IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 March 2024, accepted 9 April 2024, date of publication 17 April 2024, date of current version 24 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3390605

RESEARCH ARTICLE

Product Helpfulness Detection With Novel Transformer Based BERT Embedding and Class Probability Features

AREEBA ISHTIAQ[1], KASHIF MUNIR[1], ALI RAZA[2], NAGWAN ABDEL SAMEE[3], MONA M. JAMJOOM[4], AND ZAHID ULLAH[5]

[1]Institute of Information Technology, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan 64200, Pakistan
[2]Department of Software Engineering, The University of Lahore, Lahore 54000, Pakistan
[3]Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[4]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[5]Department of Information System, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Corresponding authors: Kashif Munir (kashif.munir@kfueit.edu.pk) and Nagwan Abdel Samee (nmabdelsamee@pnu.edu.sa)

This work was supported by Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, through Princess Nourah bint Abdulrahman University Researchers Supporting Project under Grant PNURSP2024R104.

ABSTRACT Nowadays global market products are readily accessible worldwide, and a vast array of reviews across numerous platforms are posted daily in several categories, making it challenging for customers to stay informed about their product interests. To make informed decisions regarding product quality, users require access to reviews and ratings. Owners and managers must analyze customer ratings and the underlying emotional content of reviews to enhance the product's quality, cost, customer service, and environmental impact. The primary aim of our proposed research is to accurately predict product helpfulness through customer reviews using the Large Language Model (LLM), thereby assisting customers in saving time and money. We employed a benchmark dataset, the Amazon Fine Food Reviews, to develop numerous advanced machine-learning techniques. We introduced a novel transformer approach BERF (BERT Random Forest) for feature engineering to enhance the value of user evaluations for Amazon's gourmet food products. The BERF method utilizes BERT embeddings and class probability features derived from product helpfulness online reviews textual data. We have balanced the dataset using the Synthetic Minority Over-sampling TEchnique (SMOTE) approach. Our comprehensive study results demonstrated that the Light Gradient Boosting Machine (LGBM) strategy outperformed existing state-of-the-art approaches, achieving an accuracy of 98%. The performance of each method is confirmed using a k-fold method and further improved through hyperparameter optimization. Our innovative study employing a transformer model has significantly enhanced the utility of customer reviews, substantially reducing online product scams and preventing wasted time and money.

INDEX TERMS Product helpfulness, large language model (LLM), machine learning, deep learning, text mining, BERT, transformer.

I. INTRODUCTION

The last decade has shown a significant increase in the availability of product reviews on traditional retail sites [1], in both professional and individual formats. To reduce the uncertainty associated with purchasing products, users consult these reviews and pay attention to online information such as images. According to a research study, most users prioritize customer reviews before making a purchase [2]. In other words, nearly 90% of consumers check reviews

The associate editor coordinating the review of this manuscript and approving it for publication was Jad Nasreddine.

before buying a product. Reviews are becoming increasingly important for both consumers and businesses. For consumers, online reviews are crucial in deciding whether or not to purchase a product [3]. Exclusive online retailers are focusing on improving the management of reviews, understanding that positive reviews can significantly impact profit gains, and creating an environment where customer reviews are an integral part of the business.

Amazon is a multinational technology corporation specializing in e-commerce [4], where reviews are crucial for purchasing decisions, providing vital and genuine insights. However, reading through all the reviews of a chosen article can be time-consuming. Reviews are not only important for buyers but also for vendors, who rely on them to differentiate and promote their products [5]. In the context of business growth and customer care, online ratings play a significant role. Customers can easily assess a product's appeal by reviewing its ratings. The decision to purchase a product often depends on its ratings and reviews, which can create a positive or negative impression. Most research in this area has utilized online feedback from Amazon to predict review helpfulness [6], with each review accompanied by data indicating the number of people who found it helpful. In the field of e-commerce, product ratings have become increasingly important.

The significance of review ratios has increased during the COVID-19 pandemic, which began in 2020, as commercial transactions have shifted towards being conducted electronically over the internet [7]. This shift has led to a 43% increase in the ratio of online reviews, and it continues to rise over time. Reviews offer a time-saving way to purchase products, as ratings provide helpful and valuable recommendations [8]. On the other hand, sellers are also interested in review analysis to understand customer interests better and achieve successful product sales. Researchers have increasingly focused on predicting the helpfulness of reviews by employing a range of Machine Learning (ML) methods [9]. Machine learning aims to identify significant patterns and gain knowledge from data. Information overload on internet review sites has significantly hindered buyers' ability to assess product or business quality when making purchases. The growth of social media has made it harder to differentiate between genuine content and advertising, leading to a surge in misleading evaluations in the market. The usefulness of a review depends on a voting mechanism [10].

Several neural network methods have been developed to automatically assess the usefulness of consumer product reviews [11]. Many current models rely on basic explanatory factors, particularly those derived from substandard evaluations that may be deceptive and result in uncertainty. Effective feature selection is crucial for forecasting the usefulness of online consumer reviews. The transformer-based BERT, a newly evolved language representation approach [12], can achieve state-of-the-art outcomes on many natural language processing works. Our main insight is to provide a platform where the purchase of products and customer decisions can be related to the responses of experienced buyers. As a conclusion, a methodical technique is required to handle big data.

Our primary research contributions to the helpfulness of customer reviews are as follows:

- We proposed a novel transformer-based BERF method, which generates BERT embeddings and class probability features from product helpfulness online reviews. The newly generated salient feature set is then utilized to build advanced machine-learning models.
- We employed a fine-tuned BERT model and four sophisticated machine-learning methods for detecting the product's helpfulness. K-fold validation is utilized to validate the performance results of the models, and hyperparameter adjustment is employed to improve performance efficacy. In addition, we have balanced the dataset using the SMOTE approach.

Subsequent sections of the research are structured as follows: Section II comprises the literature work analysis. Section III outlines our proposed research approach. Section IV assesses the outcomes of the approaches used in the comparison. The primary discoveries are detailed in Section V.

II. LITERATURE REVIEW

Scientists have described the distinctive characteristics of consumer product reviews and forecasted their usefulness. Large companies gather data from online sources to facilitate users with product recommendation systems. Major corporations, such as Yelp, Spotify, and Amazon, rely on and require the assessment of product review helpfulness to achieve better results and revenue from customers. The increasing usage of social media has made it difficult to differentiate between authentic, useful reviews and advertisements.

The literature study focusing on state-of-the-art applied approaches to performance is detailed in Table 1.

This research [10] utilizes the Amazon product review dataset, spanning from May 1996 to October 2018. The dataset comprises approximately 233.1 million entries across 29 distinct product categories (such as Office Products, Pet Supplies, Grocery Gourmet Food, etc.) and includes 11 columns. The research findings indicate a high accuracy in predicting the usefulness of Amazon product reviews. The improvement from the initial model, which produced an inadequate confusion matrix, to the final model, which applied various data manipulation techniques to achieve F1 scores of 0.83, serves as validation of the research methodology. The effectiveness of this pragmatic approach was demonstrated by the increasing F1 scores, and the research further identified factors that could improve the helpfulness of reviews. These factors include eliminating duplicate reviews, estimating review helpfulness based on word count, and utilizing part-of-speech (POS) tagging to incorporate lexical components into all reviews. The conclusion and result of this study underscore that review helpfulness can be predicted excellently following the deployment of the trained model. By removing duplicate

TABLE 1. The summary analysis of analyzed literature.

Ref.	Year	Dataset Used	Proposed Technique	Performance Score	Research Limitations
[3]	2022	Amazon fine food reviews	BERT Model	79%	Low performance scores were achieved.
[13]	2023	Amazon fine food reviews	Naïve Bayes	85%	Low performance scores were achieved.
[10]	2023	Amazon fine food reviews	BERT Model	86%	Low performance scores were achieved.
[14]	2022	Amazon Alexa reviews	SVM Model	91.5%	Classical Machine Learning method is build.
[15]	2023	Amazon fine food reviews	RoBERTa Model	82%	Low performance scores were achieved.
[16]	2020	Amazon fine food reviews	RoBERTa Model	82%	Low performance scores were achieved.
[17]	2022	Yelp Shopping reviews	K-NN	59.6%	Low performance scores were achieved.

reviews, the model provides a more straightforward way to predict the helpfulness of a product.

In [18], researchers presented sentiment analysis of Amazon product reviews using textual analysis and natural language processing (NLP) methods. This research utilized web-based tools for analyzing customer reviews, which were categorized in a specific manner. Similarly, the study in [5] outlined sentiment analysis of the Amazon product Alexa using neural network algorithms. A fundamental and significant element of NLP is emotion analysis. In this study, researchers employed Naïve Bayes, Random Forest, and Support Vector Machine (SVM) algorithms to facilitate sentiment analysis of Amazon products.

This study [3] has presented an analysis of Amazon fine food reviews using the BERT model. Online product reviews play an important role in predicting the helpfulness of product reviews. In this digital era, where people prefer to shop online rather than visit physical stores, many rely on online product reviews for a better purchasing experience. The helpfulness of product reviews enables customers to save time and purchase products in a cost-efficient manner. For this study, the Amazon Fine Food Products dataset, which is available for download from Kaggle, was used for experimentation. Initially, the data was cleansed, which involved removing special characters and dropping punctuation, leading to the completion of the model analysis. Natural Language Processing (NLP) and BERT techniques were employed to train the model on the dataset, which was subsequently deployed.

In this research [15], the sentiment analysis of Amazon product reviews using deep learning techniques is discussed. The ratio of online shopping increased significantly with the rise of the COVID-19 pandemic in 2020. Due to this pandemic situation, the importance of online product reviews has significantly grown in the digital world. More people now shop online, relying on product review analysis, which assists customers in making time-saving and cost-effective purchases. In this online context, not only do customers benefit from product reviews, but sellers also leverage these reviews for the revenue and growth of their businesses.

In this study [16], specific models have been used for sentiment analysis, which includes Transformers such as BERT, ROBERTa, and XLNet. After the experiment, ROBERTa achieved the highest accuracy at 82% among all the models used. Choosing the right features is essential for precisely forecasting the usefulness of online consumer feedback. The newly introduced Bidirectional Encoder Representations from Transformers (BERT) model represents a significant advancement in language processing, achieving unparalleled results across various natural language processing tasks. This work proposes a prediction model that utilizes BERT features and deep learning approaches to determine the helpfulness scores of customer reviews. The program employs a BERT-based algorithm to analyze the dataset of Amazon product reviews, aiming to assist users in making informed purchasing decisions.

This study [19] tests a dataset comprising reviews of Shopify applications. To address the aforementioned constraints, user evaluations are classified into two categories: positive and negative. These evaluations are then subjected to preprocessing to cleanse the data. Following this, different methods of feature engineering such as bag-of-words, term frequency-inverse document frequency (TF-IDF), and chi-square (Chi2) are utilized, both separately and in conjunction, to preserve essential information. Ultimately, the reviews are classified as either 'pleased' or 'dissatisfied' using AdaBoost classifier, Random Forest, and logistic regression models.

The objective of this work [20] was to develop a comprehensive method by refining the BERT foundational model through fine-tuning. The efficacy of BERT-based classifiers was evaluated by comparing their performance with traditional bag-of-words techniques. The tests, conducted using Yelp shopping product reviews, indicated that fine-tuned BERT-based classifiers outperformed bag-of-words methods in accurately categorizing reviews as useful or unhelpful. Furthermore, it was discovered that the sequence length used in the BERT-based method significantly affects classification effectiveness. The BERT method with a sequence data length of 64 achieved the lowest accuracy, at 0.668, and an F1 score of 0.685. Conversely, a more sophisticated model with a sequence length of 320 achieved the highest accuracy of 0.707 and an F1 score of 0.717. Sequence lengths of 128 and 256 yielded superior outcomes compared to those of 384 and 512. The study demonstrates that the sequence length utilized in refining and evaluating the BERT base approach has a important impact on classification accuracy.

This paper [21] presents a novel ensemble technique known as the regression vector voting method for identifying poisonous remarks on various social media networks. The ensemble merges a support vector classifier and logistic

regression using a soft voting criterion. The suggested technique is evaluated through experiments on both unbalanced and balanced datasets to analyze its performance. To address the issue of an unbalanced dataset, the SMOTE is employed to balance the data. Additionally, two feature extraction methods are employed to assess their appropriateness: TF-IDF and Bag-of-Words (BoW).

This research [22] explores the application of diverse machine learning approaches to assess the polarity of feelings expressed in user review data on the IMDb website. To achieve this objective, the reviews undergo an initial preprocessing phase to eliminate redundant information and noise. Subsequently, a range of classification methods, including support vector machines (SVM), Naïve Bayes classifiers, random forests, and gradient boosting method, are employed to predict the sentiment expressed in these reviews.

This research [23] introduced a sophisticated deep-learning model designed to accurately categorize the most favorable and unfavorable product evaluations. This model employs a neural network approach, utilizing a recurrent neural network (RNN) method design that is the first to surpass conventional text categorization algorithms in this specific problem domain. The authors evaluate the deep learning model by comparing it with a baseline classifier that employs logistic regression. Similarly, research [13] examines and evaluates methods for automatically identifying the sentiments conveyed in English texts for Amazon and Flipkart products, using Random Forest and K-Nearest Neighbor algorithms. The text offers an in-depth comparative review of current sentiment analysis algorithms and approaches, evaluating them based on five major factors. This leads to an assessment of their performance based on parameter usage and contributions.

This study [24] utilizes a web-based application to categorize and analyze customer evaluations of items, thereby saving analysts considerable time and effort that would otherwise be required to manually sift through millions of reviews. The sentiment analysis on product reviews employs NLP techniques. The developed technology comprises the following five components: Lexicon-Based Sentiment, Text Analytics, Customer Satisfaction Score, Amazon Products, and the Application Programming approach for the extraction of recent reviews. The text analytics component processes the textual data by removing any unnecessary elements and extracting the sentiment. The customer satisfaction score can be determined by calculating the average of the sentiment scores. Python-based sentiment analysis may be used to research and analyze reviews that analysts wish to evaluate.

In this study [13], Linear Regression and Convolutional Neural Network (CNN) are identified as two of the most commonly employed machine-learning algorithms. The research demonstrates that the method of representing text data plays a crucial role in performance outcomes, indicating that TF-IDF and Word2Vec methods result in the most favorable Mean Squared Error (MSE) scores. Experimental findings reveal that the bag-of-words (BoW) approach, when representing review text, yields the poorest results across both datasets, with the exception of a sampled example using the Cell Phones and Accessories dataset.

A. LIMITATIONS AND RESEARCH GAP

The limitations of the previous study are evident in the techniques and frameworks utilized. Traditional methods for manipulating text features, such as TF-IDF or BoW, were employed in earlier research. In contrast, our work leverages the advanced contextualized embeddings of the BERT (Bidirectional Encoder Representations from Transformers) model, which significantly enhances our understanding of language semantics and context. Additionally, prior studies often relied on conventional machine learning models, which lack the complexity and adaptability of more sophisticated models. By integrating advanced methodologies and state-of-the-art models, we have successfully addressed the performance limitations identified in previous studies, achieving exceptional levels of accuracy and efficacy in our experimental results.

We have discovered research gaps through a thorough literature review:

- Traditional machine learning techniques were previously used with BERT, BOW, and other feature engineering approaches to determine the helpfulness of user evaluations on Amazon items. Although these approaches showed high-performance ratings, there is still an accuracy gap that has to be resolved.
- The Amazon Fine Food dataset has complex attributes that require a sophisticated feature engineering strategy to enhance the efficacy of product review helpfulness.

III. PROPOSED METHODOLOGY

This module examines our newly proposed research methodology, as exemplified in Figure 1. Our proposed method utilizes the Amazon Fine Food Reviews dataset from Kaggle for research projects. The original textual dataset's features are then preprocessed to eliminate noise and encode the data effectively. We propose a unique feature engineering method, BERF (BERT-RF), to enhance the utility of customer evaluations of Amazon's fine food products. The newly generated dataset is divided into two parts for train and test, using an 80% train and 20% test split ratio, respectively. Sophisticated machine learning techniques are applied, and their effectiveness is evaluated using fresh data. The efficacy of each approach, along with hyperparameter optimization, is further validated using cross k-fold validation. The superior machine learning technique is then utilized to forecast the helpfulness of customer reviews for online products.

The objective of the proposed technique is to forecast the usefulness of Amazon product reviews by utilizing transformer-based feature embeddings. The focus is on the Amazon Fine Food Reviews dataset, which encompasses up to 568,454 reviews. The primary columns selected for analysis are the "Helpfulness Denominator" and the

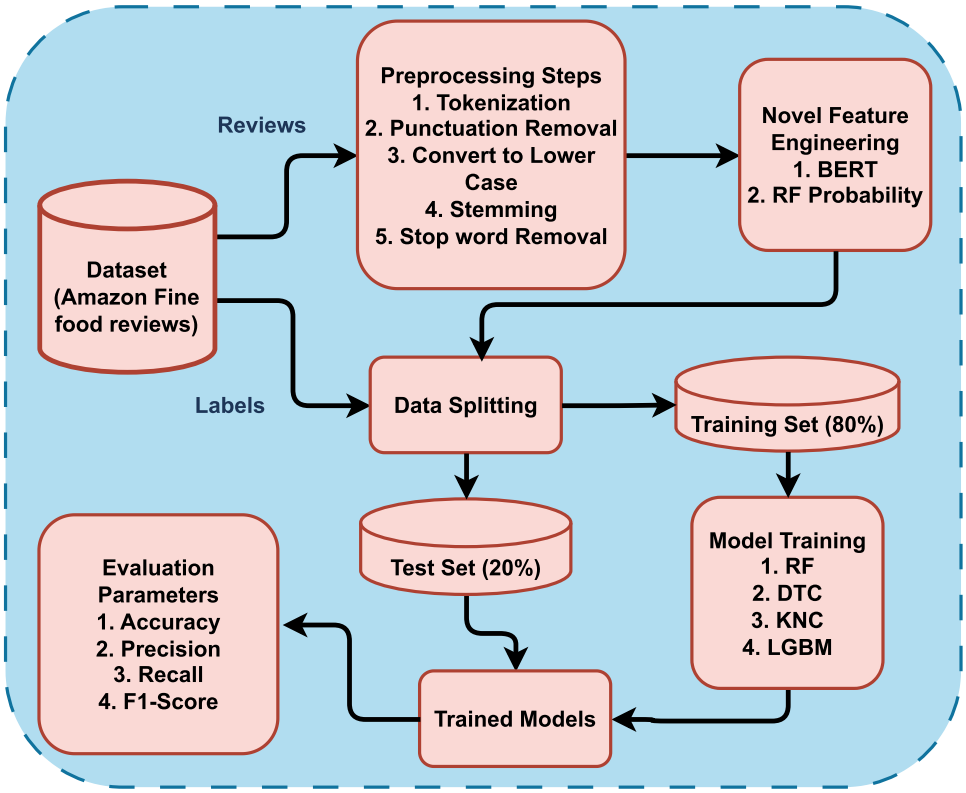

FIGURE 1. The transformer-based novel methodology workflow analysis.

"Text" columns, as they contain crucial information for understanding the helpfulness of product evaluations.

- **Step 1:** The first phase involves data preparation, where methods are employed to cleanse the data by eliminating stop words and noise. This ensures that the subsequent analysis is founded on significant and relevant data. Following this, the sanitized data is used to generate a systematically organized dataset for further research.
- **Step 2:** For feature engineering, a novel transformer-based model named BERF (BERT-RF) is utilized to capture the complex patterns and semantic linkages in the reviews.
- **Step 3:** After performing feature extraction with BERF, the dataset is divided into two portions: a train set and a test set. The train set is used to train various machine learning models, while the testing set is employed to evaluate the models' performance and generalization capabilities.
- **Step 4:** Several machine learning models, such as Random Forest, K-Nearest Neighbors, Decision Tree, and LightGBM, are utilized to forecast the helpfulness of reviews. These models are selected based on their versatility and ability to handle different types of data and interactions.
- **Step 5:** The evaluation of each model is conducted using appropriate metrics such as accuracy, precision, recall, and F1 score.

A. PHASE 1: REVIEW HELPFULNESS TEXTUAL DATA

In this study, we employed a benchmark dataset known as the Amazon Fine Food Reviews [25]. This dataset is sourced from the Kaggle website and was originally compiled and released by McAuley and Leskovec [26] in their study on online reviews. It consists of reviews for gourmet food products posted on Amazon.com, encompassing a total of 568,454 reviews for 74,258 products. These were employed in conducting our research experiments.

B. PHASE 2: TEXT PREPROCESSING AND DATA ANALYSIS

A sophisticated text preparation strategy has been developed to enhance the quality of user comments in our study. The first phase involves eliminating common stopwords, which are frequently occurring words that do not impact the overall content of the text. This step facilitates the removal of irrelevant information, simplifying the analysis process to focus on more significant elements. Additionally, punctuation marks, special characters, and numerical digits are methodically removed to improve the precision of the text. The technique includes tokenization, stemming, and lemmatization to normalize and reduce words to their fundamental or core forms. This approach not only helps in reducing noise but also ensures that different forms of words are treated consistently, thereby preserving the fundamental semantic meaning of the comments. Through the implementation of this comprehensive text

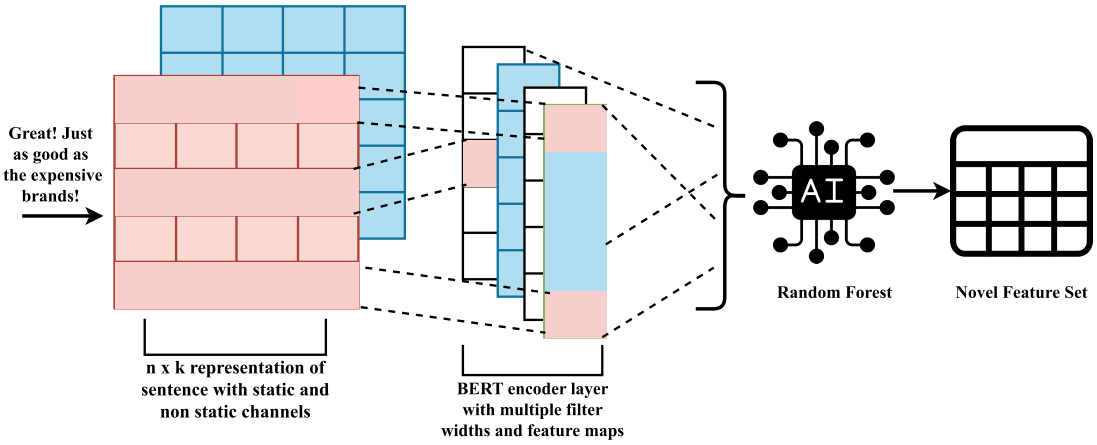

FIGURE 2. The proposed novel LLM based BERF features engineering architecture.

preparation pipeline, user comments are efficiently sanitized, enabling a more precise and insightful analysis of our research.

Figure 3 graphically represents the most common terms in the dataset, providing a concise overview of important patterns. This visualization aids in data comprehension, enabling researchers to quickly identify key themes and trends. The benefits include effective data exploration, efficient analysis, and enhanced decision-making, which collectively contribute to more insightful conclusions and robust research outcomes. We generated the Word Cloud before training our BERT model. The keywords for reviews depicted in Fig. 3 are "good," "one," and "taste," indicating that the majority of reviewers primarily focus on the product's flavor.

FIGURE 3. The word cloud data analysis.

C. PHASE 3: NOVEL LLM BASED BERT-RF TEXTUAL FEATURE EXTRACTION

In this module, we analyze our novel proposed features engineering method. The feature extraction mechanisms and architecture of the proposed BERF approach are illustrated in Figure 2. We propose a novel transformer-based BERF method, which generates BERT embeddings and class probability features from product helpfulness online reviews. The newly generated salient feature set is then used to build advanced neural network models.

1) TRANSFORMER-BASED BERT FEATURES

Transformer-based feature engineering involves the use of transformer models, specifically pre-trained ones such as BERT, to extract significant features values from raw data for various machine learning tasks. BERT represents an unsupervised pre-training technique within Natural Language Processing (NLP). Unlike conventional NLP methods, BERT benefits from being pre-trained on a large corpus of text, requiring only fine-tuning for specific tasks. This approach allows BERT to understand the context of words in sentences more effectively than previous models, making it a powerful tool for feature engineering in NLP applications.

2) RF BASED CLASS PROBABILITY FEATURES

Random Forest (RF) probability-based feature engineering is a method that leverages the inherent characteristics of Random Forest models to generate additional features based on class probabilities [27]. This approach involves extracting class probabilities as extra features and integrating them into the dataset. These probabilities represent the likelihood of each data point belonging to a specific class, offering more nuanced information than basic class labels alone. By incorporating these probability-based features into the existing feature set, machine learning models can gain insights into the distribution of class probabilities within the data.

Algorithm 1 outlines the sequential working for extracting new features.

D. PHASED 4: FEATURES DATA BALANCING

Figure 4 illustrates the study of the dataset's class distributions using histograms. The Amazon dataset has been classified into two main groups. The first category includes cases where the reviews are considered useful and labeled as "yes." This rating implies that the reviews provide

Algorithm 1 BERF LLM Algorithm

Input: Input Textual Content.
Output: Novel Feature Set.

initiate;
1- $BERT_{ce} \longleftarrow E_{BERT}(D_t)$ // here $BERT_{ce}$ are the Embedding features and D_t are input Textual Content.
2- $RF_{pf} \longleftarrow P_{RF}(BERT_{ce})$ // here RF_{pf} are the class Probability features and $BERT_{ce}$ are input Embedding features set.
3- $F_t \longleftarrow RF_{pf}$ // here F_t are the Novel feature set.
end;

important insights or assistance to future buyers. The second category is labelled as "no," indicating that the reviews in this group are not deemed useful. Following this, we balanced the dataset's features using the Synthetic Minority Over-sampling TEchnique (SMOTE) approach.

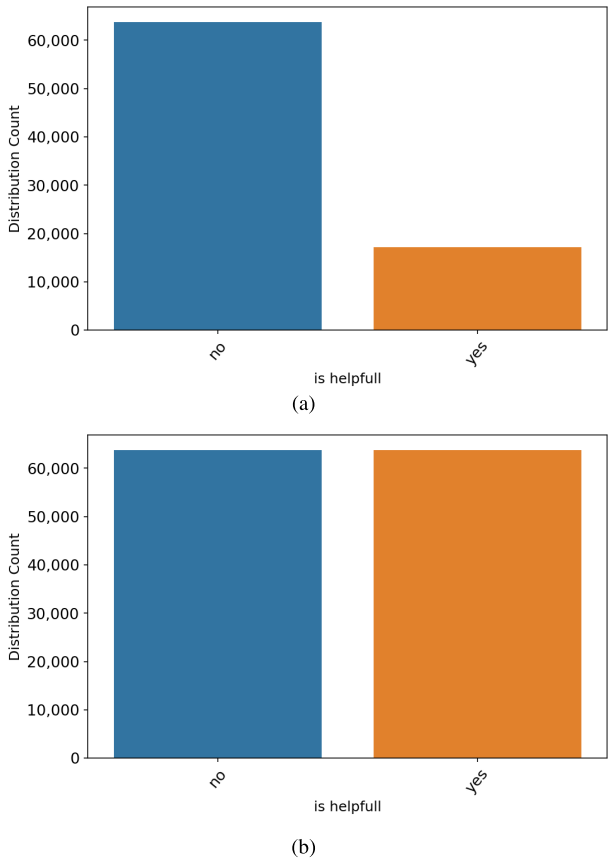

FIGURE 4. Histogram chart-based data balancing of features.

E. PHASE 5: FEATURES DATA SPLITTING

The partitioning of data into two distinct sets, known as training data and test data, plays a crucial role in the construction and assessment of machine learning models. In this research, the dataset is partitioned with a split ratio of 80% for train and 20% for test. This approach ensures that the model is not only exposed to a substantial portion of the data for learning but also reserves a separate subset for the unbiased evaluation of its performance. Sophisticated features were incorporated to enhance the learning process. Machine learning techniques were then applied, and their effectiveness was assessed using the previously unseen test data.

F. PHASE 6: APPLIED ARTIFICIAL INTELLIGENCE MODELS

This section discusses the applied machine learning algorithms, detailing their implementation and associated hyperparameters [28], [29], [30], [31]. The Scikit-learn library is utilized to implement these algorithms. Specifically, four supervised machine learning algorithms are implemented in Python, leveraging the scikit-learn module. Supervised machine learning methods are commonly employed to tackle both classification and regression tasks.

- **Bidirectional Encoder Representations from Transformers (BERT):** model for detecting product helpfulness in online reviews [32]. By leveraging the deep learning capabilities of BERT, which understands the nuances of language context, our research aims to classify reviews based on their perceived helpfulness to consumers. Utilizing a dataset of customer reviews, the BERT model is fine-tuned to identify key features that contribute to the helpfulness rating of a review. Preliminary results indicate that the BERT-based approach significantly outperforms traditional machine learning models in detecting helpful and non-helpful product reviews.
- **Random Forest Classifier (RF):** algorithm is an effective machine learning classifier [33] that falls within the ensemble learning category. It is a tree-based ensemble model that uses decision trees as weak learners to generate very accurate predictions. RF employs bootstrap aggregation, or bagging, to train various decision trees on diverse bootstrap samples, thereby enhancing prediction accuracy by aggregating the outcomes of these weak learners.
- **Decision Tree (DT):** is a widely utilized method for solving classification and prediction tasks [34]. DT serves as an effective method for understanding data characteristics and making informed decisions based on inference. Decision trees are created by iteratively splitting the data based on specific criteria, and they comprise three types of nodes: root, internal, and leaf—the primary node being the root. The structure of a decision tree can be either binary or non-binary, which is determined by the number of child nodes each parent node supports. The gain ratio is a criterion frequently utilized for splitting nodes in decision trees.
- **K-Neighbors Classifier (KNC):** is a straightforward and adaptable machine learning technique [35] that falls within the supervised learning category. The KNN

algorithm is widely used in classification due to its easy implementation and straightforward functionality. It operates effectively as long as there are none missing values (NAs) in the dataset, which must be either removed or transformed according to different principles. While most processes rely on numerical computations to produce predictions, some may also employ text to enhance model understanding. The fundamental premise of KNN is based on the notion that data points with similar characteristics are likely to yield similar outcomes.

- **Light Gradient Boosting Machine (LGBM):** is an innovative [36], scalable, precise, and efficient Gradient Boosting Decision Tree framework introduced by Microsoft. Gradient boosting classifiers, a set of machine learning techniques, aggregate multiple weak learners to create a robust predictive model. These methods are particularly utilized for machine learning tasks, including ranking and categorization. Unlike previous boosting techniques that employ depth-wise or level-wise splitting, LGBM optimizes by splitting the tree leaf-wise using the best-fit strategy. This leaf-wise splitting approach enhances efficiency and minimizes wastage by reducing loss, thereby improving accuracy.

G. PHASE 7: HYPER-PARAMETER SETTING

Table 2 outlines the optimal hyperparameters for the techniques used in this research. Optimizing hyperparameters is an essential stage in the machine learning model training process. Hyperparameter optimization seeks to identify the best hyperparameter configuration for deployed models, thereby enhancing their performance and reducing errors. We have adjusted the parameters of applied methods using recursive testing and training mechanisms, as well as the k-fold cross-validation approach.

TABLE 2. The optimal hyperparameters analysis.

Technique	Hyperparameter Description
RF	n_estimators=20, max_depth=20, random_state=0
DT	criterion="gini", splitter="best",min_samples_split=2
KNC	n_neighbors=2 , weights='uniform', leaf_size=30, p=2
LGBM	num_leaves=31, boosting_type='gbdt', learning_rate=0.1, n_estimators=100, min_child_weight=0.001, min_child_samples=20

IV. EXPERIMENTS AND OBSERVATIONS

This section presents the findings from studies conducted to address the issue of predicting helpfulness. The study employed various methods, with a particular focus on feature selection. We utilized the novel BERF approach, which resulted in enhanced outcomes in our trials. Several experiments were conducted to assess learning models using feature extraction techniques on the Amazon Fine Food Reviews dataset.

A. EXPERIMENTAL SETUP

For our study project, the experimental environment is constructed using a cloud-based Notebook, specifically Google Colab. We assessed the performance of neural network techniques by evaluating them based on accuracy, F1 score, precision, and recall metrics. The details of the environment utilized are outlined in Table 3.

TABLE 3. The experimental environment analysis.

Specification	Value
Programming language	Python 3.0
Environmental model name	Intel(R) core(TM)i5-3307U
CPU MHz	CPU@1.80GHz
RAM	4.00 GB
Cache size	64 KB
CPU cores	2

Our proposed study encompasses several assessment criteria, including accuracy, F1 score, recall, and precision. This evaluation measures parameters are utilized to assess the performance of machine learning models. Classification models can be evaluated using a confusion matrix to determine the accuracy of their predictions on the testing set. The confusion matrix, shown in the findings section, can be viewed as an error matrix that represents four values: true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

- **Accuracy:** is the proportion of correct predictions made by classifiers using test data. The highest achievable accuracy score is 1, signifying that all forecasts are correct.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

- **Precision:** Precision refers to the accuracy of our classifiers. Precision is calculated as the ratio of true positives to the sum of TPs and FPs.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

- **Recall:** Sensitivity, also known as recall, is the proportion of correctly identified positive events out of all actual positive instances.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

- **F1-Score:** is a crucial metric for assessing classifier performance and is considered more significant than accuracy and recall.

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

- **Confusion Matrix:** is a table commonly employed to depict the classifier's performance on test data. It is referred to as an error matrix that enables the viewing of an algorithm's performance.

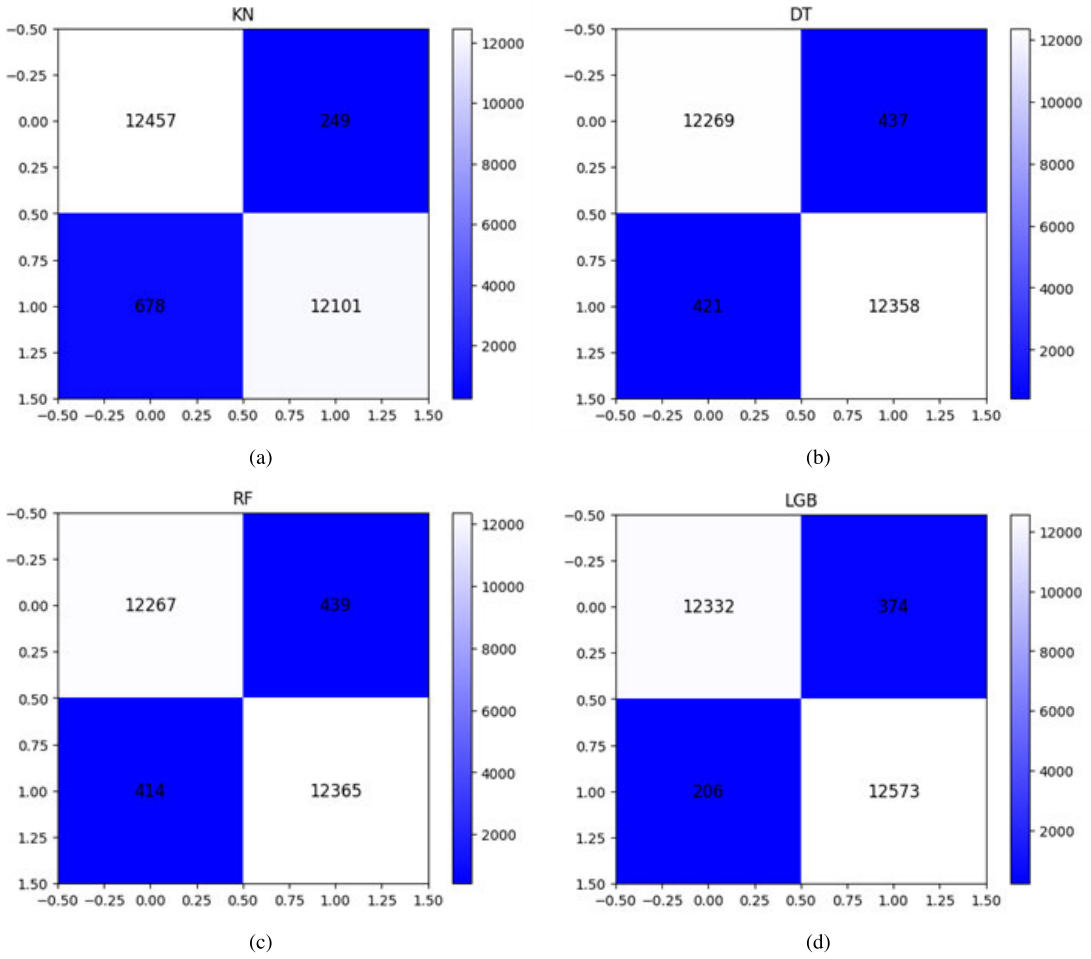

FIGURE 5. Confusion matrix analysis based performance evaluation.

B. RESULTS WITH CLASSICAL BERT EMBEDDING FEATURES

The performance outcomes of machine learning models employing classical BERT embedding features are delineated in Table 4. The highest accuracy, reaching 86.79%, is observed in the Random Forest (RF) models when evaluated on the test dataset. In comparison, other models demonstrated average performance. Specifically, Decision Trees (DT) recorded a significantly lower accuracy of 73.40%, indicating a disparity in performance. Notably, the RF models excel in terms of F1 score, accuracy, and recall when leveraging BERT features for analysis. The Random Forest algorithm, in particular, achieves the highest recall score, underscoring its efficacy in feature extraction using BERT. However, it is important to acknowledge that the performance in detecting helpful product reviews remains suboptimal. This highlights an urgent need for more advanced mechanisms to enhance performance further.

TABLE 4. Neural network models results with BERT features.

Classifier	Accuracy (%)	Target	Precision	Recall	F1
RF	86.79	No	0.83	0.92	0.87
		Yes	0.91	0.81	0.86
		Average	0.87	0.87	0.87
DT	73.40	No	0.74	0.72	0.73
		Yes	0.73	0.75	0.74
		Average	0.73	0.73	0.73
KNC	83.77	No	0.94	0.72	0.82
		Yes	0.78	0.95	0.85
		Average	0.86	0.84	0.84
LGBM	79	No	0.75	0.87	0.80
		Yes	0.85	0.71	0.77
		Average	0.80	0.79	0.79

C. RESULTS WITH NOVEL PROPOSED BERF FEATURES

After examining results using only BERT features, we analyzed the performance with the novel proposed BERF features, as described in Table 5. The learning models achieved an impressive 98% accuracy score, with LGBM standing out in the test dataset. The analysis reveals that LGBM performed exceptionally well in terms of F1 score, accuracy, and recall when utilizing both feature extraction approaches (BERT and RF) together. Among the models, Random Forest (RF) achieved the highest recall score using the BERF feature extraction method, followed by Decision Trees (DT) and K-Nearest Neighbors (KNC). This

analysis demonstrates that the proposed feature engineering techniques significantly enhance performance in detecting the helpfulness of product reviews.

TABLE 5. Machine learning models results with proposed BERT+RF features.

Method	Accuracy (%)	Target	Precision	Recall	F1
RF	96.65	No	0.97	0.97	0.97
		Yes	0.97	0.97	0.97
		Average	0.97	0.97	0.97
DT	96.63	No	0.97	0.97	0.97
		Yes	0.97	0.97	0.97
		Average	0.97	0.97	0.97
KNC	96.36	No	0.95	0.98	0.96
		Yes	0.98	0.95	0.96
		Average	0.96	0.96	0.96
LGBM	**98**	**No**	**0.98**	**0.97**	**0.98**
		Yes	**0.97**	**0.98**	**0.98**
		Average	**0.98**	**0.98**	**0.98**

D. CONFUSION MATRIX RESULTS AND HISTOGRAM ANALYSIS

The confusion matrix validation analysis of the employed machine learning method, using the proposed features engineering, is illustrated in Figure 5. A lower wrong prediction error is achieved using the proposed features engineering approach for product review helpfulness detection. The proposed Light Gradient Boosting Machine (LGBM) model made only 580 wrong predictions out of 25,485. This validates the performance score of the applied method using the novel proposed BERF features.

The histogram-based performance analysis of applied machine learning models, utilizing both feature engineering approaches as illustrated in Figure 6, demonstrates a clear distinction in representing the results. The use of only BERT features achieved lower results; however, performance scores were significantly improved with the novel proposed feature, BERF. This analysis further reveals that the superior performance results of applied methods using BERF features for product review helpfulness detection are noteworthy.

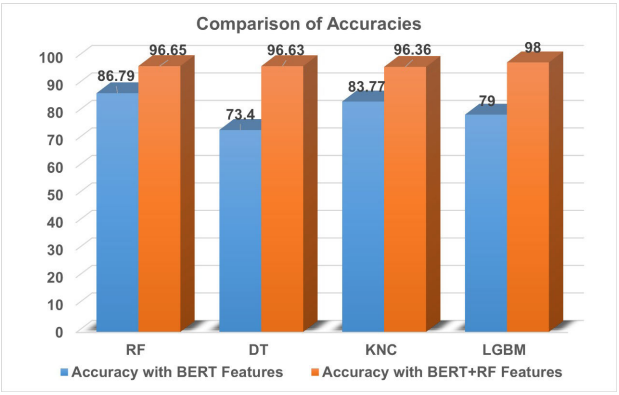

FIGURE 6. The histogram-based results comparisons of applied methods.

E. KFOLD CROSS-VALIDATION RESULTS

We utilized k-fold cross-validation methods to evaluate the generalizability and confirm the performance ratings of each methodology, as shown in Table 6. This analysis employs validation accuracy and standard deviation scores. We partitioned the entire dataset into 10 segments for validation purposes. Among the methodologies tested, only LGBM achieved high k-fold accuracy scores compared to the other approaches, which received respectable ratings. The proposed technique exhibited a high k-fold accuracy score of 0.98, indicating its effectiveness. After analyzing the data, we determined that all the strategies employed are suitable for predicting review helpfulness and demonstrate strong generalization capabilities.

TABLE 6. The k-fold cross-validation analysis.

Model	Kfolds	Kfold Accuracy
RF	10	0.97
DT	10	0.97
K-NN	10	0.96
LGBM	**10**	**0.98**

F. STATE OF THE ART RESULTS COMPARISON

The state-of-the-art performance results comparison of our proposed study is presented in Table 7. We contrasted our approach with recent research studies published in 2022 and 2023. The analysis demonstrates that our study approach surpassed the current leading studies, achieving excellent accuracy scores of 98%. This analysis definitively indicates that our research excels in detecting product helpfulness compared to previous studies.

TABLE 7. The state of the art results comparison.

Ref .	Year	Proposed Technique	Performance Accuracy
[3]	2022	BERT Model	79%
[10]	2023	Naïve Bayes	85%
Proposed	**2024**	**Novel BERF-LGBM**	**98%**

G. ABLATION STUDY

In this part of the research, we present the outcomes of an ablation study designed to evaluate the effectiveness of our employed machine learning techniques. This was achieved by selectively omitting certain elements from the system. The main goal of this study is to gauge the impact and importance of each individual component on the system's overall functionality and performance.

Specifically, we focused on evaluating the impact of our innovative approaches, BERT and BERF, as illustrated in Figure 7. The line chart-based analysis indicates that removing the BERF features results in significantly lower performance scores, as shown in Figure 7(a). This removal led to diminished performance outcomes, underscoring the importance of the BERF features. Our introduced model notably outperformed state-of-the-art methods, achieving

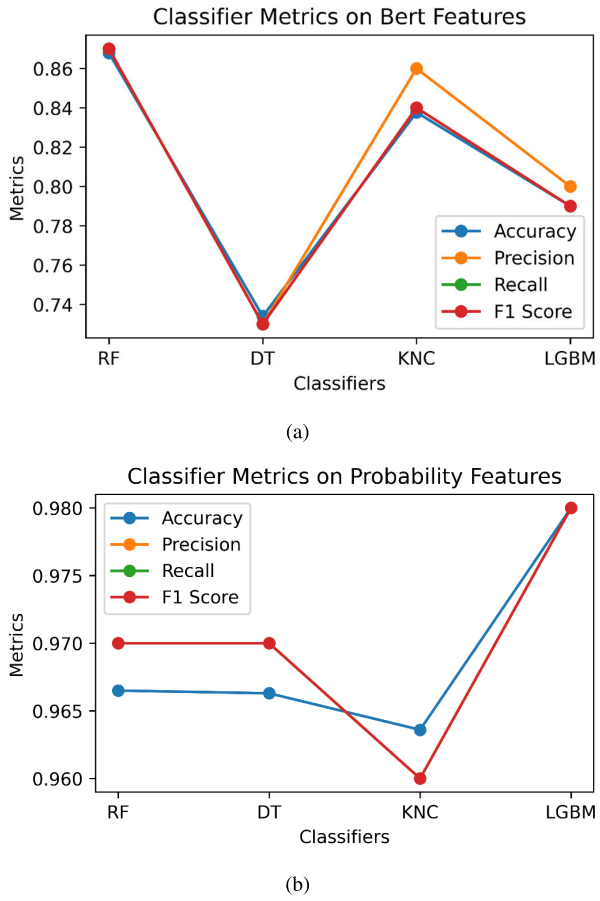

FIGURE 7. Analysis shows a comparison of accuracy, precision, recall, and F1 score for all learning models utilizing the BERT and BERF approaches.

the highest accuracy of 98% in detecting the helpfulness of product reviews. This highlights the superiority of our novel BERF approach over traditional machine learning models in this domain. The results presented in Figure 7(b) indicate that adding our BERF feature substantially improves performance outcomes. Overall, the results from the ablation study reinforce the robustness and efficacy of our approach.

H. DISCUSSIONS AND LIMITATIONS

In our study, we employed a transformer-based novel feature engineering technique, BERF, for the detection of helpfulness in product reviews using machine learning. The analysis was conducted on a state-of-the-art textual online review dataset, where the LGBM model notably outperformed other state-of-the-art methods, achieving an impressive accuracy of 98%.

Despite these promising results, it is important to acknowledge that there are still gaps in performance that need to be addressed. One limitation of our study is the potential for overfitting, given the high accuracy rate, which may not generalize well to unseen data. Additionally, while the LGBM model showed superior performance, the reliance on a single model may not capture the complexity of language and sentiment present in online reviews as effectively as a more diverse ensemble of models might.

Furthermore, our approach with BERF, although innovative, may not fully encapsulate the nuances of human language, suggesting that further refinement of feature engineering techniques is necessary. Future research should also consider the impact of evolving language and context in online reviews, which may affect the longevity and adaptability of the proposed method. Overall, while our findings are significant, continuous efforts in improving and testing the robustness of these methods are essential for closing the identified performance gaps.

V. CONCLUSION AND FUTURE DIRECTIONS

This study utilizes several machine learning methods to address issues related to categorizing the helpfulness of user product reviews. Various feature engineering approaches, including BERT and RF probability, are applied. The classifiers RF, DT, LGB, and KN are trained on text reviews to predict the helpfulness of product reviews. This study introduces an innovative method to accurately predict the helpfulness of Amazon product evaluations. We utilized a benchmark dataset, known as the Amazon Fine Food Reviews, to develop sophisticated machine-learning techniques. Four advanced machine-learning methods were evaluated to determine the helpfulness of product reviews. We propose a new method called BERF (BERT-RF) for feature engineering to enhance the usefulness of customer reviews for Amazon's gourmet food products. A unique feature set is developed using the suggested BERF method. The BERF strategy utilizes class probabilities derived from the review helpfulness denominator dataset as features for constructing applicable machine learning models. The extensive research findings showed that the LGBM methodology surpassed the current leading methods, achieving an accuracy of 98%. The performance of each applicable technique is verified using a k-fold method and further enhanced by hyperparameter optimization.

A. FUTURE WORK

In our upcoming projects, we plan to develop a graphical user interface (GUI) tailored for clients who prefer online shopping. Additionally, we aim to enhance the utility of customer reviews for products by creating a model designed to improve their usefulness. Our efforts will also focus on increasing the accuracy of the model's predictions.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this manuscript. Any affiliations, or relationships with organizations or entities that might pose a conflict of interest with the subject matter discussed in this work are hereby disclosed.

REFERENCES

[1] Y. Jiang, A. Huang, S. Gao, and S. Yu, "Relationship between the terminal built environment and airport retail revenue," *J. Air Transp. Manage.*, vol. 116, Apr. 2024, Art. no. 102568.
[2] A. Alabaidi, "The impact of work life balance on employee attitudes and behavior in health care sector," Tech. Rep., 2024.
[3] X. Zhao and Y. Sun, "Amazon fine food reviews with BERT model," *Proc. Comput. Sci.*, vol. 208, pp. 401–406, Jan. 2022.
[4] J. Ballerini, A. Ključnikov, D. Juárez-Varón, and S. Bresciani, "The e-commerce platform conundrum: How manufacturers' leanings affect their internationalization," *Technol. Forecasting Social Change*, vol. 202, May 2024, Art. no. 123199.
[5] M. S. Akin, "Enhancing e-commerce competitiveness: A comprehensive analysis of customer experiences and strategies in the Turkish market," *J. Open Innov., Technol., Market, Complex.*, vol. 10, no. 1, Mar. 2024, Art. no. 100222.
[6] M. S. I. Malik and A. Nawaz, "SEHP: Stacking-based ensemble learning on novel features for review helpfulness prediction," *Knowl. Inf. Syst.*, vol. 66, no. 1, pp. 653–679, Jan. 2024.
[7] S. Negoita, H. Chen, P. V. Sanchez, R. L. Sherman, S. J. Henley, R. L. Siegel, H. Sung, S. Scott, V. B. Benard, B. A. Kohler, A. Jemal, and K. A. Cronin, "Annual report to the nation on the status of cancer—Part 2: Early assessment of the COVID-19 pandemic's impact on cancer diagnosis," *Cancer*, vol. 130, no. 1, pp. 117–127, Jan. 2024.
[8] H. Zhang, J. Zhao, R. Farzan, and H. A. Otaghvar, "Risk predictions of surgical wound complications based on a machine learning algorithm: A systematic review," *Int. Wound J.*, vol. 21, no. 1, p. e14665, Jan. 2024.
[9] M. Hussain, T. Zhang, M. Chaudhry, I. Jamil, S. Kausar, and I. Hussain, "Review of prediction of stress corrosion cracking in gas pipelines using machine learning," *Machines*, vol. 12, no. 1, p. 42, Jan. 2024.
[10] T. Hudgins, S. Joseph, D. Yip, and G. Besanson, "Identifying features and predicting consumer helpfulness of product reviews," *SMU Data Sci. Rev.*, vol. 7, no. 1, p. 11, 2023.
[11] S. Park and H. Kim, "Extracting product design guidance from online reviews: An explainable neural network-based approach," *Expert Syst. Appl.*, vol. 236, Feb. 2024, Art. no. 121357.
[12] J. Ryu, S. Lim, O. Kwon, and S. Na, "Transformer-based reranking for improving Korean morphological analysis systems," *ETRI J.*, vol. 46, no. 1, pp. 137–153, Feb. 2024.
[13] F. Hjalmarsson, "Predicting the helpfulness of online product reviews," Tech. Rep., 2021.
[14] B. Yu, "Comparative analysis of machine learning algorithms for sentiment classification in Amazon reviews," *Highlights Bus., Econ. Manag.*, vol. 24, pp. 1389–1400, Jan. 2024.
[15] A. Iqbal, R. Amin, J. Iqbal, R. Alroobaea, A. Binmahfoudh, and M. Hussain, "Sentiment analysis of consumer reviews using deep learning," *Sustainability*, vol. 14, no. 17, p. 10844, Aug. 2022.
[16] S. Xu, S. E. Barbosa, and D. Hong, "BERT feature based model for predicting the helpfulness scores of online customers reviews," in *Proc. Future Inf. Commun. Conf. Adv. Inf. Commun.*, Springer, 2020, pp. 270–281.
[17] M. K. Shaik Vadla, M. A. Suresh, and V. K. Viswanathan, "Enhancing product design through AI-driven sentiment analysis of Amazon reviews using BERT," *Algorithms*, vol. 17, no. 2, p. 59, Jan. 2024.
[18] A. Haseeb, R. Taseen, M. Sani, and Q. G. Khan, "Sentiment analysis on Amazon product reviews using text analysis and natural language processing methods," in *Proc. Int. Conf. Eng., Natural Social Sci.*, vol. 1, 2023, pp. 446–452.
[19] F. Rustam, A. Mehmood, M. Ahmad, S. Ullah, D. M. Khan, and G. S. Choi, "Classification of shopify app user reviews using novel multi text features," *IEEE Access*, vol. 8, pp. 30234–30244, 2020.
[20] M. Bilal and A. A. Almazroi, "Effectiveness of fine-tuned BERT model in classification of helpful and unhelpful online customer reviews," *Electron. Commerce Res.*, vol. 23, no. 4, pp. 2737–2757, Dec. 2023.
[21] V. Rupapara, F. Rustam, H. F. Shahzad, A. Mehmood, I. Ashraf, and G. S. Choi, "Impact of SMOTE on imbalanced text features for toxic comments classification using RVVC model," *IEEE Access*, vol. 9, pp. 78621–78634, 2021.
[22] M. Z. Naeem, F. Rustam, A. Mehmood, Mui-Zzud-Din, I. Ashraf, and G. S. Choi, "Classification of movie reviews using term frequency-inverse document frequency and optimized machine learning algorithms," *PeerJ Comput. Sci.*, vol. 8, p. e914, Mar. 2022.
[23] J. Wei, J. Ko, and J. Patel, "Predicting Amazon product review helpfulness," *IEEE Trans. Neural Netw.*, vol. 5, no. 1, pp. 3–14, 2016.
[24] M. B. Kursa and W. R. Rudnicki, "The all relevant feature selection using random forest," 2011, *arXiv:1106.5112*.
[25] SNProject. *Amazon Fine Food Reviews*. Accessed: Feb. 28, 2024. [Online]. Available: https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews
[26] J. J. McAuley and J. Leskovec, "From amateurs to connoisseurs: Modeling the evolution of user expertise through online reviews," in *Proc. 22nd Int. Conf. World Wide Web*, May 2013, pp. 897–908.
[27] A. Raza, A. M. Qadri, I. Akhtar, N. A. Samee, and M. Alabdulhafith, "LogRF: An approach to human pose estimation using skeleton landmarks for physiotherapy fitness exercise correction," *IEEE Access*, vol. 11, pp. 107930–107939, 2023.
[28] F. Rustam, A. Raza, M. Qasim, S. K. Posa, and A. D. Jurcut, "A novel approach for real-time server-based attack detection using meta-learning," *IEEE Access*, vol. 12, pp. 39614–39627, 2024.
[29] A. M. Qadri, M. S. A. Hashmi, A. Raza, S. A. J. Zaidi, and A. U. Rehman, "Heart failure survival prediction using novel transfer learning based probabilistic features," *PeerJ Comput. Sci.*, vol. 10, p. e1894, Mar. 2024.
[30] A. Naseer, M. Amjad, A. Raza, K. Munir, N. A. Samee, and M. A. Alohali, "A novel transfer learning approach for detection of pomegranates growth stages," *IEEE Access*, vol. 12, pp. 27073–27087, 2024.
[31] A. Raza, F. Rustam, B. Mallampati, P. Gali, and I. Ashraf, "Preventing crimes through gunshots recognition using novel feature engineering and meta-learning approach," *IEEE Access*, vol. 11, pp. 103115–103131, 2023.
[32] Y.-T. Peng and C.-L. Lei, "Using bidirectional encoder representations from transformers (BERT) to predict criminal charges and sentences from Taiwanese court judgments," *PeerJ Comput. Sci.*, vol. 10, p. e1841, Jan. 2024.
[33] A. F. Amiri, H. Oudira, A. Chouder, and S. Kichou, "Faults detection and diagnosis of PV systems based on machine learning approach using random forest classifier," *Energy Convers. Manage.*, vol. 301, Feb. 2024, Art. no. 118076.
[34] J. Zhou, Z. Su, S. Hosseini, Q. Tian, Y. Lu, H. Luo, X. Xu, C. Chen, and J. Huang, "Decision tree models for the estimation of geo-polymer concrete compressive strength," *Math. Biosci. Eng.*, vol. 21, no. 1, pp. 1413–1444, 2023.
[35] A. Raza, K. Munir, M. S. Almutairi, and R. Sehar, "Novel transfer learning based deep features for diagnosis of down syndrome in children using facial images," *IEEE Access*, vol. 12, pp. 16386–16396, 2024.
[36] B. Abu-Salih, S. Alotaibi, R. Abukhurma, M. Almiani, and M. Aljaafari, "DAO-LGBM: Dual annealing optimization with light gradient boosting machine for advocates prediction in online customer engagement," *Cluster Comput.*, pp. 1–27, Jan. 2024.

AREEBA ISHTIAQ received the Bachelor of Science and master's degrees in information technology from the Khwaja Fareed University of Engineering and Information Technology (KFUEIT), Rahim Yar Khan, Pakistan, in 2021. During the master's degree, she focused on data analytics, cybersecurity, machine learning, deep learning, web development, and emerged technologies. She looked for internships and part-time jobs during the master's program in order to obtain real-world experience. Employers were pleased by her ability to adapt academic knowledge to practical situations. She hopes to work in IT research and development and contribute to inventions that will influence technology when she finishes the master's degree.

KASHIF MUNIR received the B.Sc. degree in mathematics and physics from The Islamia University of Bahawalpur, Pakistan, in 1999, the M.Sc. degree in information technology from Universiti Sains Malaysia, in 2001, the M.S. degree in software engineering from the University of Malaya, Malaysia, in 2005, and the Ph.D. degree in informatics from Malaysia University of Science and Technology, in 2015. He has been engaged in higher education, since 2002, he taught initially with the Binary College, Malaysia, for a semester, followed by approximately four years with Stamford College, Malaysia. Later, he moved to Saudi Arabia, to work with the King Fahd University of Petroleum and Minerals, from September 2006 to December 2014. In January 2015, he transitioned to the University of Hafr Al-Batin, Saudi Arabia, and in July 2021, he joined the Khwaja Fareed University of Engineering and IT, Rahim Yar Khan, as an Assistant Professor with the IT Department. With a substantial publication record, including journal articles, conference papers, books, and book chapters, he has served on technical program committees for numerous peer-reviewed conferences and journals, contributing to the review of numerous research papers. His research interests include cloud computing security, software engineering, and project management.

ALI RAZA received the Bachelor of Science and M.S. degrees in computer science from the Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology (KFUEIT), Rahim Yar Khan, Pakistan, in 2021 and 2023, respectively. He is currently a Lecturer with the Faculty of Information Technology, Department of Software Engineering, The University of Lahore, Pakistan. He has published several articles in reputed journals. His current research interests include data science, artificial intelligence, data mining, natural language processing, machine learning, deep learning, and image processing.

NAGWAN ABDEL SAMEE received the B.S. degree in computer engineering from Ain Shams University, Egypt, in 2000, and the M.S. degree in computer engineering and the Ph.D. degree in systems and biomedical engineering from Cairo University, Egypt, in 2008 and 2012, respectively. Since 2013, she has been an Assistant Professor with the Information Technology Department, CCIS, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Her research interests include data science, machine learning, bioinformatics, and parallel computing. Her awards and honors include the Takaful Prize (Innovation Project Track), the Princess Nourah Award in innovation, the Mastery Award in predictive analytics (IBM), the Mastery Award in big data (IBM), and the Mastery Award in cloud computing (IBM).

MONA M. JAMJOOM received the Ph.D. degree in computer science from King Saud University. She is currently an Associate Professor with the Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Her research interests include artificial intelligence, machine learning, deep learning, medical imaging, and data science. She has published several research articles in her field.

ZAHID ULLAH received the Ph.D. degree from the University of Kuala Lumpur, Malaysia. He is currently an Assistant Professor with King Abdulaziz University, Jeddah, Saudi Arabia. He is an experienced educator and a Researcher of computer science and information systems. His research interests include machine learning, deep learning, medical imaging, and data science. He has published various articles in his field of specialization.