921%, and 78.205%. FACO-DT captures nuanced sentiments across domains, showcasing adaptability and efficiency through optimization and decision-making synergy.

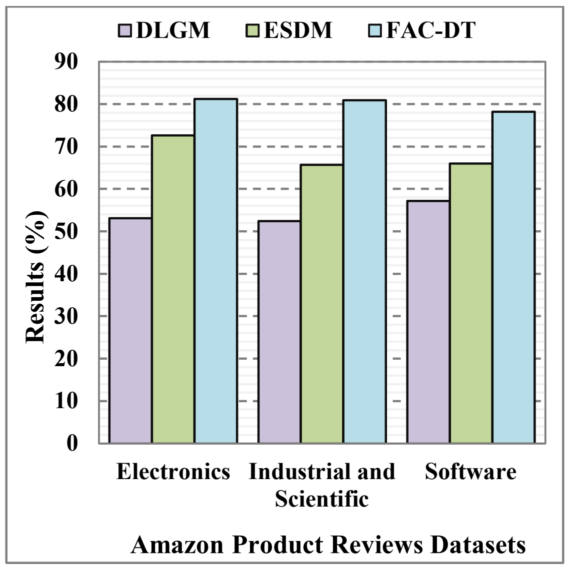

Figure 1. Precision

Table 2. Precision Result Values

DATASET	DLGM	ESDM	FACO-DT
Electronics	53.098	72.652	81.218
Industrial and Scientific	52.435	65.722	80.921
Software	57.175	65.998	78.205

6.2 Recall Analysis

The recall is a pivotal performance measure, highlighting an algorithm's ability to identify all relevant positive sentiment instances within a dataset. It addresses the potential of missing positive instances, thus minimizing false negatives.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The recall values depicted in Figure 2 stem from the data presented in Table 3, which outlines recall scores for each algorithm across three Amazon product review datasets: Electronics, Industrial and Scientific, and Software.Figure 2 offers a comparative view of recall scores, a critical sentiment analysis metric derived from evaluating three sentiment analysis algorithms: DLGM, ESDM, and FACO-DT. The assessment is conducted within the domain of sentiment analysis, which involves gauging the algorithms' proficiency in identifying positive sentiment instances accurately.

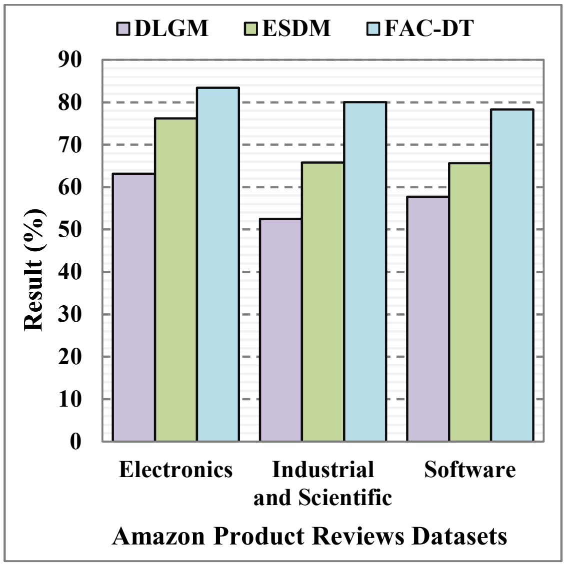

Figure 2. Recall

Analyzing the trends from Figure 2, it's evident that DLGM showcases recall scores of 63.189%, 52.534%, and 57.749% for Electronics, Industrial and Scientific, and Software datasets, respectively. ESDM excels with 76.186%, 65.805%, and 65.666%, while FACO-DT consistently demonstrates strong recall abilities with scores of 83.414%, 80.019%, and 78.328%. This visual representation in Figure 3 aids in understanding the algorithms'performance in capturing positive sentiments across various domains, thereby assisting researchers and practitioners in making informed choices when selecting an appropriate sentiment analysis model for specific applications.

Table 3. Recall

DATASET	DLGM	ESDM	FACO-DT
Electronics	63.189	76.186	83.414
Industrial and Scientific	52.534	65.805	80.019
Software	57.749	65.666	78.328

6.3. Classification Accuracy Analysis

Figure 3 provides a comparative view of classification accuracy scores, a fundamental measure in sentiment analysis. The analysis is grounded in the context of three prominent sentiment analysis algorithms: FACO-DT, DLGM, and ESDM. These algorithms were evaluated using datasets within the field of sentiment analysis, which focuses on understanding and categorizing sentiments expressed in textual data.

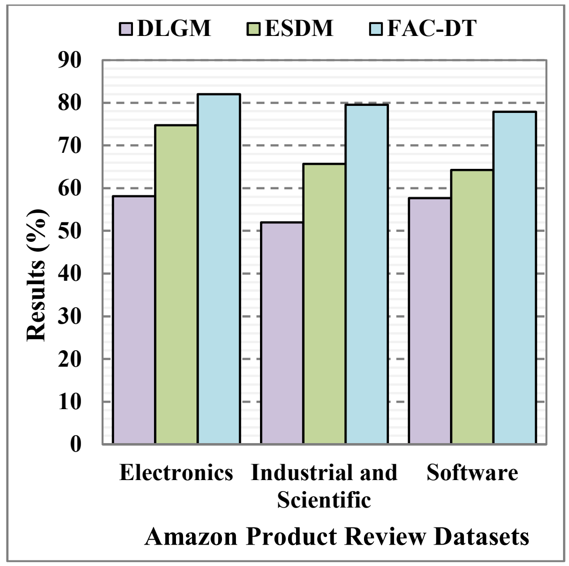

Figure 3. Classification Accuracy

As depicted in Figure 3, classification accuracy is a pivotal performance indicator. It gauges how accurately the algorithms can classify sentiments, reflecting their overall effectiveness in sentiment analysis tasks. The classification accuracy values presented in Figure 3 are extracted from the comprehensive data detailed in Table 4. This table comprehensively summarises classification accuracy scores for each algorithm across three dataset categories: Electronics, Industrial and Scientific, and Software.

Table 4. Classification Accuracy

DATASET	DLGM	ESDM	FACO-DT
Electronics	58.110	74.748	82.005
Industrial and Scientific	51.976	65.701	79.514
Software	57.638	64.271	77.852

DLGM registers classification accuracy scores of 58.110%, 51.976%, and 57.638% for the Electronics, Industrial and Scientific, and Software domains, respectively. DLGM's strength lies in its

Journal of Theoretical and Applied Information Technology
31st October 2023. Vol.101. No 20

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ability to decipher intricate linguistic relationships and contextual nuances. Its performance underscores its holistic approach to sentiment analysis, aiding in accurate sentiment categorization. ESDM showcases classification accuracy scores of 74.748%, 65.701%, and 64.271% across Electronics, Industrial, Scientific, and Software datasets. ESDM's success is attributed to its profound understanding of sentence structures, allowing it to capture sentiment nuances based on grammatical dependencies. This linguistic insight bolsters its accuracy in sentiment classification. FACO-DT yields classification accuracy scores of 82.005%, 79.514%, and 77.852% for the Electronics, Industrial, Scientific, and Software domains, respectively. FACO-DT's exceptional performance arises from its synergistic blend of optimization techniques and structured decision-making. The algorithm effectively refines features through optimization and captures sentiment complexities through decision trees, culminating in high classification accuracy.

Figure 3 visually encapsulates the classification accuracy achievements of DLGM, ESDM, and FACO-DT algorithms across distinct sentiment analysis datasets. Each algorithm's unique strengths, be it DLGM's holistic language understanding, ESDM's syntactic insight, or FACO-DT's optimization and decision-making fusion, contribute to their distinctive accuracy profiles. This visual insight empowers researchers and practitioners to make informed decisions, selecting the most suitable sentiment analysis algorithm tailored to the intricacies of their specific dataset.

6.4. F-Measure Analysis

F-measure, as portrayed in Figure 4, combines the significance of precision and recall, offering a balanced assessment of an algorithm's performance. It gauges how well an algorithm balances correctly identifying positive instances (precision) and capturing all positive instances (recall). The F-measure values illustrated in Figure 4 are extracted from Table 5, which meticulously delineates F-measure scores for each algorithm across three distinct Amazon product review datasets: Electronics, Industrial and Scientific, and Software.

Figure 4 presents a comparative visual representation of F-measure scores, a comprehensive sentiment analysis metric derived from the data provided in Table 5. The focus is on three prominent sentiment analysis algorithms: Fervent Ant Colony Optimization-based Decision Trees (FACO-DT), Disentangled Linguistic Graph Model (DLGM), and Effective Syntactic Dependency Model (ESDM). These algorithms underwent rigorous evaluation within the scope of sentiment analysis research aimed at understanding and categorizing sentiments within textual data.

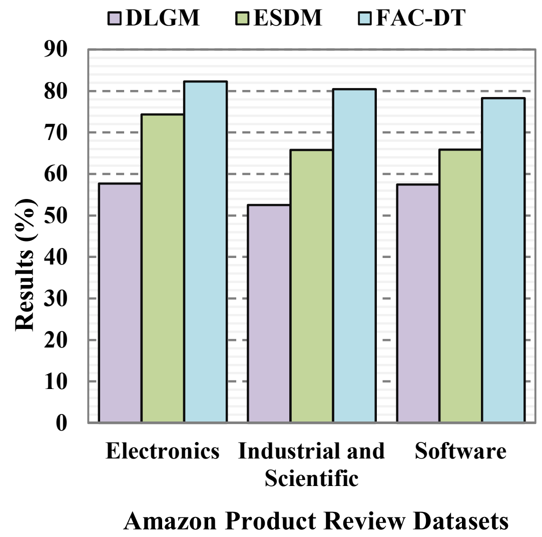

Figure 4. F-Measure

DLGM showcases F-measure scores of 57.706%, 52.484%, and 57.461% for Electronics, Industrial and Scientific, and Software datasets, respectively. DLGM's strength lies in its capacity to capture intricate linguistic relationships, contributing to its balanced performance. However, domain fluctuations suggest domain-specific complexities' influence on its performance. ESDM achieves F-measure scores of 74.377%, 65.763%, and 65.832% across the Electronics, Industrial, Scientific, and Software domains, respectively. ESDM's prowess in deciphering syntactic dependencies within sentences underpins its ability to balance precision and recall harmoniously. FACO-DT registers F-measure scores of 82.301%, 80.467%, and 78.266% for Electronics, Industrial and Scientific, and Software datasets, respectively. FACO-DT's robust performance results from its synergy between optimization techniques and decision-making strategies, enabling a well-balanced trade-off between precision and recall.

Table 5. F-Measure

DATASET	DLGM	ESDM	FACO-DT
Electronics	57.706	74.377	82.301
Industrial and Scientific	52.484	65.763	80.467
Software	57.461	65.832	78.266

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure 4 visually encapsulates the F-measure achievements of DLGM, ESDM, and FACO-DT algorithms across distinct sentiment analysis datasets. These algorithms' strengths, be it DLGM's linguistic insight, ESDM's syntactic understanding, or FACO-DT's optimization-driven decision-making, contribute to their unique F-measure profiles. This visual insight empowers researchers and practitioners to make informed decisions when selecting a sentiment analysis algorithm aligned with their specific dataset's characteristics.

7. CONCLUSION

The evolution of online shopping has elevated the significance of customer reviews as an essential source of information for potential buyers. These reviews' diverse and complex sentiments present formidable challenges for accurate sentiment classification. Traditional approaches often fail to capture the intricate nuances of sentiment embedded within the context of online product reviews. The proposed "Fervent Ant Colony Optimization-based Decision Trees (FACO-DT)" approach has introduced a novel and effective solution to this problem. By synergizing the optimization capabilities of Ant Colony Optimization with the structured decision-making framework of Decision Trees, FACO-DT achieves remarkable improvements in sentiment classification accuracy. This hybrid approach optimizes feature selection and captures the complex sentiment hierarchies prevalent in online reviews, enhancing the overall sentiment analysis process. Through evaluation of an Amazon product review dataset spanning different domains, FACO-DT consistently outperforms traditional methods. Its adaptability to diverse domains and superior performance metrics underscore its potential for practical applications in sentiment analysis. In the dynamic landscape of online commerce, where informed decision-making is crucial, the accuracy and depth of sentiment classification offered by FACO-DT hold great promise in aiding consumers and businesses. In a world driven by online interactions, the FACO-DT approach paves the way for improved sentiment understanding and informed decision-making in online shopping, thereby contributing to a more satisfying and productive online purchasing experience. Future enhancements in this domain hold the potential for exciting developments. Refinements in feature optimization strategies and decision tree structures could further amplify FACO-DT's accuracy.

REFERENCES:

[1] M. Polignano, V. Basile, P. Basile, G. Gabrieli, M. Vassallo, and C. Bosco, "A hybrid lexicon-based and neural approach for explainable polarity detection," *Inf. Process. Manag.*, vol. 59, no. 5, 2022, doi: 10.1016/j.ipm.2022.103058.

[2] M. C. Pereira, J. C. Ferreira, S. Moro, and F. Gonçalves, *University Digital Engagement of Students*, vol. 13198 LNCS. 2022. doi: 10.1007/978-3-030-98388-8_33.

[3] S. Salawu, Y. He, and J. Lumsden, "Approaches to Automated Detection of Cyberbullying: A Survey," *IEEE Trans. Affect. Comput.*, vol. 11, no. 1, pp. 3–24, 2020, doi: 10.1109/TAFFC.2017.2761757.

[4] K. Quoc Tran, A. Trong Nguyen, P. G. Hoang, C. D. Luu, T. H. Do, and K. Van Nguyen, "Vietnamese hate and offensive detection using PhoBERT-CNN and social media streaming data," *Neural Comput. Appl.*, vol. 35, no. 1, pp. 573–594, 2023, doi: 10.1007/s00521-022-07745-w.

[5] J. Wang and C. Li, "Research on User Satisfaction of Video Education Application Based on Reviews," in *ICEIEC 2020 - Proceedings of 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication*, 2020, pp. 340–343. doi: 10.1109/ICEIEC49280.2020.9152252.

[6] L. Moraes, R. M. Marcacini, and R. Goularte, "Video Summarization using Text Subjectivity Classification," in *ACM International Conference Proceeding Series*, 2022, pp. 133–141. doi: 10.1145/3539637.3556998.

[7] J. Ramkumar and R. Vadivel, "Improved frog leap inspired protocol (IFLIP) – for routing in cognitive radio ad hoc networks (CRAHN)," *World J. Eng.*, vol. 15, no. 2, pp. 306–311, 2018, doi: 10.1108/WJE-08-2017-0260.

[8] J. Ramkumar and R. Vadivel, "Performance Modeling of Bio-Inspired Routing Protocols in Cognitive Radio Ad Hoc Network to Reduce End-to-End Delay," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 1, pp. 221–231, 2019, doi: 10.22266/ijies2019.0228.22.

[9] R. Jaganathan and V. Ramasamy, "Performance modeling of bio-inspired routing protocols in Cognitive Radio Ad Hoc Network to reduce end-to-end delay," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 1, pp. 221–231, 2019, doi: 10.22266/IJIES2019.0228.22.

[10] R. Jaganathan and R. Vadivel, "Intelligent Fish

Swarm Inspired Protocol (IFSIP) for Dynamic Ideal Routing in Cognitive Radio Ad-Hoc Networks," *Int. J. Comput. Digit. Syst.*, vol. 10, no. 1, pp. 1063–1074, 2021, doi: 10.12785/ijcds/100196.

[11] R. Vadivel and J. Ramkumar, "QoS-enabled improved cuckoo search-inspired protocol (ICSIP) for IoT-based healthcare applications," *Inc. Internet Things Healthc. Appl. Wearable Devices*, pp. 109–121, 2019, doi: 10.4018/978-1-7998-1090-2.ch006.

[12] J. Ramkumar and R. Vadivel, "CSIP—cuckoo search inspired protocol for routing in cognitive radio ad hoc networks," in *Advances in Intelligent Systems and Computing*, 2017, vol. 556, pp. 145–153. doi: 10.1007/978-981-10-3874-7_14.

[13] M. Lingaraj, T. N. Sugumar, C. S. Felix, and J. Ramkumar, "Query aware routing protocol for mobility enabled wireless sensor network," *Int. J. Comput. Networks Appl.*, vol. 8, no. 3, pp. 258–267, 2021, doi: 10.22247/ijcna/2021/209192.

[14] J. Ramkumar and R. Vadivel, "Whale optimization routing protocol for minimizing energy consumption in cognitive radio wireless sensor network," *Int. J. Comput. Networks Appl.*, vol. 8, no. 4, pp. 455–464, 2021, doi: 10.22247/ijcna/2021/209711.

[15] J. Ramkumar, S. Samson Dinakaran, M. Lingaraj, S. Boopalan, and B. Narasimhan, "IoT-Based Kalman Filtering and Particle Swarm Optimization for Detecting Skin Lesion," 2023, pp. 17–27. doi: 10.1007/978-981-19-8353-5_2.

[16] R. J, "Meticulous Elephant Herding Optimization based Protocol for Detecting Intrusions in Cognitive Radio Ad Hoc Networks," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 8, pp. 4548–4554, 2020, doi: 10.30534/ijeter/2020/82882020.

[17] J. Ramkumar and R. Vadivel, "Multi-Adaptive Routing Protocol for Internet of Things based Ad-hoc Networks," *Wirel. Pers. Commun.*, vol. 120, no. 2, pp. 887–909, Apr. 2021, doi: 10.1007/s11277-021-08495-z.

[18] J. Ramkumar, "Bee inspired secured protocol for routing in cognitive radio ad hoc networks," *Indian J. Sci. Technol.*, vol. 13, no. 30, pp. 2159–2169, 2020, doi: 10.17485/ijst/v13i30.1152.

[19] J. Ramkumar, C. Kumuthini, B. Narasimhan, and S. Boopalan, "Energy Consumption Minimization in Cognitive Radio Mobile Ad-Hoc Networks using Enriched Ad-hoc On-demand Distance Vector Protocol," *2022 Int. Conf. Adv. Comput. Technol. Appl. ICACTA 2022*, pp. 1–6, Mar. 2022, doi: 10.1109/ICACTA54488.2022.9752899.

[20] P. Menakadevi and J. Ramkumar, "Robust Optimization Based Extreme Learning Machine for Sentiment Analysis in Big Data," *2022 Int. Conf. Adv. Comput. Technol. Appl. ICACTA 2022*, pp. 1–5, Mar. 2022, doi: 10.1109/ICACTA54488.2022.9753203.

[21] A. Senthilkumar, J. Ramkumar, M. Lingaraj, D. Jayaraj, and B. Sureshkumar, "Minimizing Energy Consumption in Vehicular Sensor Networks Using Relentless Particle Swarm Optimization Routing," *Int. J. Comput. Networks Appl.*, vol. 10, no. 2, pp. 217–230, 2023, doi: 10.22247/ijcna/2023/220737.

[22] D. Jayaraj, J. Ramkumar, M. Lingaraj, and B. Sureshkumar, "AFSORP: Adaptive Fish Swarm Optimization-Based Routing Protocol for Mobility Enabled Wireless Sensor Network," *Int. J. Comput. Networks Appl.*, vol. 10, no. 1, pp. 119–129, Jan. 2023, doi: 10.22247/ijcna/2023/218516.

[23] L. Mani, S. Arumugam, and R. Jaganathan, "Performance Enhancement of Wireless Sensor Network Using Feisty Particle Swarm Optimization Protocol," *ACM Int. Conf. Proceeding Ser.*, pp. 1–5, Dec. 2022, doi: 10.1145/3590837.3590907.

[24] J. Ramkumar, K. S. Jeen Marseline, and D. R. Medhunhashini, "Relentless Firefly Optimization-Based Routing Protocol (RFORP) for Securing Fintech Data in IoT-Based Ad-Hoc Networks," *Int. J. Comput. Networks Appl.*, vol. 10, no. 4, p. 668, Aug. 2023, doi: 10.22247/IJCNA/2023/223319.

[25] M. AL-Smadi, M. M. Hammad, S. A. Al-Zboon, S. AL-Tawalbeh, and E. Cambria, "Gated recurrent unit with multilingual universal sentence encoder for Arabic aspect-based sentiment analysis," *Knowledge-Based Syst.*, vol. 261, 2023, doi: 10.1016/j.knosys.2021.107540.

[26] Q. Yan, Y. Sun, S. Fan, and L. Zhao, "Polarity-aware attention network for image sentiment analysis," *Multimed. Syst.*, vol. 29, no. 1, pp. 389–399, 2023, doi: 10.1007/s00530-022-00935-5.

[27] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, and P. Cotae, "Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on COVID-19 vaccination Twitter dataset," *Expert Syst. Appl.*, vol. 212, 2023, doi: 10.1016/j.eswa.2022.118715.

[28] G. D. Pilar, S. B. Isabel, P. M. Diego, and G. Á. José Luis, “A novel flexible feature extraction algorithm for Spanish tweet sentiment analysis based on the context of words,” *Expert Syst. Appl.*, vol. 212, 2023, doi: 10.1016/j.eswa.2022.118817.

[29] Y. Qian, J. Wang, D. Li, and X. Zhang, “Interactive capsule network for implicit sentiment analysis,” *Appl. Intell.*, vol. 53, no. 3, pp. 3109–3123, 2023, doi: 10.1007/s10489-022-03584-3.

[30] M. B. Shelke, D. D. Sawant, C. B. Kadam, K. Ambhure, and S. N. Deshmukh, “Marathi SentiWordNet: A lexical resource for sentiment analysis of Marathi,” *Concurr. Comput. Pract. Exp.*, vol. 35, no. 2, 2023, doi: 10.1002/cpe.7497.

[31] L. Xu, X. Pang, J. Wu, M. Cai, and J. Peng, “Learn from structural scope: Improving aspect-level sentiment analysis with hybrid graph convolutional networks,” *Neurocomputing*, vol. 518, pp. 373–383, 2023, doi: 10.1016/j.neucom.2022.10.071.

[32] A. Pratama, R. I. Alhaqq, and Y. Ruldeviyani, “Sentiment Analysis of the Covid-19 Booster Vaccination Program As a Requirement for Homecoming During Eid Fitr in Indonesia,” *J. Theor. Appl. Inf. Technol.*, vol. 101, no. 1, pp. 248–261, 2023, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85146273659&partnerID=40&md5=4dfcdee44fc836245cceaa1e9ca045c1

[33] X. Mei, Y. Zhou, C. Zhu, M. Wu, M. Li, and S. Pan, “A disentangled linguistic graph model for explainable aspect-based sentiment analysis,” *Knowledge-Based Syst.*, vol. 260, 2023, doi: 10.1016/j.knosys.2022.110150.

[34] B. Huang, J. Zhang, J. Ju, R. Guo, H. Fujita, and J. Liu, “CRF-GCN: An effective syntactic dependency model for aspect-level sentiment analysis,” *Knowledge-Based Syst.*, vol. 260, 2023, doi: 10.1016/j.knosys.2022.110125.

[35] D. Moreno-Blanco, B. Ochoa-Ferreras, F. J. Gárate, J. Solana-Sánchez, P. Sánchez-González, and E. J. Gómez, “Evaluation and Comparison of Text Classifiers to Develop a Depression Detection Service,” in *IFMBE Proceedings*, 2020, vol. 76, pp. 1205–1212. doi: 10.1007/978-3-030-31635-8_146.

[36] J. A. P. Lalata, B. Gerardo, and R. Medina, “A sentiment analysis model for faculty comment evaluation using ensemble machine learning algorithms,” in *ACM International Conference Proceeding Series*, 2019, pp. 68–73. doi: 10.1145/3341620.3341638.

[37] S. Deepa, N. Revathi, K. Sivakami, and P. K. Pareek, “Machine Learning based Education System with Sentiment Analysis for Students,” in *MysuruCon 2022 - 2022 IEEE 2nd Mysore Sub Section International Conference*, 2022. doi: 10.1109/MysuruCon55714.2022.9972555.

[38] Priyanka and K. Walia, “Popularity of Optimization Techniques in Sentiment Analysis,” *Lecture Notes in Networks and Systems*, vol. 494. pp. 523–534, 2023. doi: 10.1007/978-981-19-4863-3_52.

[39] S. Kumar and R. Singh, “Design of ensemble classifier selection framework based on ant colony optimization for sentiment analysis and opinion mining,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 12, pp. 5112–5117, 2019, doi: 10.35940/ijitee.L2755.1081219.