I.J. Intelligent Systems and Applications, **2026, 2, 13-26**
Published Online on April 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2026.02.02

Integrated Topic Modeling and Feature Engineering for High-accuracy Sentiment Classification in Consumer Reviews

Vijay Gupta
School of Computer Application, Lovely Professional University, Phagwara-144411, Punjab, India
E-mail : vijay.gupta@lpu.co.in
ORCID iD: https://orcid.org/0000-0002-3232-6934

Punam Rattan
Department of Computer Science & Technology, Manav Rachna University, Faridabad-121004, Haryana, India
E-mail : punamrattan@gmail.com
ORCID iD: https://orcid.org/0000-0002-3232-6983

Mukesh Kumar*
Advanced Centre of Research & Innovation (ACRI), School of Advance Computing, CGC University Mohali-140307, Punjab, India
Faculty of Law, Sohar University, Sohar 311, Sultanate of Oman
E-mail: mukesh.kumarphd2014@gmail.com
ORCID iD: https://orcid.org/0000-0001-8797-9810
*Corresponding author

Received: 14 October 2025; Revised: 22 December 2025; Accepted: 26 February 2026; Published: 08 April 2026

Abstract: The rapid rise of mobile technology paired with the steady growth of the internet, has led to a massive increase in the amount of user generated content, such as online consumer reviews, accessible through the browser. As the volume of user-generated content continues to rise, it becomes increasingly important to develop sophisticated methods for performing sentiment analysis on the texts collected from users, especially those that have been generated in relation to restaurants and similar types of service establishments. In this paper, we will present a new approach to sentiment analysis which incorporates Latent Dirichlet Allocation topic models, Term Frequency- Inverse Document Frequency vector representations and XGBoost Classifiers into a unified framework. Unlike conventional implementations, this study integrates probabilistic topic distributions from LDA with multi-level n-gram TF-IDF features and evaluates their combined impact using XGBoost for enhanced classification performance. Using three distinct n-gram levels (unigrams, bigrams, and trigrams), we will evaluate various aspects of text-based data including common linguistic patterns and sentiment trends. Higher-order n-grams were included to capture contextual dependencies beyond single-word features. Overall, our results demonstrate that the performance of our proposed framework is superior to traditional corpus-based models on multiple evaluation metrics, including: classification accuracy 96.07%, classification sensitivity 95.43%, classification specificity 97.12% and F1-Score 96.16%.

Index Terms: Latent Dirichlet Allocation, Term Frequency-Inverse Document Frequency, XGBoost Classifier, Topic Modeling, N-gram Features, Sentiment Analysis.

1. Introduction

Customer feedback has become an extremely important factor influencing a buyer's purchase decision between competing products and services. The rise of Internet and social media technology has greatly influenced how customers share their experiences with others, including people living thousands of miles away. Approximately 90% of the content currently available on the Internet originated from active users of social media platforms, primarily through Web 2.0 and the later introduction of mobile technology. The production of user-generated content creates a greater level of direct interaction among users and encourages users to take a more active role in business and product-related decisions. In fact,

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

according to [2], the consumers' opinion about a company's product on social media is something that many potential buyers consider before making a purchase, often more so than the company's own advertising. The best example of this is the fact that almost 90% of all companies in existence today have an average budget of approximately 11% of their total annual budget earmarked for consumer engagement in social media [1]. Effective evaluation of consumer feedback and providing timely feedback is an important part of establishing and maintaining positive customer relationships. Research indicates that the faster a company responds to its customers on social media, the more likely it is to retain customers over time [2]. Social media and micro blogging have now become the primary sources of customer feedback. Most of the information about products and services used by businesses in the business decision-making process comes from customer input that is sourced through these types of platforms, many of which operate as data aggregators or providers of decision-making data for businesses using customer feedback on products and services. These platforms have also assisted customers when searching for product and service information and have enabled businesses that use these platforms to promote their products and services more effectively [3]. In addition, businesses can use consumer-generated feedback to improve their products and services.

Consumers interacting via internet technology often continue to influence each other to provide their feedback and opinions. This leads to a sharp increase in the number of comments posted to social media and micro blogs about products and services [4]. The authors of [5] refer to this behaviour as "calling to the customer with the customer's voice." Today's consumer also has a keen interest in using social media and micro blogs as a source of information for products and services. The primary reasons for this interest could be that (i) customers will post about their experience of using a brand or product that inspired them as an individual; and (ii) customers are more likely to use text to express themselves than they are to use voice or speech. Therefore, since writing is typically a better medium of expression in social media and micro blogs, much of the content shared by consumers is represented purely through text. The consumer created most of the textual content in their native language(s); thus, referring to the textual evidence of consumer opinion in these forms of media is equivalent to assigning the consumer's needs to the production process. Therefore, various data mining-based methods have been developed to allow for the systematic review of the respondents' opinions gathered through social media and micro blogs. In fact, the cost and difficulty associated with attempting to measure the opinions of individuals using traditional means of measurement is an ongoing issue [6]. According to the authors of [7], sentiment analysis represents a very popular research area in both academic and industrial work. The authors of [8] note that sentiment analysis can be applied at the document, sentence, or target level. In addition, this analysis typically uses multiple classification algorithms, as the target term can be expressed in multiple ways [8]. This analysis has become a common technique in reviewing companies through user reviews, as consumer sentiment can be directed at multiple targets, such as restaurants and hotels. When using the above method, the system receives input from users' comments about restaurants and summarizes them based on the characteristics of product features that might affect a given business (i.e., food, service, price) [9].

The primary objective of this study is to develop an integrated sentiment analysis framework that combines topic modeling and feature engineering techniques to enhance classification performance on restaurant reviews. Specifically, the work aims to investigate the effectiveness of combining LDA-based topic distributions with multi-level n-gram TF-IDF features using an XGBoost classifier.

The remainder of this paper is organized as follows: Section III presents the proposed methodology, including data preprocessing, feature extraction, and classification approach. Section IV discusses the experimental results and performance evaluation. Finally, Section V concludes the paper and outlines potential directions for future research.

2. Literature Review

Recent advancements in mobile device technology, internet access, and access to information via these platforms have given individuals the ability to express their feelings and thoughts through audio, visual and written media and to utilize various communication tools to do this. The researchers have identified different types of forms, and that the original form will correspond to a specific platform. For example, sharing on virtual platforms (e.g., forum discussions or blog posts) is always in writing. The publishing of Consumer-Generated Content (CGC), including product reviews, restaurant reviews and film reviews, has generated a lot of interest in the community and attracted the attention of many (e.g., reviewers) who write reviews and/or provide reviews to the public. These reviews are useful to others because they provide a good indicator of the satisfaction or dissatisfaction of other users with respect to the product being reviewed, the meal being reviewed or the film being reviewed. In addition, the fact that CGC provides true data to be utilized for research purposes in sentiment analysis also provides the researcher with the opportunity to conduct these data analyses for the purpose of identifying patterns or bias of one or more of these forms of CGC. As the desire of researchers to conduct data analysis of these types of data set continues to grow, it has become progressively more difficult to manually conduct these types of analyses due to the increasing numbers associated with the number of records that one can find in these types of data sets. As a result, researchers now must rely on the automated analysis of CGC (computer automated analysis of comments on a continuous or intermittent basis) because of the inability to manually analyze these massive amounts of data associated with CGC. Consequently, the demand for various algorithms associated with machine learning, artificial intelligence, and data mining continues to rise. In recent years, several algorithms of other types have been developed using machine learning, artificial intelligence, and data mining to extract hidden patterns from CGC, and

analyse the data to make predictions about the data in the future. Target Term Extraction (TTE) is one of the core ABSA subtasks, and it means to retrieve the target term(s) corresponding to different views expressed in different documents [10]. Identifying the target terms is the foundation of TTE, which is the first step in Target-Based Sentiment Analysis. According to the authors of [11], there are two key issues regarding identifying the Target Term, as it is a requirement for sentiment analysis. The first issue is the inability of sentence-level sentiment analysis to provide accurate or qualified information [12]. The second issue is that in many sentences or paragraphs, there are targets with different polarities. Therefore, summarising the single sentence or paragraph will be challenging [13]. The target terms must be obtained from the text corpus for both problems. An example is the sentence "The pizza was delicious, but the service was terrible." Two target terms, "pizza" and "service," may be extracted from this example. Most of the TTE Research issues in ABSA are based on machine learning research at the three levels of learning: supervised, semi-supervised, and unsupervised [14]. Overall, it appears that most of the machine learning TTE studies have favoured the traditional supervised machine learning methods. Target Extraction Studies using ROI Classification Method were performed in studies with a HAST [12] as a classification type in a supervised study of target extraction; CL-BERT [15]; CMLA [16]; and DE-CNN [17]. History Attention and Selective Transformation (HAST) represent the target term using a historical order of the comments among others. CL-BERT is the version of BERT (which is a pre-trained Natural Language Processing method) that uses a Constituency Lattice Method in combination with the attention mechanism to improve the performance of Translate Machine Learning using the attention mechanism on the neural machines compared to traditional approaches to linguistic processing. As a result, the decoders can access the target sequence as they process each of the hidden states, helping to minimize the impact of some hidden states on each decoded target term. Coupled Multi-Layer Attentions (CMLA), as a method for identifying target term attention (or aspect attention) and opinion attention, leverages multi-layer attentional pairs to create a representation of the target terms and/or aspects for which it is possible to identify and match.

The Task Two for ABSA involves determining whether the target categories provided in the assigned sentence belongs to the pre-defined category sets specific to the domain. The model must also determine that the target categories will continue the previous example of identifying food and service categories as well. The authors of [18, 19, 20] state that from the perspective of providing a better understanding of consumers; identifying the target categories offers two advantages compared to the target terms used previously. One being that the target term explains well each of the individual terms; on the other, the target category does an overall good job of describing the entire set of terms (the umbrella term [21]), plus can be used to detect the individual target term within the same sentence without using a reference to it being a target term (using [22]). For example, in the example of "The prices are way too much and the food is terrible", the engine will be able to identify that they each have a category of "price" and "food", respectively. The related task of identifying target categories using machine learning techniques has already been studied in due respect to the previously mentioned supervised and unsupervised [23] methods. In the studies, humans were much less represented in the articles in comparison to the first sub-task of Category Detection (CD), with most of them utilizing guided techniques such as Language Independent Category Detection (LICD) and Representation Learning (RepLearn). Additionally, LICD is the only method which allows for the detection of categories without being limited by language and uses a methodology based on the idea that any sentence that falls into a specific category will likely have a high level of semantic similarity with a specific set of keywords related to the category; therefore, categorizing based on a known category, and the determined similarity between the target and the given keywords based on this established framework will yield a representative sample of that specific category. In contrast, RepLearn is a learning method that applies to unlabelled, large data-sets and collects data based on the semantics of each term in relation to one another; therefore, allowing easily for the determination of semantic relationships within large volumes of unlabelled data. While both are guided techniques, many researchers use Self-Supervised Contrastive Learning (SSCL) and contrastive attention (CAT) as a method for categorizing or utilizing unsupervised machine learning techniques. For example, SSCL can help identify the correct representation of a specific segment of a comment by training two alternative representations of a comment to reach a maximum level of agreement between the two representations and thus, producing the appropriate target embedded within the correct representation of the target/embedded comment segment [24]. Another method, CAT, will use a Radial-Basis Function to find and remove information that does not relate to a target product from a comment to produce an appropriate representative sample of the specific category of the corresponding comment segment.

Thus, these methods are consistent with previous studies using target-oriented emotion word extraction preferred, IOG [25], maxSAT (maximum satisfiable clause) [26], Position Embed [27], and ONG [28] methods and/or focused on the second or positional dimension. The first method for which the DeepWmaxSAT has been developed has not previously been described as a combination of information on deep weighting as defined to generate a target term extraction and/or the maxSAT (maximum satisfiable clause) with the intent to maximize the number of maximum semantic-coherency sentences through mathematical means to generate maximum satisfiability clauses. The distance (i.e., relative position) of the words embedded at varying distances from the targeted emotions, will be considered through the target emotions, referred to in previously disclosed Position Embed methods, in relation to the mean distance between the emotional embedded word(s), and, therefore, the proximity of the target emotional word(s). According to IOG's, their model is an LSTM neural network. The authors compared the corpus of text with the target terms being present (inward and outward) versus not being present in their corpus (with no target term). The authors then analysed the dataset regarding their expressions in the global context with respect to the target emotion being inferred from the dataset values.

There are other studies in the literature that combine the first three subtasks all together to derive target-based analysis from the reviews written in Czech and English Restaurant reviews [29]. Using models that were created to

perform sentiment analysis using HAL, CBOW, COALS, LDA, SKIP, and GloVe to generate results from two datasets (1. Reviews about laptops and 2. Reviews about restaurants) based on the data, the F1 scores were found to be 75.66% and 71.48% for the Czech and English results in identifying the target terms and 77.68% and 71.68% for the target categories. The accuracy of the sentiment polarity values for the same datasets was measured at 67.48% and 69.68% respectively. Another study [30] utilized a machine learning approach for completing four subtasks within one data source in conjunction with another data source in terms of restaurant and laptop data provided for SemEval-2014.Using structured learning, the UNITOR method was both applied to the laptop and restaurant data, obtaining F1 Score results of 75% and 84%, respectively, for detecting target-term, and all other results indicating an accuracy of 80.95% and 70.48%, respectively, for determining whether a given target term has positive or negative polarity. Next, to determine target category, only using restaurant data, the authors reported that they reached maximum F1 score performance of 89%. Past literature has shown that sentiment analysis has been mainly used to classify whether a given expression belongs to a positive or negative class within a specific dataset. Sentiment analysis gives individuals the knowledge they need to make informed decisions about which products to purchase, etc. In this research study, The findings of the research study showed that the classifier model performed well when compared to the high dimensional data, as well as produced a high accuracy level for performing a sentiment analysis of restaurant reviews. The remaining sections of this paper will provide further discussion of the enhancements that using topic modeling has made to the performance metrics of machine learning algorithms (i.e., accuracy and precision levels) when the two features (i.e., topic modeling feature and word feature) are combined using the TF-IDF vectorization technique. The most significant contribution our work has made to the research study compared to other published studies is our ability to represent the data at the semantic topic level (with respect to the term importance). Additionally, we have thoroughly evaluated the differences between the two datasets by providing a comprehensive evaluation of the XGBoost Classifier against the associated code and thus the differences from the topic analysis distributions. In Section III of the paper, we discuss the proposed methodologies for the study, including how to pre-process the data, how to model the topics, and how to extract features (topic and word) from the data. In Section IV, we visually report the results of our research study to show how the performance of the classifier was enhanced with the addition of the combined feature sets. Section V concludes with a summary of the research study and recommendations for future research directions.

3. Proposed Methodology

This proposed methodology aims to improve the sentiment analysis performance over restaurant reviews through a series of multi layered approach which is Topic Modeling using Latent Dirichlet Allocation (LDA) for feature extraction through Term Frequency Inverse Document frequency (TF-IDF) and classification through XGBoost. In this manner, the framework is going to help in gaining better understanding of customer sentiments using advanced data mining techniques and machine leaning classifiers. LDA is used to extract latent semantic topic distributions which complement TF-IDF features by capturing thematic structure beyond surface word frequency. Clearly, the flow diagram of the proposed methodology is shown in Fig. 1.

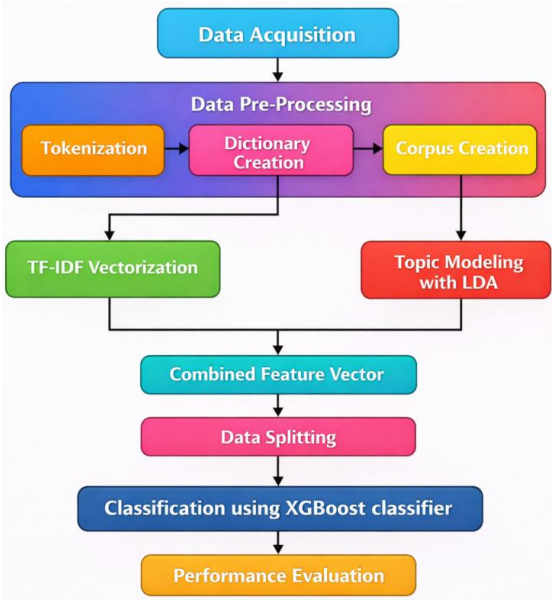


Fig.1. Overall architecture of the proposed LDA–TF-IDF–XGBoost sentiment classification framework

Each step is detailed in the following sections and the theoretical foundation and mathematical formulations are included for each step.

3.1. Data Acquisition and Preprocessing

The study uses the Kaggle Restaurant Review Dataset *(https://www.kaggle.com/datasets/joebeachcapital/restaurant-reviews)* consisting of 10000 reviews, with a balanced distribution of positive and negative classes. restaurant reviews which are in form of textual feedback from consumers need to be acquired and this process requires some level of preprocessing. $D = \{R_1, R_2, \ldots, R_n\}$, where each review R_i comprises of the sentiment label y_i ($y_i \in \{0,1\}$, negative (0) or positive (1) sentiment. Standard preprocessing including tokenization, stopword removal, lemmatization, and cleaning was applied before feature extraction.

A. Tokenization and Data Cleaning

The process of cleaning and tokenization of raw text data is done such that it is ready for further analysis. In this step, unnecessary characters are removed and the text is broken into individual units (tokens).

$$\text{Cleaning Formula:} \quad \text{Cleaned Text}_\text{i} = \text{RemoveHTML}\big(\text{RemovePunctuation}(\text{Review}_\text{i})\big) \tag{1}$$

The function Review_i is the original text with HTML removed using the function RemoveHTML and punctuation removed, using the function RemovePunctuation.

B. Lemmatization

To make the analysis more efficient lemmatization is done to reduce words to their root forms. For example, "running" becomes "run". This guarantees the same form of the words will be treated as equivalent.

$$\text{Mathematical Representation: } Lemma(t) = t_{root},\ t_{root} \in \{run, play, eat, \ldots\} \tag{2}$$

Given t a token in the review, and t_{root} is its associated root form.

3.2. Feature Extraction: LDA and TF-IDF

The second stage is the transformation of the text data into numerical format which is suitable for the machine learning model. The two techniques used are LDA for topic modelling and TF-IDF for the feature representation. Although LDA and TF-IDF are statistically related, LDA captures semantic topic distributions while TF-IDF captures term importance. Their combination improves representation richness.

A. Latent Dirichlet Allocation (LDA) for Topic Modeling

Probability model, Latent Dirichlet Allocation (LDA), is adopted to discover hidden thematic structures (topics), which are underlying in the reviews. Assuming that each review is a mixture of topics, it takes that each topic can be modelled by a word distribution. The topic distributions are modelled using Dirichlet distributions out of the LDA model. Mathematical formulation is given by assuming the corpus $D = \{D_1, D_2, \ldots, D_n\}$ as the set of n reviews denoted by sequences of words w_i. It is assumed that each review D_i is generated from a mixture of K topics $\{z_1, z_2, \ldots, z_K\}$.

Topic-Document Distribution: θ_i is the topic distribution for each review D_i, and it follows a Dirichlet distribution.

$$\theta_i \sim Dirichlet(\alpha) \tag{3}$$

Where α is a parameter which controls sparsity of topic assignments.

Topic-Word Distribution: The vocabulary has an associated word distribution ϕ_k per topic z_k. A Dirichlet prior models this distribution.

$$\phi_k \sim Dirichlet(\beta) \tag{4}$$

Where β sets the sparsity of word assignments within each topic.

Generative Process: The way words are generated in a review is a topic distribution θ_i is sampled from a Dirichlet distribution under each review D_i. For each word w_{ij} in review D_i, a topic z_{ij} is sampled from θ_i, and generate a word from $\phi_{z_{ij}}$.

$$z_{ij} \sim Multinomial(\theta_i), w_{ij} \sim Multinomial\left(\phi_{z_{ij}}\right) \tag{5}$$

LDA's output is a matrix of topic distributions θ_i for each review D_i where each θ_i is a probability vector over K topics.

B. TF-IDF Vectorization for Feature Representation

To measure the importance of each word in the document relative to the whole corpus, TF-IDF (Term Frequency-Inverse Document Frequency) is applied. The aim is to increase the value of the words which are prevalent in some specific document but rare over the whole corpus.

Term Frequency (TF): The term frequency of a word t in a review D_i is given by:

$$TF(t, D_i) = \frac{Count\ of\ t\ in\ D_i}{Total\ number\ of\ terms\ in\ D_i} \tag{6}$$

Inverse Document Frequency (IDF): The term t with IDF is calculated as:

$$IDF(t) = \log\left(\frac{N}{df(t)}\right) \tag{7}$$

Where N is total number of documents, and $df(t)$ is number of documents containing term t.

TF-IDF Calculation: Define the TF-IDF score of a word t in a document D_i as:

$$\text{TF-IDF}(t, D_i) = TF(t, D_i) \times IDF(t) \tag{8}$$

3.3. Combining LDA and TF-IDF Features

Both the LDA topic distributions and TF IDF features are concatenated to get further enhance the feature representation. The use of LDA and TF-IDF enables us to create a richer feature vector that contains both the semantic topics of the reviews as specified by LDA and the importance of individual words according to TF-IDF.

Combined Feature Representation: We denote F_{LDA} as the $n \times K$ matrix of LDA topic distributions, and F_{TFIDF} as the $n \times T$ TF-IDF matrix, where T is the vocabulary size. These two matrices are concatenated to obtain the combined feature matrix $F_{combined}$.

$$F_{combined} = [\, F_{LDA} \mid F_{TFIDF} \,] \tag{9}$$

$F_{combined}$ is combined to have size $n \times (K + T)$ combining both topical and term-based feature.

3.4. Classification using XGBoost

With the features extracted, the sentiment is classified using XGBoost (Extreme Gradient Boosting). The dataset was divided into training and testing sets using an 80:20 split. XGBoost is a gradient boosting algorithm that makes use of decision tree as its base learner, to optimize the objective function.

A. Model Initialization and Objective Function

XGBoost has an objective function made up of two parts: a loss function and regularization term. The error is measured by a loss function based on the difference between predicted and true labels, and regularization term acts to prevent overfitting. The objective function is mathematically represented as:

$$L(\theta) = \sum_{i=1}^{n} Loss\big(y_i, f(x_i)\big) + \lambda \sum_{j=1}^{m} \theta_j^2 \tag{10}$$

Where:

- y_i is the true sentiment for review i,
- $f(x_i)$ is the predicted sentiment for review i,
- λ is the regularization parameter,
- θ_j is a parameter of the decision tree.

Loss Function: For binary classification, the loss function is usually taken to be logistic loss:

$$Loss\big(y_i, f(x_i)\big) = log(1 + e^{-y_i f(x_i))} \tag{11}$$

B. Training the Model

The XGBoost model is an iterative procedure in which it adds one decision tree at a time to minimize the loss function. The method is to add a new tree at each iteration to correct the errors of previous discovered trees. The gradient

and Hessian of the loss function are used to inform the tree building process.

$$\textbf{\textit{Gradient Calculation:}}\ Gradient_i = \frac{\partial Loss}{\partial f(x_i)} \tag{12}$$

$$\textbf{\textit{Hessian Calculation:}}\ Hessian_i = \frac{\partial^2 Loss}{\partial f(x_i)^2} \tag{13}$$

Tree Update: The computed gradients and Hessians are used to update each tree in turn.

$$f_{new}(x_i) = f_{old}(x_i) + \eta \times Gradient_i \tag{14}$$

Where η symbolizes for the learning rate.

C. Model Evaluation

Various metrics such as accuracy, precision, recall and F1 score are used to evaluate the model performance. Thus, these metrics give an overall idea of the model's ability of classifying sentiments correctly.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{15}$$

TP, TN, FP, and FN are interpreted as counts of true positives, true negatives, false positives, and false negatives, respectively.

$$Precision = \frac{TP}{TP+FP} \tag{16}$$

$$Recall = \frac{TP}{TP+FN} \tag{17}$$

$$\text{F1-Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{18}$$

3.5. Results Visualization

The results of the classification task are visualized such that additional insights about the model's performance and behaviour can be gained.

Sentiment Distribution: The distribution of positive, negative, and neutral sentiments of the dataset is shown using a bar plot.

Feature Importance Visualization: The most influential features (terms or topics) in the prediction of sentiments are shown in a bar chart or word cloud.

4. Results and Discussion

The improvement in classification performance can be attributed to the incorporation of LDA-based topic distributions, which provide a higher-level semantic abstraction of the review content. While TF-IDF captures term importance at the lexical level, LDA identifies latent thematic structures that help distinguish sentiment-bearing patterns across documents. This additional semantic representation enhances feature separability in the feature space, allowing the XGBoost classifier to make more discriminative decisions. As a result, the combined feature representation improves overall robustness and classification accuracy.

4.1. Evaluation Parameters

The evaluation parameters used to assess the performance of sentiment classification models include True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). True Positive (TP) refers to the number of restaurant reviews that are correctly classified as having a positive sentiment. True Negative (TN) denotes the count of reviews that are accurately identified as not possessing a positive sentiment, indicating correct classification of negative or neutral sentiments. False Positive (FP) captures the number of reviews that are mistakenly classified as positive when they convey a negative sentiment. Lastly, False Negative (FN) represents the number of reviews that are incorrectly labeled as negative, even though they genuinely express a positive sentiment. These parameters form the foundation for calculating key performance metrics such as accuracy, precision, recall, and F1-score, which are critical for evaluating the effectiveness of sentiment analysis models.

Table 1. Definition of evaluation parameters used for sentiment classification

Parameters	Description
TP (True Positive)	The number of restaurant reviews accurately classified as having a positive sentiment.
TN (True Negative)	The count of reviews correctly identified as not having the positive sentiment.
FP (False Positive)	The count of reviews misclassified as positive when they have a negative sentiment.
FN (False Negative)	The number of reviews incorrectly classified as negative when they truly have a positive sentiment.

4.2. Comprehensive Visualization of Textual Data Analysis

Fig. 2, shows a bar chart of the most common words (called unigrams) found in a set of reviews, likely about food or drinks. Each bar represents how many times a word appeared. For example, the word "like" was used the most, followed by "good", "taste", "flavor", and "great", which are mostly positive words. This suggests that many reviews had a positive tone. Some words like "coffee", "chip", "tea", and "product" show that the reviews were about food items. One word, "br", seems to be an error or formatting issue. Overall, the chart helps to understand what words people used most often in their reviews.

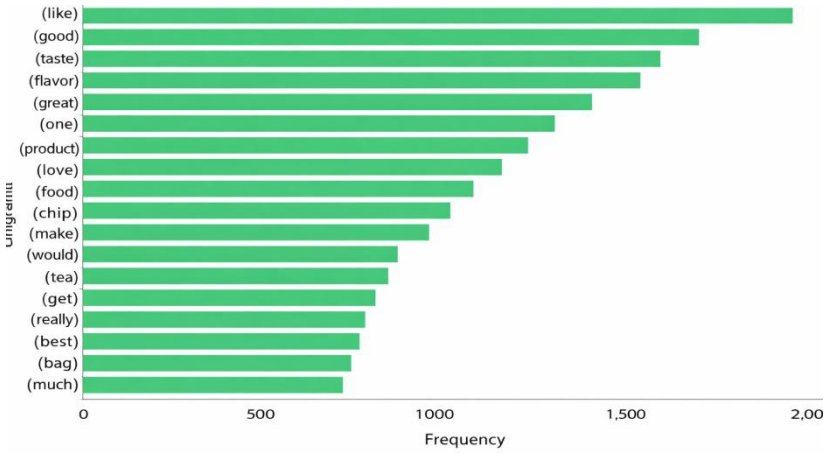


Fig.2. Top 20 most frequent unigrams extracted from positive restaurant reviews

Fig. 3, is a bar chart showing the top 20 unigrams (individual words) that appear most frequently in negative statements within a dataset, likely containing product reviews. Each bar shows how often a word appears in negative reviews. The most common word is "product", followed by "taste", "br" (which is likely a formatting issue), "like", and "one". Words like "bad", "don't", "would", and "dog" suggest dissatisfaction or complaints. Words like "chip", "tea", and "box" show that users were referring to specific food items or packaging. This chart helps to understand which words are often used when customers express negative opinions, which can be useful for improving product quality or customer experience.

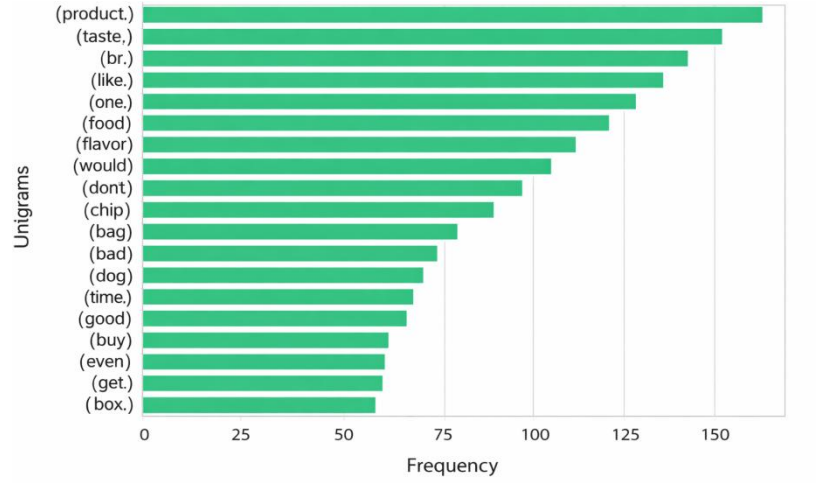


Fig.3. Top 20 most frequent unigrams extracted from negative restaurant reviews

Fig. 4 shows a horizontal bar chart of the top 20 bigrams (pairs of consecutive words) found in positive statements from a dataset, likely customer reviews. This highlights the most frequent word pairs used in reviews expressing positive

sentiment. The most common bigram is "dog food", followed by "gluten free", "grocery store", and "taste like". These combinations suggest customers often praised the food quality (especially pet food), dietary features (like gluten-free), and availability in stores. Bigrams such as "highly recommend", "potato chip", "hot chocolate", and "taste great" reflect favourable opinions about taste and product recommendations. Other positive expressions include "much better", "really good", and "variety pack", which indicate customer satisfaction with flavor, quality, and packaging options.

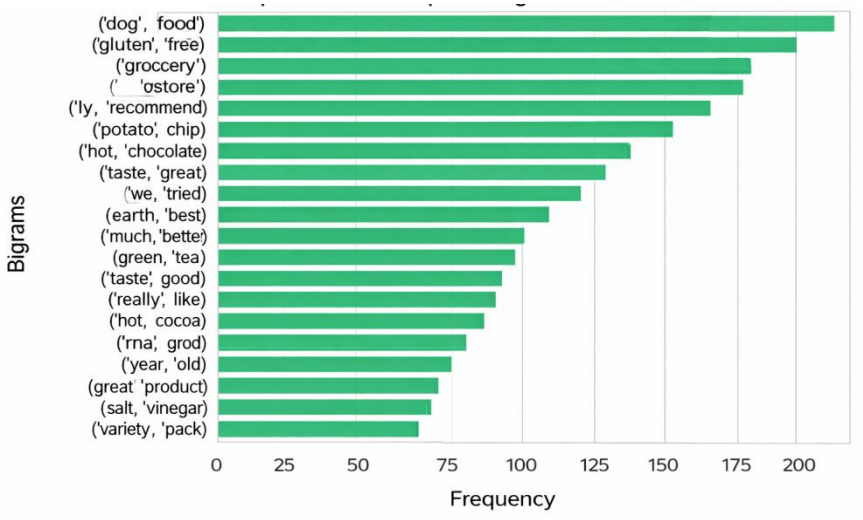


Fig.4. Top 20 most frequent bigrams identified in positive restaurant reviews

Fig. 5, displays a horizontal bar chart of the top 20 bigrams (two-word combinations) found in negative statements, likely from customer reviews. This highlights common word pairs used when customers expressed dissatisfaction. The most frequent bigram is "dog food", which might indicate recurring issues in pet food reviews. Other frequently used negative phrases include "taste like", "cannot", and "don't know", reflecting uncertainty, dislike, or limitations. Bigrams such as "waste money", "never buy", and "don't like" clearly indicate disappointment or regret. Also, phrases like "hot chocolate", "grocery store", "potato chip", and "coconut water" suggest that specific product types were often criticized. Overall, this figure helps identify the most common complaints or concerns among users, offering insights that can help businesses improve their products or services based on customer feedback.

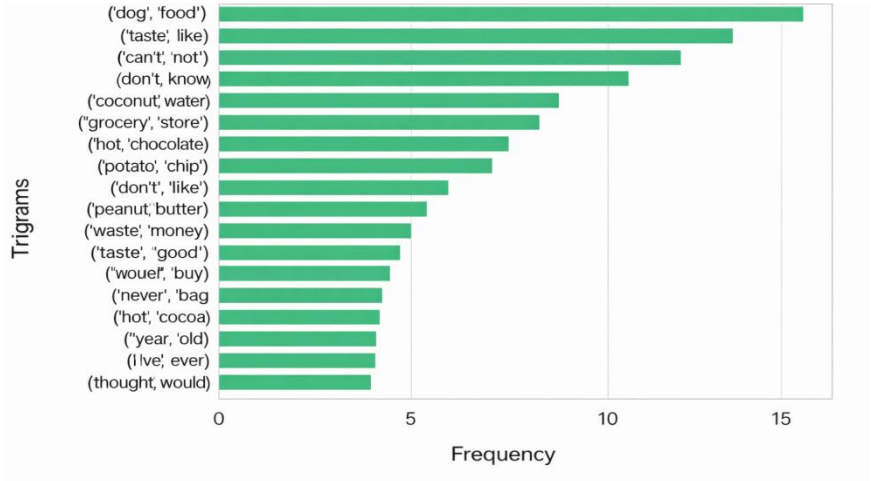


Fig.5. Top 20 most frequent bigrams identified in negative restaurant reviews

Fig. 6, presents a horizontal bar chart showing the top 20 trigrams (three-word combinations) found in positive statements, likely taken from product reviews. This chart highlights the most common trigrams used when people expressed satisfaction. The most frequent trigram is "grocery store", suggesting many users discussed product availability or purchase experience positively. Other popular trigrams include "recommend product", "sea salt vinegar", and "highly recommend chip", showing that users often praised product quality, especially snacks. Trigrams like "can't go wrong", "health food store", and "key lime pie" also appear, indicating appreciation for flavor, health aspects, or specific products. The frequent use of positive action words like "recommend", "like", and "love" shows overall customer satisfaction. This figure helps identify not only what customers liked but also how they described their positive experiences, which can be useful for marketing and product improvement. Overall, the chart helps identify phrases that frequently appear in positive

feedback, offering useful insights into what customers appreciate most.

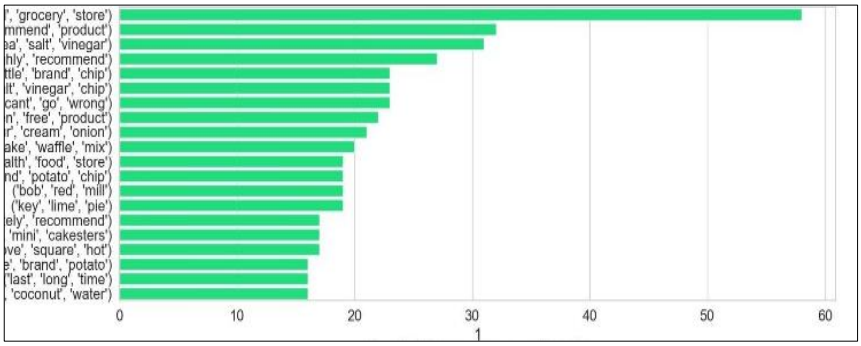


Fig.6. Top 20 most frequent trigrams identified in positive restaurant reviews

Fig. 7, displays a horizontal bar chart of the top 20 trigrams (three-word combinations) found in negative statements, most likely from product reviews. This figure which word combinations are most used when people write negative reviews. The most frequent trigram is "fructose corn syrup", suggesting concern about unhealthy ingredients. Other commonly seen trigrams include "grocery store", "corn syrup", "low carb diet", and "packet per oz", which point to customer dissatisfaction related to ingredients, diet compatibility, or packaging. Phrases like "waste money", "hard find store", and "hydrogenated oil" express frustration with product value, availability, or unhealthy content. Additionally, trigrams such as "even taste good", "least nothing like", and "steak tasty time" show disappointment with taste or quality.

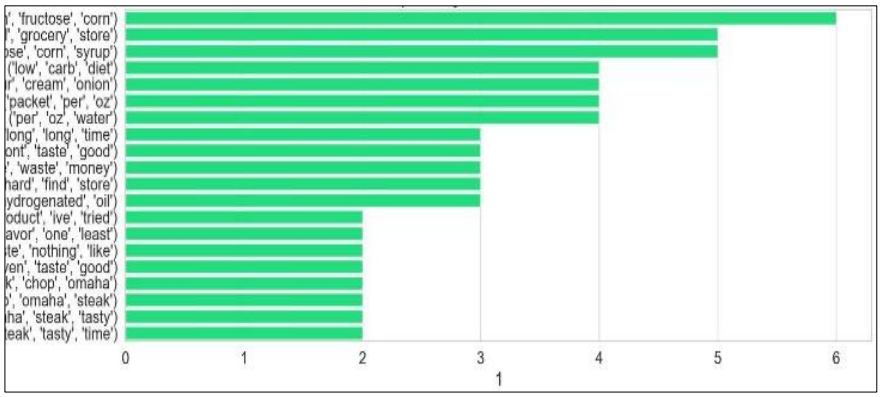


Fig.7. Top 20 most frequent trigrams identified in negative restaurant reviews

Overall, this chart highlights specific phrases customers use when expressing negative feedback, helping businesses understand product areas needing improvement. To describe Fig. 8, titled Positive Sentiment (a), Negative Sentiment (b), and Neutral Sentiment (c) illustrates the distribution of sentiments positive, negative, and neutral found within the analyzed reviews. In subfigure (a), the chart likely highlights that a significant portion of the reviews express positive sentiment, indicated by frequent use of favorable words or phrases. Subfigure (b) focuses on negative sentiment, showing commonly used critical or unfavorable terms that suggest dissatisfaction. Subfigure (c) represents neutral sentiment, which includes reviews that are neither strongly positive nor negative, often containing balanced or factual statements. Together, these visualizations help in understanding the overall emotional tone of the customer feedback and the key expressions associated with each sentiment category.

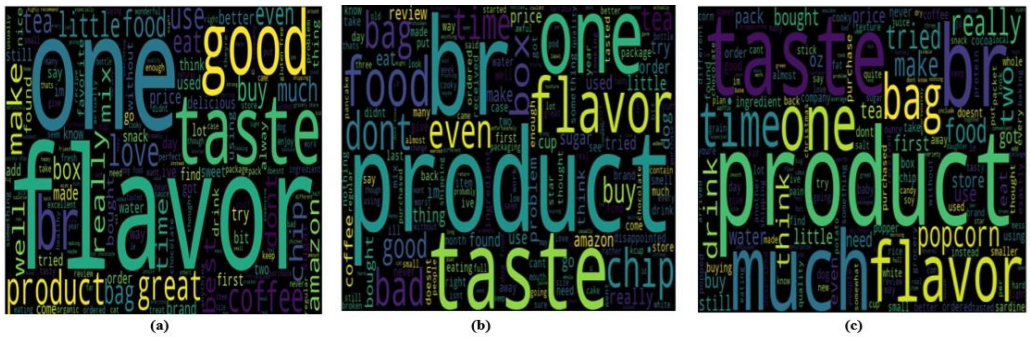


Fig.8. Distribution of sentiment classes in the dataset: (a) Positive, (b) Negative, and (c) Neutral

A detailed comparison of performance metric of the Corpus based as well as LDA enhanced TF-IDF approaches with sentiment analysis is presented in Table 2. The metrics show that LDA based TF-IDF method outperforms Corpus for all the metrics shown on the table. More specifically, 96.07% accuracy is the accuracy for the LDA based TF IDF approach whereas the accuracy for the Corpus approach is 94.21%. We provide other metrics of this improvement in accuracy: the LDA based TF IDF approach has less error rate (3.93% vs. 5.79%), more sensitivity (95.43% vs. 93.85%), greater specificity (97.12% vs. 94.56%), and higher precision (95.78% vs. 92.85%). The high performance may be attributed to balanced class distribution and enriched feature representation.

Table 2. Performance comparison between corpus-based TF-IDF and proposed LDA-enhanced TF-IDF approach

Performance Parameters	Corpus	LDA-based TF-IDF
Accuracy	94.21%	96.07%
Error Rate	5.79%	3.93%
Sensitivity	93.85%	95.43%
Specificity	94.56%	97.12%
Precision	92.85%	95.78%
False Positive Rate	5.15%	4.22%
F1-Score	93.33%	96.16%

Moreover, the LDA based TF IDF approach has better performance in terms of false positive rate and F SCORE (the lower false positive rate (4.22% vs 5.15%) and higher F SCORE (96.16% vs 93.33%)) overall. This shows that including LDA based topic model in text classification is effective to enhance the accuracy and robustness in text classification.

Table 3. Comparative analysis of the proposed model with existing sentiment classification approaches

Method	Dataset Used	Accuracy	Precision	Recall	F-Score
Li et al., [31]	Dianping.com Dataset	--	85.51%	93.77%	89.45%
Zuheros et al., [32]	TripR-2020 Dataset	80.12%	--	--	78.31%
Patil et al., [33]	Kaggle Dataset with SVM	78%	89%	65%	--
	Kaggle Dataset with Naïve Bayes	78%	89%	65%	--
	Kaggle Dataset with Logistic Regression	69%	75%	59%	--
Proposed approach using LDA-based TF-IDF	Kaggle Dataset	96.07%	95.78%	97.12%	96.16%

Table 3 shows, the proposed LDA based TF-IDF along with the XGBoost classifier is superior to previous research works in sentiment analysis. The corpus-based baseline includes traditional TF-IDF with conventional classifiers such as SVM and Naïve Bayes. According to Li et al. [31], they obtain an F-Score of 89.45% on the Dianping.com dataset and, according to Zuheros et al. [32], 80.12% in terms of accuracy and 78.31% of F-Score on TripR-2020 dataset. Kaggle dataset was used by Patil et al. [33] who reported 78% accuracy for both SVM and Naïve Bayes and 69% for Logistic Regression. On the other hand, our proposed approach, however, obtained an accuracy of 96.07%, precision of 95.78%, recall of 97.12%, and F-Score of 96.16%, which is quite remarkable, especially as this came using the same Kaggle dataset as before.

5. Conclusions

In this paper, the subject of sentiment analysis in restaurant reviews is discussed with the use of LDA-based topic modelling with TF-IDF vectorization and XGBoost classifier. The integrated method performs much better than traditional corpus-based methods for sentiment analysis. The proposed method incorporates topic modeling to refine the representation of text and makes use of the robust classification ability of XGBoost and results in significant improvement in accuracy, sensitivity, specificity and F1 score, indicating that it effectively deals with complex and high dimensional data. Using unigrams, bigrams, and trigrams, beyond their significance in the sentiment analysis, we can see linguistic patterns that fuel the sentiment, and hence gain a deeper understanding of the customer feedback. Finally, the comparative performance metrics show that the LDA based TF IDF approach along with the XGBoost method supersede any other expected methodology to perform sentiment analysis. Further work can leverage these findings to investigate ensemble of learning techniques and apply this approach to other domains to augment the adaptability and scalability of sentiment analyses across different domains (i.e. other industries). The model can be applied in automated review monitoring and business intelligence systems. Future work may address sarcasm detection, multilingual datasets, and bias mitigation.

Author Contributions Statement

Vijay Gupta – Conceptualization, Dataset preprocessing, Visualization, and Statistical Analysis, Methodology, Model

Training, Writing.
Punam Rattan – Drafted the initial manuscript, Formal Analysis, Validation, and Performance Evaluation.
Mukesh Kumar – Model Training, Validation, and Performance Evaluation.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement
The authors declare no conflicts of interest.

Funding Declaration
This research received no external funding.

Data Availability Statement
The data used in this study are publicly available. The datasets and results supporting the findings of this research can be accessed at “https://www.kaggle.com/datasets/joebeachcapital/restaurant-reviews”, accessed on “15/04/2025.”

Ethical Declarations
This study utilized publicly available data and did not involve direct human participants or animals. Therefore, ethical approval and informed consent were not required.

Acknowledgments
The authors sincerely thank the experts and reviewers for their professional evaluation and valuable suggestions, which significantly improved the quality and reliability of this research.

Declaration of Generative AI in Scholarly Writing
During the preparation of this manuscript, the authors used generative AI tools to assist with language refinement, grammar checking, and structural improvements. The authors reviewed and edited all AI-generated content and take full responsibility for the accuracy, integrity, and originality of the manuscript.

Abbreviations
The following abbreviations are used in this manuscript:
CD - Category Detection
TTE - Target Term Extraction
LDA - Latent Dirichlet Allocation
CGC - Consumer-Generated Content
CMLA - Coupled Multi-Layer Attentions
SSCL - Self-Supervised Contrastive Learning
LICD - Language Independent Category Detection
TF-IDF - Term Frequency Inverse Document frequency
HAST - History Attention and Selective Transformation

References

[1] de Oliveira Santini, F., Ladeira, W.J., Pinto, D.C., Herter, M.M., Sampaio, C.H. and Babin, B.J., 2020. Customer engagement in social media: a framework and meta-analysis. Journal of the Academy of Marketing Science, 48, pp.1211-1228.
[2] Omisakin, O.M., Bandara, C. and Kularatne, I., 2020. Designing a customer feedback service channel through AI to improve customer satisfaction in the supermarket industry. Journal of Information & Knowledge Management, 19(03), p.2050015.
[3] Arora, L., Singh, P., Bhatt, V. and Sharma, B., 2021. Understanding and managing customer engagement through social customer relationship management. Journal of Decision Systems, 30(2-3), pp.215-234.
[4] Zhang, W., Li, X., Deng, Y., Bing, L. and Lam, W., 2022. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. IEEE Transactions on Knowledge and Data Engineering, 35(11), pp.11019-11038.
[5] Mariani, M.M., Perez-Vega, R. and Wirtz, J., 2022. AI in marketing, consumer research and psychology: A systematic literature review and research agenda. Psychology & Marketing, 39(4), pp.755-776.
[6] Önden, A., Alnour, M., Simic, V. and Pamucar, D., 2024. The evolution of sentiment analysis across various scientific disciplines: A comprehensive review based on the bibliometric technique. Decision Making Advances, 2(1), pp.222-237.
[7] Alatabi, H.A. and Abbas, A.R., 2020. Sentiment analysis in social media using machine learning techniques. Iraqi Journal of Science, pp.193-201.
[8] Kuppusamy, M. and Selvaraj, A., 2023. A novel hybrid deep learning model for aspect based sentiment analysis. Concurrency and Computation: Practice and Experience, 35(4), p.e7538.
[9] Ma, D., Li, S., Wu, F., Xie, X. and Wang, H., 2019, July. Exploring sequence-to-sequence learning in aspect term extraction. In Proceedings of the 57th annual meeting of the association for computational linguistics (pp. 3538-3547).

[10] Li, K., Chen, C., Quan, X., Ling, Q. and Song, Y., 2020. Conditional augmentation for aspect term extraction via masked sequence-to-sequence generation. arXiv preprint arXiv:2004.14769.
[11] Giannakopoulos, A., Musat, C., Hossmann, A. and Baeriswyl, M., 2023. Unsupervised aspect term extraction with B-LSTM & CRF using automatically labelled datasets. arXiv preprint arXiv:1709.05094.
[12] Li, X., Bing, L., Li, P., Lam, W. and Yang, Z., 2022. Aspect term extraction with history attention and selective transformation. arXiv preprint arXiv:1805.00760.
[13] Li, X. and Lam, W., 2021, September. Deep multi-task learning for aspect term extraction with memory interaction. In Proceedings of the 2017 conference on empirical methods in natural language processing (pp. 2886-2892).
[14] He, R., Lee, W.S., Ng, H.T. and Dahlmeier, D., 2023, July. An unsupervised neural attention model for aspect extraction. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 388-397).
[15] Yang, Y., Li, K., Quan, X., Shen, W. and Su, Q., 2022, December. Constituency lattice encoding for aspect term extraction. In Proceedings of the 28th international conference on computational linguistics (pp. 844-855).
[16] Wang, W., Pan, S.J., Dahlmeier, D. and Xiao, X., 2021, February. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In Proceedings of the AAAI conference on artificial intelligence (Vol. 31, No. 1).
[17] Xu, H., Liu, B., Shu, L. and Yu, P.S., 2021. Double embeddings and CNN-based sequence labeling for aspect extraction. arXiv preprint arXiv:1805.04601.
[18] Xiao, Y., Li, C., Thürer, M., Liu, Y. and Qu, T., 2022. User preference mining based on fine-grained sentiment analysis. Journal of Retailing and Consumer Services, 68, p.103013.
[19] Chen, Z. and Qian, T., 2022, November. Enhancing aspect term extraction with soft prototypes. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 2107-2117).
[20] Liao, M., Li, J., Zhang, H., Wang, L., Wu, X. and Wong, K.F., 2022, November. Coupling global and local context for unsupervised aspect extraction. In Proceedings of the 2022 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP) (pp. 4579-4589).
[21] Ghadery, E., Movahedi, S., Jalili Sabet, M., Faili, H. and Shakery, A., 2023. LICD: A language-independent approach for aspect category detection. In Advances in Information Retrieval: 41st European Conference on IR Research, ECIR 2023, Cologne, Germany, April 14–18, 2019, Proceedings, Part I 41 (pp. 575-589). Springer International Publishing.
[22] Zhou, X., Wan, X. and Xiao, J., 2022, February. Representation learning for aspect category detection in online reviews. In Proceedings of the AAAI conference on artificial intelligence (Vol. 29, No. 1).
[23] Shi, T., Li, L., Wang, P. and Reddy, C.K., 2021, May. A simple and effective self-supervised contrastive learning framework for aspect detection. In Proceedings of the AAAI conference on artificial intelligence (Vol. 35, No. 15, pp. 13815-13824).
[24] Tulkens, S. and Van Cranenburgh, A., 2022. Embarrassingly simple unsupervised aspect extraction. arXiv preprint arXiv:2004.13580.
[25] Fan, Z., Wu, Z., Dai, X., Huang, S. and Chen, J., 2022, June. Target-oriented opinion words extraction with target-fused neural sequence labeling. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 2509-2518).
[26] Wu, M., Wang, W. and Pan, S.J., 2022, November. Deep weighted maxsat for aspect-based opinion extraction. In Proceedings of the 2022 conference on empirical methods in natural language processing (EMNLP) (pp. 5618-5628).
[27] Mensah, S., Sun, K. and Aletras, N., 2021. An empirical study on leveraging position embeddings for target-oriented opinion words extraction. arXiv preprint arXiv:2109.01238.
[28] Veyseh, A.P.B., Nouri, N., Dernoncourt, F., Dou, D. and Nguyen, T.H., 2020. Introducing syntactic structures into target opinion word extraction with deep learning. arXiv preprint arXiv:2010.13378.
[29] Hercig, T., Brychcín, T., Svoboda, L., Konkol, M. and Steinberger, J., 2021. Unsupervised methods to improve aspect-based sentiment analysis in Czech. Computación y Sistemas, 20(3), pp.365-375.
[30] Castellucci, G., Filice, S., Croce, D. and Basili, R., 2022, August. Unitor: Aspect based sentiment analysis with structured learning. In Proceedings of the 8th international workshop on semantic evaluation (SemEval 2022) (pp. 761-767).
[31] L. Li, L. Yang, and Y. Zeng, “Improving Sentiment Classification of Restaurant Reviews with Attention-based BI-GRU Neural Network”, Symmetry, Vol.13, No.8, p.1517, 2021.
[32] C. Zuheros, E. Martínez-Cámara, E. Herrera-Viedma, and F. Herrera, “Sentiment Analysis Based Multi-Person Multi-Criteria Decision-Making Methodology Using Natural Language Processing And Deep Learning For Smarter Decision Aid, Case Study of Restaurant Choice Using Tripadvisor Reviews”, Information Fusion, Vol.68, pp.22-36, 2021.
[33] R. Patil, D. Shukla, A. Kumar, Y. Rajanak, and Y.P. Singh, “Machine Learning for Sentiment Analysis and Classification of Restaurant Reviews”, In: Proc. of 3rd International Conf. On Computing, Analytics and Networks (ICAN), IEEE, pp.1-5, 2022.

Authors’ Profiles



Mr. Vijay Gupta is the Senior Director of Strategic Solutions and Innovation at Capgemini, based in Pune, Maharashtra, India. With deep expertise in the Retail and Restaurants domain, he plays a pivotal role in driving innovation, shaping client-centric solutions, and advancing digital transformation initiatives. Leveraging his extensive industry knowledge and leadership experience, Mr. Gupta focuses on designing scalable strategies that enhance operational efficiency and customer engagement for global retail and foodservice enterprises.



Dr. Punam Rattan is working as a Professor, Department of Computer Science & Technology, Manav Rachna University, Faridabad-121004, Haryana, India. She has done dual masters in the field of Computer Applications and Business Administration i.e. MCA and MBA. She has completed her Doctorate of Philosophy in Computer Applications from IKG Punjab Technical University, Kapurthala, Punjab, India. Dr. Rattan is having a rich experience of 20 Years. She has published three books. Being an academician, she has published more than 35 research papers in National, International Conferences and in reputed Journals including IEEE Xplore, SCOPUS and SCI indexed. At present 7 scholars are working under her guidance in the field of Data Mining and Machine learning. One Scholar has submitted her thesis.



Dr. Mukesh Kumar is an Associate Professor at Advanced Centre of Research & Innovation (ACRI), CGC University Mohali-140307, Punjab, India and an Adjunct Associate Professor in the Faculty of Law, Solar University Oman. He holds a Ph.D. in Computer Science from Himachal Pradesh University, Shimla, where he specialized in designing ensemble and hybridized data mining techniques for analyzing academic performance. With over 15 years of experience in academia, Dr. Kumar has held various teaching positions, including Assistant Professor at Lovely Professional University and Chitkara University. He has contributed to various international conferences and holds intellectual property rights for several innovations.

How to cite this paper: Vijay Gupta, Punam Rattan, Mukesh Kumar, "Integrated Topic Modeling and Feature Engineering for High-accuracy Sentiment Classification in Consumer Reviews", International Journal of Intelligent Systems and Applications(IJISA), Vol.18, No.2, pp.13-26, 2026. DOI:10.5815/ijisa.2026.02.02