Kashf Journal of Multidisciplinary Research
Vol:03 – Issue 03 (2026)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

EXPLOITING QUALITY OF PRODUCTS TO IDENTIFY CUSTOMER'S ATTITUDE USING MACHINE LEARNING METHODS

Dr. Amanullah[1], Dr. Muhammad Javed[2], Dr. Muhammad Nawaz Khan[3], *Shoukat Ullah[4], Raja Iqbal[5]

[1]Department of Commerce Education and Management Sciences, Higher Education, Archives and Libraries Department, Peshawar, Khyber Pakhtunkhwa, 25130, Pakistan.
[2, 4]University of Science & Technology, Bannu, 28100, Pakistan.
[3]Department of Smart Security, Gachon University, Seongnam, 13120, Republic of Korea.
[5]Department of Computer Science, Khushal Khan Khattak University, Karak, Pakistan.

[*]***Corresponding Author:***(shoukatscholar@yahoo.com)
DOI:(*https://doi.org/10.71146/kjmr856*)

Article Info

Abstract

With the rise of e-commerce and online shopping, consumers are increasingly relying on the reviews and feedback of other customers to make informed decisions about the products they are considering purchasing. While product descriptions, pictures, and videos can be helpful, they may not provide enough detail or nuance for consumers to fully evaluate a product's quality and complexity. As a result, online shops and independent review websites are collecting and aggregating customer feedback to provide a more comprehensive picture of a product's strengths and weaknesses, which can be especially helpful for consumers who are unable to see or handle a product in person before making a purchase. The proposed work is focused on using natural language processing and machine learning methods to automatically identify the quality and complexity of products using consumers' reviews. The proposed study involves a semantic and contextual analysis of consumers' reviews using an appraisal framework. This approach aims to identify and evaluate the opinions and attitudes expressed by consumers in their reviews of products. The appraisal framework is a linguistic approach that considers how language is used to convey judgments, feelings, and evaluations, and can be used to analyze the opinions expressed in written or spoken text. Using machine learning to analyze customer feedback can be an effective way to extract valuable insights and identify patterns that might be difficult to discern using manual methods. The main contribution of this work is preparation of manually annotated datasets for machine learning methods to automatically identify product quality and complexity. The datasets were tested using machine learning methods: Logistic Regression (LR), Support Vector Machine (SVM), Multinomial Naïve Bayes (NB) and Bernoulli NB. In experiments Multinomial NB with over sampling technique ADASYN is compared with other models and achieving average macro F1 score of 0.77 in respect of quality of product.

Keywords: *Product Quality; Appraisal Framework; Machine Learning Methods; Sampling Technique.*

INTRODUCTION

Online shopping and the internet have revolutionized the traditional offline business by providing web-based platforms for businesses to operate on. E-commerce sites such as Amazon, etc. have made it easier for customers to access a wide range of products from the comfort of their own homes. Customers no longer need to visit physical stores to purchase products as they can easily place an order online and have it delivered to their doorstep [1].

One of the advantages of online shopping is that it provides customers with fast access to a variety of items, a better selection of products, and detailed information about products that they might not be able to find in a physical store. However, one of the problems associated with online shopping is that customers cannot physically test the quality and complexity of a product before making a purchase. To address this issue, businesses have implemented reviewing and feedback systems that allow customers to share their experiences with a product, which can be beneficial for both prior and potential customers. These feedback systems can be implemented through various methods, such as surveys on social media, live chat with customers, offering incentives to customers for their responses, or creating feedback systems on the website itself.

The main aim of this study is to identify product quality and complexity from customers' reviews using machine learning methods. Quality means presence of features of a particular product that satisfy customers' needs. Quality assessment enables stakeholders to improve the standard of their products and make their products more marketable, compete more effectively, gain market share and generate sales revenue. Higher quality enables businesses to lower error rates, rework, field failures, warranty costs, customer unhappiness and other costs. It also reduces the time it takes to launch new products [2]. Based on the appraisal framework, product complexity related with specific aspects of the product that users find difficult or challenging [3]. Here are some examples:

"*This mobile phone has a steep learning curve. It took me a while to figure out how to use all the features.*"

"*The windows setup process was very complicated. I had to read the instructions several times to get everything working.*"

"*I love the functionality of this product, but it's definitely not for beginners. You need to have a good understanding of the technology to make the most of it.*"

"*This product has a lot of moving parts, and it can be challenging to keep everything in good working order.*"

"*The software that comes with this product is very complex. It took me a while to understand how to use all the different menus and settings.*"

The authors in [4] suggested different aspects of product, i.e., internal features, external features, industry standards, reliability, life time, services, customers' response, exterior finish and past performance of the product. In [5] authors employed survey method and examined hypothesis with the help of multiple regression analysis to check the influence of product quality on Pay TV customer happiness and this study concluded that customer satisfaction with Pay TV is positively and significantly correlated with reception quality, content quality, and customer service.

The significant contributions of this paper are summarized as under:

i. To extract different features of product like reliability, durability, function, aesthetics and perceived quality.
ii. To develop annotation guidelines for identification of quality based on quality subcategory in appreciation (reaction) of attitude system in appraisal framework.
iii. To prepare re-annotated datasets regarding nine products customers' reviews in respect to quality of products as high and low quality where high quality related with concepts like good, fine, beautiful, welcome, lovely, durable, reliable, flawless, superb etc. and low quality is associated with concepts like bad, ugly, plain, repulsive, defective, useless etc. of products.
iv. To apply machine learning methods for automatic identification of the quality of product.
v. Sampling Techniques are used to handle class imbalance problems in order to improve the performance of machine learning methods.

This study exploits appraisal framework to prepare manually annotated dataset for detection of quality of product from customers' response which is explained in subsequent section.

APPRAISAL FRAMEWORK

According to language perspective, appraisal framework presents emotional state of the author toward given event and the level of participation through linguistic terms [6]. Attitude is somewhat durable organization of beliefs, thoughts and behavior proclivities in connection with important social things, groups, occurrences or symbol [7]. As explained in Fig. 1, appraisal framework examines the attitude of people in a written text. There are three basic sub-domains of attitude system in appraisal system, i.e., Appreciation, Affect and Judgment. Affects are emotion based, Judgment deals with evaluation of persons to extract their opinions and Appreciation copes with extraction of opinions related with physical things, process etc. Appreciation is further divided into three main categories i.e. reaction, composition and valuation. Impact and quality are the two main sub-domains of reaction. Impact associated with opinions regarding attractiveness and unattractiveness of things. Quality concerns with liking and disliking of things, e.g. beautiful, elegant, hideous etc. Composition is further divided into two main parts, i.e., complexity and balance. Complexity is about opinions containing complication and simplicity of things. Balance dealt with opinions like balanced and unbalanced of things. Valuation, as its name implies, containing opinion related to concepts like innovations, useless, cost, shoddy etc. of things [3]. As given in Fig. 1, attitude of customers toward quality of product is predicted based on appraisal framework. For instance, in the customer's review "Software that came with MicroMP3 is awful", where adjective term "awful" belongs to the quality subcategory of attitude system in appraisal framework. Therefore, this review is annotated as low quality. For instance, in the customer's review "picture quality of the camera is good in automatic mode", where adjective term "good" belongs to the quality subcategory of attitude system in appraisal framework. Therefore, this review is annotated as high quality.

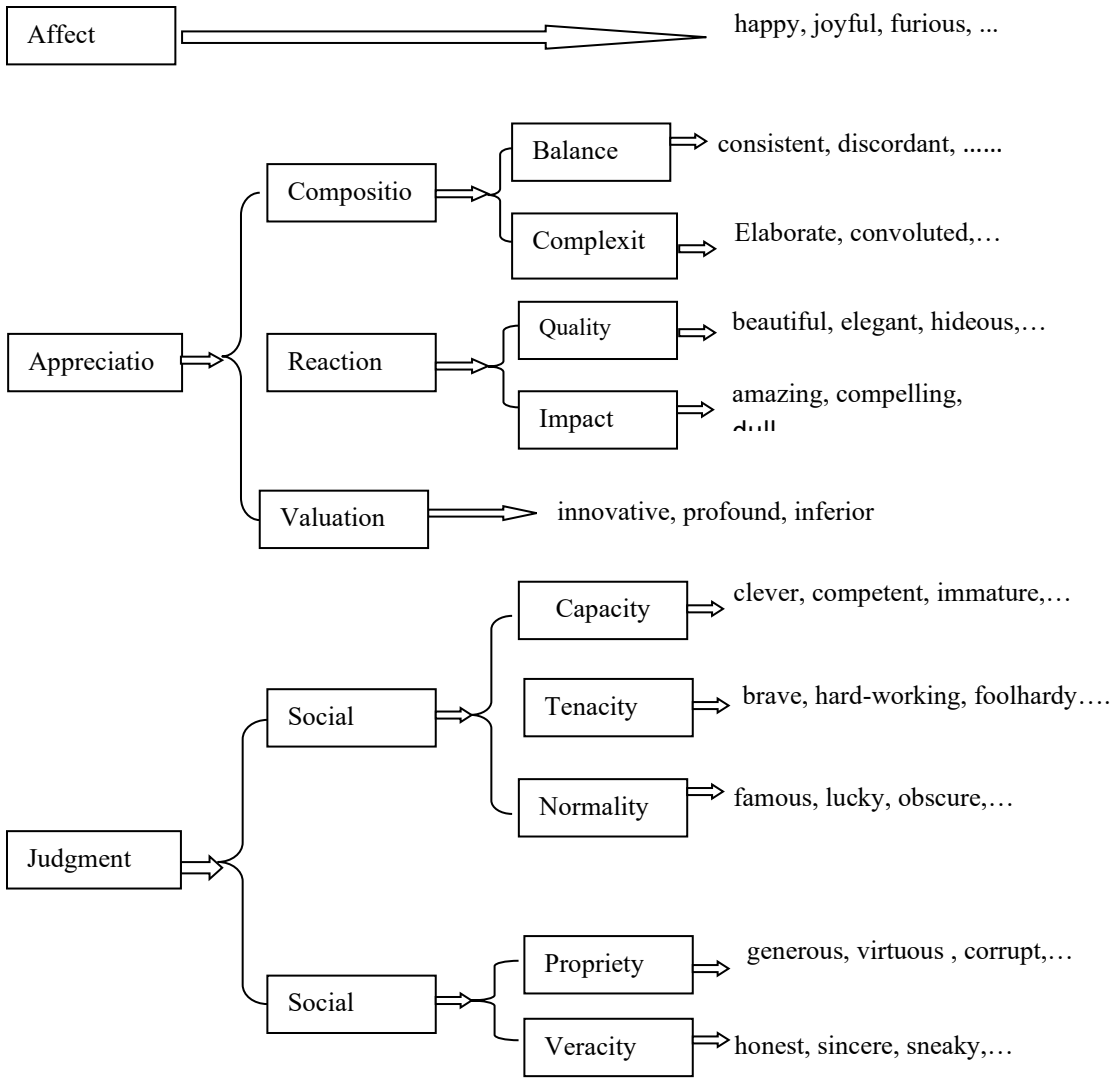

Figure 1: Attitude System of Appraisal Framework

RELATED WORK

Social media is a big source of information. To analyze its contents is very laborious task through traditional data mining tools because their contents are huge, noisy, unstructured as well as not similar to each other's. Conventional tools of data mining are slow, costly, depend on size and biased as well [8]. Mining social media contents deals with people's opinion, attitude and emotion identification. Social media contents mining revolves around two main concepts, i.e., sentiment analysis and opinion mining. The word sentiment analysis first introduced in [9] and the term opinion mining was initially used in [10]. Though, there is major disagreement about the limitation between these two fields. Authors also believed that under opinion mining text mining techniques are used to discover exciting and intuitive correlation among opinions of authors [11]. While, sentiment analysis is considered as sentiment classification, related with categorization of text, or part of text, based on computing the amount of individual's opinion and accurate information contained in the text and orientation [12]. Electronic word-of-mouth is about sharing of information by customers unceremoniously on internet in respect to particular features of products, services or about vendor [13]. Customers in their reviews state their knowledge, approval and opinions about goods or services on difference web-based platforms e.g. social webs, blogs etc. [14]. Web-based reviews generate an opinion, gives user detail and useful in making assessment for perspective buyers [15]. Processing of text automatically in order to evaluate attitude, emotions etc. of the peoples in the main purpose of the natural language processing. Opinion mining is related with extractions of sentiment and topic discussed within text. The authors in [16] used text mining approach to predict consumers' attitude based on web reviews and also explore words to find positive and negative attitude of consumers toward the hotel. In this approach important features related with positive and negative attitude of consumers were extracted. These important features significantly assisted the marketers to plan keywords selections in their marketing policies. While recognition of negative keywords associated with the challenging areas of the business. In [17] point of view, development of business is mainly based on recommendations which are the sole major predictors.

Due to familiarity of Internet and attainability of big data, customers depend on web-based reviews to find details of products or services. Customers connect to web-based reviews to avert indecision on purchasing alternatives [18, 19]. Rapid upsurge of social networks and its related amount of data generate a no more static activity in the fields of sentiment analysis and methods evolved can be generally dissected into lexicon-based approach [20] and machine learning approach [21]. In [22] authors presented a classifier based on two steps for twitter data. This classifier classified the tweets into two main groups, i.e., subjective and objective and then differentiated the subjective group as positive or negative. In order to minimize the cost of labeling manually it uses the noisy labels offered by some websites which deals with detecting sentiment on Twitter data. It proposed feature-set comprised of Meta data about words of tweet and different attributes of writing of tweets to form a most compact representation of tweets. In[23] authors examined the selection of online product by the consumers from web-based product recommendation system. The authors found that consumers recommended products twice as frequently as compare to consumers who did not check the recommendations from web-based product recommendations system. Online recommendations of products from consumers are very beneficial for marketers too.

Large quantities of information regarding reviews of consumers are present on web. Recommendations of consumers are also present in these reviews. These textual reviews give information to money-spinner and it shows set of assessment variables about presence of recommendations or not. With the huge information on web, data mining approaches are used to extract concealed information regarding customer behaviours [24, 25]. The

authors in [26] exploited Twitter data to assess sentiment of consumers in respect to renowned brands. [27, 28] obtained features of the products and envision of the market formation by using text mining techniques. [29, 30] also implemented text mining technique to complement numerical data in order to envisage sales of the product. In [31] authors assessed new quality value by gathering web-based comments of the consumers in regard to shopping and then implemented text classification method with fuzzy comprehensive evaluation method. This evaluation procedure assisted the consumers to improve their decision making for the purchase of suitable products. The authors in [32] prescribed innovative architecture for inventor by exploiting web-based reviews for design of the product depending on the reality based online item reviews of a sample product relating to precision, comparison and rationality. The authors [33] endeavoured to originate some significant data from asset reviews to increase aspects of an asset and assists in enhancing support quality and customer understanding. Author applied different text mining techniques to analyze reviews of customers of a product to see regularly occurred problems and tendency of problems over a particular duration of time. [34] identified the part of text mining to extricate diverse parameters of item accessible on diverse sources in unstructured frame and show machine learning method for improving the item quality. [35] investigated the relation between quality of products and satisfaction of customers. The appraisal theory examined by [6] for social media contents analysis by going deep into conventional sentiment analysis and opinion mining. The main theme focused on appraisal categories like attitude etc, to explore the extreme attitude, extreme beliefs and thoughts of web users related to cyberspace in order to support defense and homeland protection. Building semantic resource, model sheer categories and draw specific attention to provisions and words related to each category were three main systems based on appraisal theory. Keywords related to hate, racism etc were collected from tweets and analysis using resource. The authors in [36] explored improving personalized recommendation systems by leveraging Graph Attention Networks (GAT), which dynamically adapt to user preferences. The approach is driven by perceived complexity and innovation, enhancing prediction accuracy and capturing evolving user interests. In [37], authors examined the evolution of sentiment analysis techniques in product reviews to enhance intelligent recommendation systems. It highlighted how advanced sentiment models help in accurately capturing user opinions, leading to more personalized and effective product recommendations. Similar studies in [38] investigated the perceived quality of mobile apps by applying the Distil BERT model to analyze user reviews. It demonstrates how the model effectively captures user sentiments and provides insights into app quality, aiding developers in improving user experiences.

Author explored social media contents with the help of appraisal system developed by [39]. Our focal point of this research is to exploit reviews of the customers in order to evaluate their responses toward product quality based on appraisal framework. We used the benchmark customers' reviews datasets of nine Amazon products and then performed preprocessing, selection of features etc. to make data suitable for machine learning methods. Major recommendations of customers toward quality of product are high and low. Independent variables in machine learning model are different combinations of words based on quality category of appraisal theory in reviews of users. Different machine learning models are employed for predictive accuracy which was high by examining at different situations. From literature it is observed that most of existing research for product reviews system has been done using traditional sentiment analysis and opinion mining but, in this research, we develop a novel idea of identifying quality of product as low- or high-quality using appraisal framework from customers' responses.

METHODOLOGY

This section presents step by step methods of the proposed technique. The processes workflow is represented in Fig. 2 and is further explained in the following subsections.

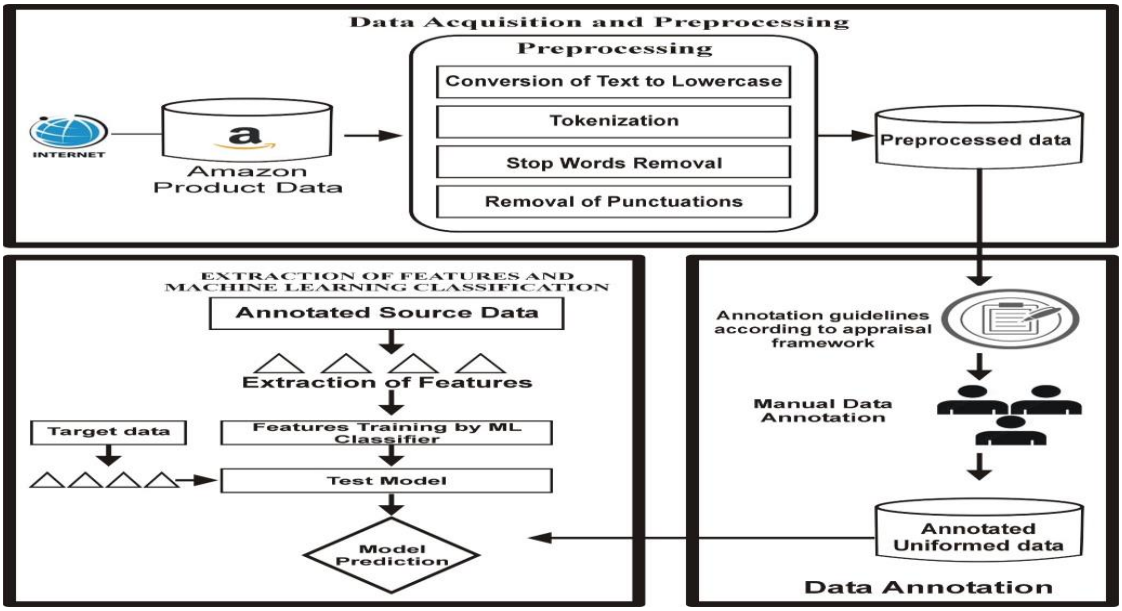

Figure 2: Methodology for Customers' Responses toward Product Quality

Data Acquisition and Preprocessing

In this research we used benchmark nine Amazon product reviews datasets [40]: Linksys, Hitachi Routers, Creative Zen MicroMP3 Player, Canon Power Shot SD500, Canon S100, Diaper Champ, iPod, Nokia 6600 and Norton Antivirus. These datasets, which originally served for the purpose of sentiment analysis, included 2656 customer reviews. Using an appraisal framework, we re-annotated the aforementioned datasets to assess product quality.

Dataset Development

Creating annotation guidelines

To ensure that the dataset is consistently annotated, standards and criteria are defined. If clear rules are in place, the expert may quickly add correct labels to the data, producing high-quality data annotation. To gauge the quality of an annotation, the highest inter-annotator agreement is employed. The primary step in developing annotation rules is to explicitly define the labels of class quality as low quality and high quality or differentiating between customers' reviews about product quality as low quality or high quality. We developed following annotation criteria to determine product quality based on the appraisal framework as part of this study. These guidelines enabled the experts who annotated the data to differentiate between low-quality and high-quality data and to label the data accordingly.

Guidelines for High Quality

A review is assigned high quality if it has dominating number of positive sentiment words explicitly or implicitly as compared to negative sentiment words for instances; "Audio is mono and fairly acceptable, quiet also with little distortions. I have been fairly impressed with its ability to capture decent details in low light situations such as theaters, inside a bus at night."

Guidelines for Low Quality

A review is assigned low quality if it has dominating number of negative sentiment words as compared to positive sentiment words according to appraisal framework; for instance; "I've been using Norton Anti-Virus in its many

incarnations for over 5 years. Norton Internet Security 2004 is one of the buggier implementations. Even for computer super-heroes it's a pain to maintain when there is a problem. It failed whenever connecting to get updates while it has a rather effective firewall."

Guidelines for No Class Label

No Quality label is reserved for reviews that have no clear customers' responses about the quality of the product, and should be treated as undetermined; for instance; "The world of MP3 players is blinding, and since most people don't understand the technology, myself included."

Guidelines to omit the reviews

Leave out the kinds of customer reviews that are written in a language other than English. Also omit the reviews which are ambiguous and unclear to annotate.

Manual annotation

By employing the aforementioned annotation criteria, we created a base annotated dataset in .csv files that still needed to be verified by an expert team.

Verification of manual annotation by experts

The datasets were made available to two annotators to annotate the data as high quality and low quality. The annotators received the annotation guidelines developed for this study.

Annotation completion based on inter-annotator consensus

The final annotation was chosen based on the assessment of the highest inter-annotator agreement after obtaining dataset annotations from two annotators. Cohen's kappa was used to determine the Inter-Annotator Agreement in this study because there are two annotators.

Preprocessing

The core objective of natural language processing, a branch of artificial intelligence, is to comprehend human-comprehensible languages. For obligatory computational task, text data required to be converted into format that computer understands [41]. Preprocessing of text is concerned with eliminating unimportant features and extracting important features from the text so that machine learning model can build up efficiently and accurately. To improve the performance of machine learning classifier datasets were preprocessed by the help of following steps and then fed into the machine learning model

Tokenization

Each sentence of the customer review was divided into token called word using delimiters like semicolon, colon, comma, dot etc. Each token did not depend on other token and also has a separate meaning.

Conversion to lowercase

Convert all the text of reviews into lower case for extraction of same meaning for both upper and lower case [42].

Punctuations removal

For analyzing the text, punctuations have no importance, therefore we omitted the punctuations in the text.

Removal of stop words

We removed the stop words because of their no important role in classifier.

Product Quality Identification

As described in Table 1, features of products: reliability, durability, function, aesthetic and perceived based on the quality subcategory of attitude system in appraisal framework from Amazon nine product customers' reviews datasets are extracted and label each feature as low quality or high quality. Fig. 3 presented customers' response toward product' features based on appraisal framework as high quality and low quality.

- Canon PowerShot SD 500 has 85% and 15% high quality and low-quality features respectively from customers' responses.
- iPod has 71% and 29% high quality and low-quality features respectively from customers' responses.
- Create Zen MicroMP3 has 68% and 32% high quality and low-quality features respectively from customers' responses
- Diaper Champ has 73% and 27% high quality and low-quality features respectively from customers' responses.
- Link Sys Router has 71% and 29% high quality and low-quality features respectively from customers' responses.
- Hitachi Router has 73% and 27% high quality and low-quality features respectively from customers' responses.
- Canon S100 has 73% and 27% high quality and low-quality features respectively from customers' responses.
- Norton Antivirus has 30% and 70% high quality and low-quality features respectively from customers' responses.

The bar graph as shown in Fig. 4 presents the comparison between percentages of nine products customers' reviews datasets as high and low quality.

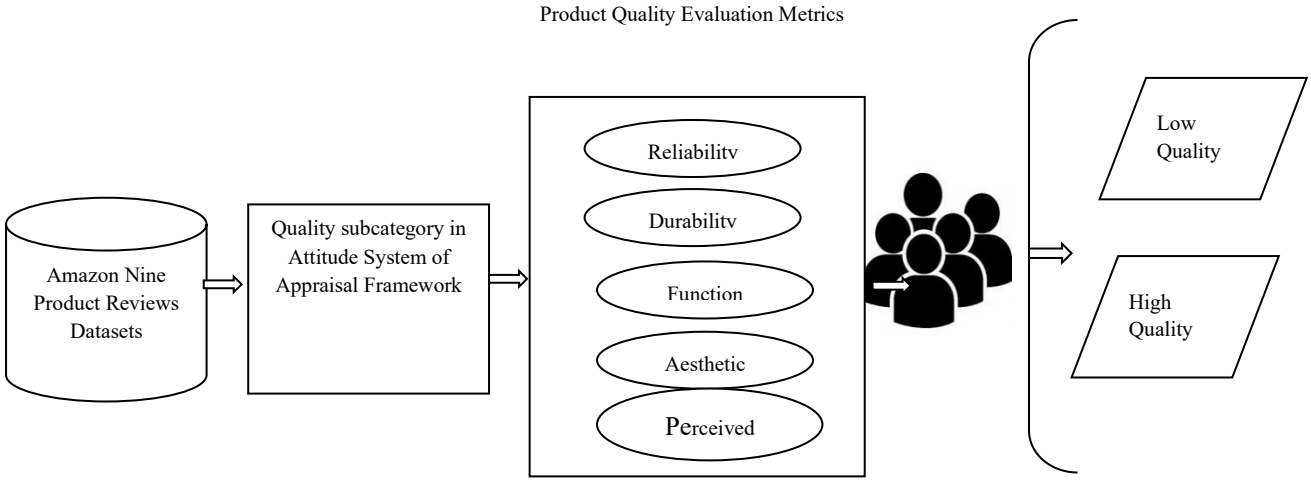

Figure 3: Evaluation of Product Quality based on Appraisal Framework

Table 1: Detail about features of products as high quality and low quality

Product Name/ Quality->	Reliability		Durability		Function		Aesthetics		Perceived		High%	Low%
	High	Low	High	Low	High	Low	High	Low	High	Low		
Canon PowerShot SD 500	5	2	1	1	3	1	38	5	10	1	85	15
iPod	3	2	4	1	7	3	73	29	1	1	71	29
Creative Zen MicroMP3	3	4	3	8	10	6	189	79	3	1	68	32
Diaper Champ	1	1	1	2	3	2	57	18	0	0	73	27
Link Sys Router	2	1	2	2	10	5	57	22	1	0	71	29
Hitachi Router	4	3	5	1	7	3	51	18	1	0	73	27
Canon S100	5	2	6	3	4	2	43	15	1	0	73	27
Norton Antivirus	10	11	1	3	0	1	23	65	1	0	30	70
Nokia 6600	2	2	9	7	5	2	144	49	3	0	73	27

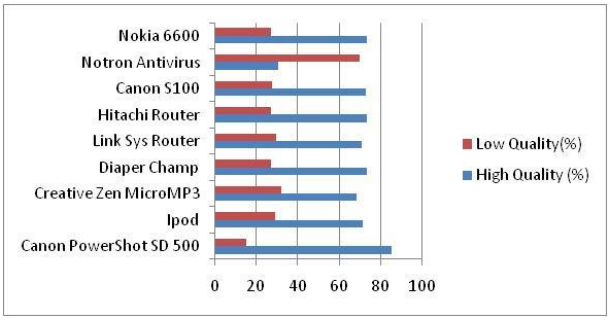

Figure 4: Visualization of Products Quality in Percentages

Feature Extraction Technique

The key process of extracting features comes after pre-processing in the analysis of raw data. The computer does not directly alter the raw data; rather, it transforms it into derived numerical values while preserving the original data's information. In this result we used TF-IDF (Term Frequency-Inverse Document Frequency) technique for features extraction.

Term Frequency-Inverse Document Frequency (TF-IDF)

Machine learning methods customarily work way better with number. TF-IDF gives assistance to machine learning methods by assigning vectors to each word in the document which also shows importance of each word is in that document. The document with comparative, important words will have comparable vectors. TF-IDF is the multiplication of number of instances of each word in the document and logarithm of add up to number of documents dividing the document containing the word. TF-IDF algorithm revolutionized machine learning methods. In the proposed model we applied TF-IDF on the data to improve the performance result of machine learning methods. Given a dataset C, word w, and document records d ∈ C, the formula to find TF-IDF is shown

below [43] where w_d indicates the significance of the provided term, and *f*(w, d) represents the frequency with which a word appears in the dataset C.

$$w_d = f(w,d) \text{ x } log \frac{|C|}{f(w,C)} \qquad (1)$$

Machine Learning

Machine learning is the branch of artificial intelligence that seeks to teach a computer to learn new things, (primarily data patterns) from earlier data, just like humans learn new knowledge from their prior experience. The methods of machine learning can be either supervised or unsupervised. In supervised learning methods, learning is done by using features and class of new dataset will be predicted from previously learnt features. In unsupervised learning methods, there are no target features to learn. This type of learning examine data for intriguing patterns or natural groups [44]. In this research we employ machine learning methods including LR, SVM, Multinomial NB and Bernoulli NB.

Logistic Regression

Logistic regression is a supervised machine learning technique. It is the extension of linear regression for performing classification. It uses the ideas' sigmoid function and linear regression and mainly suited for binary classification. The result of linear regression is in the range between 0 and 1 [45]. Logistic regression is calculated by following formula.

$$\text{s(x)} = \frac{1}{1 + e^{-x}} \qquad (2)$$

where X is an input variable for sigmoid function s(x)

Support Vector Machine

SVM is a machine learning technique. It is well suited for binary classification. It creates hyper plane that can divide different types of data that are characterized having more than one dimension into similar groups. Kernel function like polynomial etc. in SVM created the vectors. In order to enhance the performance of classification, user selects the best vectors to be fitted on the particular corpus [46]. Fig. 5 shows the basic structure of SVM.

Figure 5: Basic structure of Support Vector Machine

Multinomial Naïve Bayes and Bernoulli Naïve Bayes

Multinomial NB is a variation of NB machine learning algorithm for solving classification problems. It utilized the NB algorithm to handle multinomial distributed data [47]. Main theme of Multinomial NB machine learning algorithm is to check the existence of term in a document and if the term exists then count its number of occurrences. While multivariate Bernoulli NB machine learning algorithm is associated with autonomous binary concepts which only check the existence of term in the under-discussion document [48].

EXPERIMENTS

For experimental results we used python language using Anaconda, Jupiter Notebook. This section comprises detail of datasets, evaluation metrics, sampling techniques and results of experiment and discussions.

Dataset

In this research study we employ publically available Amazon nine product datasets. The datasets comprised of 2656 customers' reviews as shown in Table 2.

Table 2: Description of Overall Datasets

Product Name	Total Reviews
Canon PowerShot SD 500	160
iPod	246
Creative Zen MicroMP3	629
Diaper Champ	238
Link Sys Router	222
Hitachi Router	254
Canon S100	219
Norton Antivirus	240
Nokia 6600	448

Within these nine product reviews datasets, only 45.03% of all reviews are related with quality of products. Each of aforementioned dataset is split into two sets, a training set and a testing set, for the experiment. In this study, we used the "train_test_split" function of the Python Sklenar package to divide each dataset into Training Set and Testing Set with a ratio of 80:20, respectively.

Training Set

In order to uncover hidden patterns in the data, a training dataset was employed to train the machine learning model. Large amounts of data should be included in the training dataset so that the model may be trained on all conceivable scenarios and be able to predict hidden data sample accurately. In this research, 80% of each dataset was used as training data.

Selection of machine learning models

To identify quality of product automatically, we applied machine learning models i.e. LR, SVM, Bernoulli NB and Multinomial NB.

Testing Set

Each trained model is given a 20% testing dataset for model evaluation, and the result in each case are recorded.

Evaluation Metrics

The researcher classified the customers' reviews based on the appraisal framework in order to evaluate the quality of products as high quality or low quality. Majority of Nine product datasets are highly imbalanced in which high accuracy is predicted based on the majority class. [49] used different metrics to assess the performance regarding model prediction involving f1_micro, f1_macro, f1_weighted, accuracy, and confusion matrix. These metrics are regularly used to handle problems related with unbalanced data. The F1_micro refers to computing global average F1 score. The researcher has computed the overall True Positive, False Positive and False Negative and then fed it into F1 equation to achieve micro F1 score. F1_weighted score is computed by taking into account each class's support with mean of f1 scores of each class**.** Main theme of the precision is to minimize false positives and the main focus of the recall is to minimize false negatives. The main goal of the imbalanced learning is to increase recall score without hurting precision. F1-score combines the properties of recall and precision. F1_macro represents the arithmetic mean of per-class F1-scores.

Table 3: Confusion Matrix

		Actual Result	
		1	0
Predicted Result	1	TP	FP
	0	FN	TN

where TP= True Positive, FP = False Positive, FN= False Negative, and TN=True Negative

Confusion matrix as shown as Table 3 in the binary classifier is the combination of rows and columns (2 x 2) where columns indicate actual result and rows represent predicted result of the classifier. TP refers to the correct prediction of the positive class instances, FP deals with incorrect prediction of positive class instances, FN refers to incorrect prediction of negative class instances and TN shows correct prediction of negative class instances. Recall alludes to the rate of quality related sentences accurately distinguished by a predictive model. Recall is calculated when researcher divides total number of true positive sentences by sum of true positive and false negative sentences. Precision alludes to the sentences showing low and high quality amid all the sentences predicted by the model as quality related sentences based on classification. Precision is achieved when we divide total number of true positive sentences by sum of true positive and false positive sentences. F1 score is the harmonic mean of Precision and Recall to evaluate the model performance.

Sampling Techniques for Imbalanced Classes

Class unbalanced is a common issue in machine learning, where number of samples in one class is far lessened than the other class [50]. In this research the datasets we used are highly imbalanced as shown in Table 4 where majority class is much bigger than the minority class. Due to which computational model is unable to predict accurately. In order to overcome this problem, we applied over sampling techniques like random over sampler, SMOTE, ADASYN and under sampling techniques like random under sampler and Tomek Links with machine learning methods to enhance the performance level of the machine learning classifiers. In these approaches, we will either increase the samples of minority class or decrease the samples from majority class. [51] compared various oversampling techniques like SMOTE (Synthetic minority oversampling approach), ADASYN, Borderline-SMOTE, Safe-Level SMOTE in combination with classification techniques.

Table 4: Statistics about Annotated Dataset

Product name	Reviews related with quality of product	Majority class reviews	Minority class reviews
Canon PowerShot SD 500	67	57	10
iPod	124	88	36
Creative Zen MicroMP3	306	208	98
Diaper Champ	85	62	23
Link Sys Router	102	72	30
Hitachi Router	93	68	25
Canon S100	81	59	22
Norton Antivirus	115	80	35
Nokia 6600	223	163	60

Synthetic Minority Oversampling Technique (SMOTE)

SMOTE is the most dominant technique that produces synthetic data based on the similitude among space containing all feature vectors of the presented minority class observations. This technique generalizes the decision range of minority class by generating arbitrary points among two already mentioned vector position [52].

Adaptive synthetic (ADASYN) sampling technique

In case of hardship in learning adaptive synthetic (ADASYN) sampling technique robustly make a decision regarding number of artificial observations required by the minority class in comparison of those minority observations which are not difficult to learn. This algorithm decreases the bias level that is occurred due to imbalanced classes and increases the learning level adaptively [53].

Tomek Links

Tomek Links is an under-sampling technique, which is develop by [54]. This technique is used to search all those instances from the majority class that are having minimal Euclidean distance with the class having smaller number of instances. If the data found in majority class closer with data of minority, then it will be ambiguous to separate from each other and then consequently remove the data.

Random Over Sampler and Random Under Sampler

Random Oversampler chooses instances of minority class, with substitution and then included to the training set or we can say it makes copies of the instances from minority class arbitrarily and then added to the training set. Random Under Sampler selected observations from majority class arbitrarily and then erasing them from the training set.

RESULTS AND DISCUSSIONS

The detail of experimental setting, experimental setup and findings are included in this section.
We carried out this study by using Jupiter Notebook in Anaconda with python version 3.0, 16 GB of RAM and a 250 GB hard drive. We applied different Python libraries, along with Sklenar, Gensim, Pandas, NumPy, and NLTK.

In this study, we employed SKlearn package to conduct experiment on the datasets and customers' reviews toward product quality are categorized as low quality and high quality. Features are retrieved through TF-IDF vectorizer, and later four classifiers LR, SVM, Bernoulli NB and Multinomial NB are applied for the classification of customers' reviews about the product quality. These classifiers are trained and tested on Amazon nine product reviews datasets. Moreover, we employed over sampling and under sampling techniques with aforementioned models because of highly imbalanced datasets. The training and testing datasets are split 80/20 correspondingly. For experiment default setting is applied for machine learning classifiers.

Our research study contributes to the existing related literature by providing a novel approach to identify customers' responses toward product quality. To exhibit the superiority and to check the predictive power of our proposed model to classify customers' responses about the product as high quality or low quality based on appraisal framework, we compared the performance of four machine learning methods i.e., LR, SVM, Multinomial NB and Bernoulli NBs with various over sampling techniques i.e. Random Over Sampler, SMOTE, ADASYN and under sampling techniques i.e. Random Under Sampler and Tomek Links. We adopted for sampling techniques in our model to address the class imbalance problem because majority class has highest ratio of reviews as compared to minority class. Table 5 show the performance results of machine learning methods with sampling techniques on the basis of evaluation metrics: f1 micro, f1 macro, f1 weighted and accuracy where f1 macro score is commonly used metric for learning from imbalanced data and to evaluate the performance of classification model. Figure 6 graphically presented the average f1 macro score of each of the machine learning method in combination with sampling techniques.

Table 5: Performance results of Machine Learning Methods with Sampling Techniques

Classifiers	Sampling Techniques	f1 micro	f1 macro	f1 weighted	accuracy
Logistic Regression	Random Over Sampler	0.73	0.58	0.68	0.73
	SMOTE	0.73	0.57	0.67	0.73
	ADASYN	0.76	0.61	0.71	0.76
	Random UnderSampler	0.61	0.58	0.60	0.61
	Tomek Links	0.68	0.42	0.57	0.68
SVM	Random OverSampler	0.74	0.58	0.68	0.74
	SMOTE	0.75	0.59	0.69	0.75
	ADASYN	0.76	0.62	0.71	0.76
	Random UnderSampler	0.61	0.58	0.61	0.61
	Tomek Links	0.70	0.49	0.61	0.70
Bernoulli Naïve Bayes	Random OverSampler	0.70	0.56	0.65	0.70
	SMOTE	0.70	0.48	0.61	0.70
	ADASYN	0.70	0.47	0.60	0.70
	Random UnderSampler	0.61	0.56	0.59	0.61
	Tomek Links	0.69	0.50	0.61	0.69
Multinomial Naïve Bayes	Random OverSampler	0.76	0.71	0.76	0.76
	SMOTE	0.80	0.76	0.80	0.80
	ADASYN	0.81	**0.77**	0.80	0.81
	Random UnderSampler	0.68	0.65	0.69	0.68
	Tomek Links	0.68	0.40	0.56	0.68

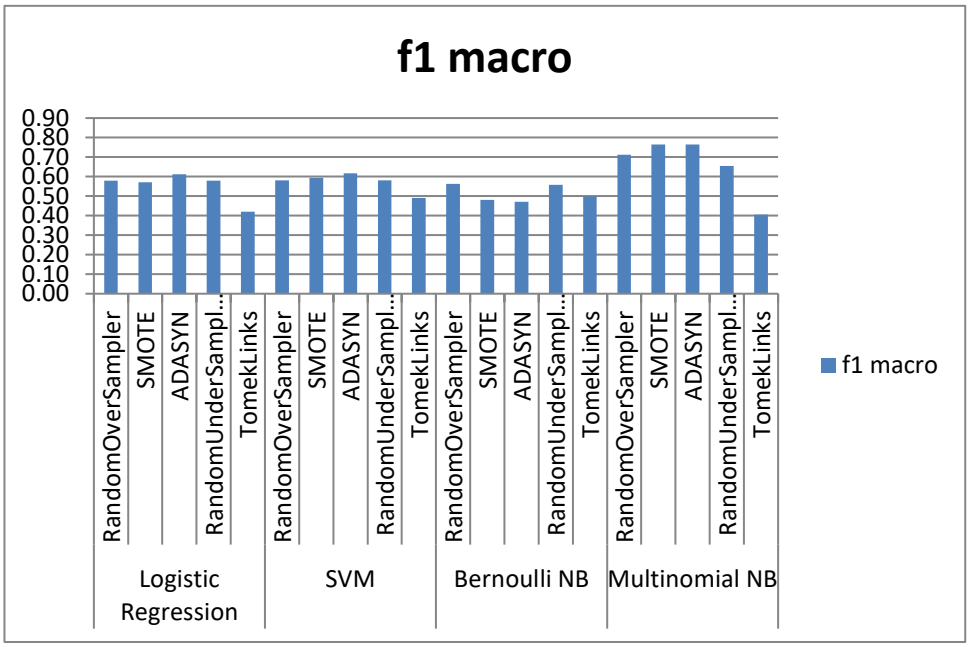

Figure 6: Visualization of f1 macro score of machine learning methods with sampling techniques

Average f1 macro score of 0.77 of Multinomial NB machine learning model in combination with ADASYN over sampling technique outperforms other three state-of-the-art machine learning models with sampling techniques to classify customers' responses toward quality of product based on appraisal framework, as described in Table 5.

LIMITATIONS AND FUTURE RESEARCH

Our research study has also few restrictions. Our labeled data comprised of 1196 out of 2656 are related with quality of product based on appraisal framework in nine diverse product reviews datasets. There is also major issue of class imbalance in our nine product datasets. Future work should be regarding more data collection in respect to quality-based product reviews under the domain of appraisal framework and various combination of embedding methods with LSTM model in order to more semantic and contextual depiction of words related with quality of the product could improve the performance of model.

CONCLUSIONS

In this paper, we propose an effective computational model for mining customers' responses toward product quality in reviews based on appraisal framework. First, we present an idea of mining quality of product based on customers' responses. Second, we extract different features regarding quality of product. Third, we provide a novel idea of identifying customers' responses toward product quality based on appraisal framework. We compare various over sampling and under sampling methods with machine learning methods. Our proposed model comprises of ADASYN over sampling technique with Multinomial NB machine learning method to detect quality of product by experimenting on Amazon nine produce reviews datasets. Experiment results shows that our proposed model exhibit greater performance among different models to classify customers' responses toward quality of product as low quality and high quality.

REFERENCES

[1] A. Ullah, K. Khan, A. Khan, and S. Ullah, "Understanding quality of products from customers' attitude using advanced machine learning methods," *Computers,* vol. 12, no. 3, p. 49, 2023.

[2] S. Khoja, A. Prajapati, S. S. Khoja, K. R. Panjwani, and J. Ray, "A review on quality aspects, evolution of quality, dimension of quality and action plan for enhancing quality culture," *Pharma Science Monitor,* vol. 8, no. 4, pp. 335-352, 2017.

[3] B. Liu, *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge university press, 2020.

[4] D. A. Garvin, *Managing quality: The strategic and competitive edge*. Simon and Schuster, 1988.

[5] C. Udeh, C. Ifekanandu, E. Idoko, C. Ugwuanyi, and C. Okeke, "Pay TV Product quality and customer satisfaction: An investigation," *International Journal of Information Systems and Informatics,* vol. 3, no. 1, pp. 25-35, 2022.

[6] V. Dragos, D. Battistelli, and E. Kelodjoue, "Beyond sentiments and opinions: exploring social media with appraisal categories," in *2018 21st International Conference on Information Fusion (FUSION)*, 2018, pp. 1851-1858: IEEE.

[7] M. V. Hogg, "Social Psychology " *Prentice Hall: London - chapter 5.,* vol. 4th Edition, 2005.

[8] B. Liu, *Sentiment Analysis and Opinion Mining*. 2012.

[9] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in *Proceedings of the 2nd international conference on Knowledge capture*, 2003, pp. 70-77.

[10] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," in *Proceedings of the 12th international conference on World Wide Web*, 2003, pp. 519-528.

[11] S. Argamon, K. Bloom, A. Esuli, and F. Sebastiani, "Automatically determining attitude type and force for sentiment analysis," in *Language and technology conference*, 2007, pp. 218-231: Springer.

[12] C. Whitelaw, N. Garg, and S. Argamon, "Using appraisal groups for sentiment analysis," in *Proceedings of the 14th ACM international conference on Information and knowledge management*, 2005, pp. 625-631.

[13] S. W. Litvin, R. E. Goldsmith, and B. Pan, "Electronic word-of-mouth in hospitality and tourism management," *Tourism management,* vol. 29, no. 3, pp. 458-468, 2008.

[14] K.-H. Yoo and U. Gretzel, "Influence of personality on travel-related consumer-generated media creation," *Computers in human behavior,* vol. 27, no. 2, pp. 609-621, 2011.

[15] B. Von Helversen, K. Abramczuk, W. Kopeć, and R. Nielek, "Influence of consumer reviews on online purchasing decisions in older and younger adults," *Decision Support Systems,* vol. 113, pp. 1-10, 2018.

[16] S. Zhao, "Thumb Up or Down? A Text-Mining Approach of Understanding Consumers through Reviews," *Decision Sciences,* vol. 52, no. 3, pp. 699-719, 2021.

[17] F. F. Reichheld, "The one number you need to grow," *Harvard business review,* vol. 82, no. 6, pp. 133-133, 2004.

[18] T. Hennig-Thurau and G. Walsh, "Electronic word-of-mouth: Motives for and consequences of reading customer articulations on the Internet," *International journal of electronic commerce,* vol. 8, no. 2, pp. 51-74, 2003.

[19] H. Tan, X. Lv, X. Liu, and D. Gursoy, "Evaluation nudge: Effect of evaluation mode of online customer reviews on consumers' preferences," *Tourism Management,* vol. 65, pp. 29-40, 2018.

[20] R. Socher *et al.*, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proceedings of the 2013 conference on empirical methods in natural language processing*, 2013, pp. 1631-1642.

[21] E. Cambria, C. Havasi, and A. Hussain, "Senticnet 2: A semantic and affective resource for opinion mining and sentiment analysis," in *Twenty-Fifth international FLAIRS conference*, 2012.

[22] L. Barbosa and J. Feng, "Robust sentiment detection on twitter from biased and noisy data," in *Coling 2010: Posters*, 2010, pp. 36-44.

[23] S. Senecal and J. Nantel, "The influence of online product recommendations on consumers' online choices," *Journal of retailing,* vol. 80, no. 2, pp. 159-169, 2004.

[24] S. J. Lee and K. Siau, "A review of data mining techniques," *Industrial Management & Data Systems,* 2001.

[25] P. Hoontrakul and S. Sahadev, "Application of data mining techniques in the on-line travel industry: A case study from Thailand," *Marketing Intelligence & Planning,* 2008.

[26] M. M. Mostafa, "More than words: Social networks' text mining for consumer brand sentiments," *Expert systems with applications,* vol. 40, no. 10, pp. 4241-4251, 2013.

[27] T. Y. Lee and E. T. Bradlow, "Automated marketing research using online customer reviews," *Journal of Marketing Research,* vol. 48, no. 5, pp. 881-894, 2011.

[28] O. Netzer, R. Feldman, J. Goldenberg, and M. Fresko, "Mine your own business: Market-structure surveillance through text mining," *Marketing Science,* vol. 31, no. 3, pp. 521-543, 2012.

[29] A. Ghose, P. G. Ipeirotis, and B. Li, "Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content," *Marketing Science,* vol. 31, no. 3, pp. 493-520, 2012.

[30] A. Ghose and P. G. Ipeirotis, "Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics," *IEEE transactions on knowledge and data engineering,* vol. 23, no. 10, pp. 1498-1512, 2010.

[31] N. Z. a. L. Y. L. He, "Research on the Evaluation of Product Quality Perceived Value Based on Text Mining and Fuzzy Comprehensive Evaluation," *International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI),* pp. 563-566, 2016.

[32] S. Güneş, "Extracting Online Product Review Patterns and Causes: A New Aspect/Cause Based Heuristic for Designers," *The Design Journal,* vol. 23, no. 3, pp. 375-393, 2020.

[33] S. C. a. S. R. V. C. Rangu, "Text Mining Approach for Product Quality Enhancement: (Improving Product Quality through Machine Learning," *IEEE 7th International Advance Computing Conference (IACC), 2017,* pp. 456-460,, 2017.

[34] D. K. Mishra, J. Kumar, J. K. Chaudhary, A. K. Upadhyay, and S. Sharma, "Role of Text Mining to Enhance the Quality of Product Using an Unsupervised Machine Learning Approach," *ECS Transactions,* vol. 107, no. 1, p. 12553, 2022.

[35] A. V. Cruz, "Relationship between product quality and customer satisfaction," 2015.

[36] S. Ullah, A. Khan, K. U. Khan, Amanullah, R. T. Chohan, and M. N. Khan, "Improving personalized recommendations system using graph attention networks driven by perceived complexity and innovation," *Scientific Reports,* 2026.

[37] S. Ullah, A. Khan, K. Khan, and A. Ullah, "EVOLUTION OF SENTIMENT ANALYSIS IN REVIEWS FOR INTELLIGENT PRODUCT RECOMMENDATIONS," *Kashf Journal of Multidisciplinary Research,* vol. 2, no. 01, pp. 142-166, 2025.

[38] S. Ullah, R. T. Chohan, and M. N. Khan, "Understanding Perceived Quality of Mobile Apps using DistilBERT Model," *IEEE Access,* 2025.

[39] J. R. Martin and P. R. White, *The language of evaluation*. Springer, 2003.

[40] H. RahmathP and T. Ahmad, "Fuzzy based sentiment analysis of online product reviews using machine learning techniques," *International Journal of Computer Applications,* vol. 99, no. 17, pp. 9-16, 2014.

[41] K. Ryan, "The role of natural language in requirements engineering," in *[1993] Proceedings of the IEEE International Symposium on Requirements Engineering*, 1993, pp. 240-242: IEEE.

[42] M. Mayo, "A general approach to preprocessing text data," *KDnuggets, https://www. kdnuggets. com/2017/12/general-approach-preprocessing-text-data. html,* 2017.

[43] M. Rathi, A. Malik, D. Varshney, R. Sharma, and S. Mendiratta, "Sentiment analysis of tweets using machine learning approach," in *2018 Eleventh international conference on contemporary computing (IC3)*, 2018, pp. 1-3: IEEE.

[44] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning* (no. 4). Springer, 2006.

[45] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013.

[46] V. Vapnik, "The support vector method of function estimation," in *Nonlinear modeling*: Springer, 1998, pp. 55-85.

[47] S. Xu, Y. Li, and Z. Wang, "Bayesian multinomial Naïve Bayes classifier to text classification," in *Advanced multimedia and ubiquitous engineering*: Springer, 2017, pp. 347-352.

[48] G. Singh, B. Kumar, L. Gaur, and A. Tyagi, "Comparison between multinomial and Bernoulli naïve Bayes for text classification," in *2019 International Conference on Automation, Computational and Technology Management (ICACTM)*, 2019, pp. 593-596: IEEE.

[49] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert systems with applications,* vol. 73, pp. 220-239, 2017.

[50] A. K. Chatterji and K. Fabrizio, "How do product users influence corporate invention?," *Organization Science,* vol. 23, no. 4, pp. 971-987, 2012.

[51] A. Gosain and S. Sardana, "Handling class imbalance problem using oversampling techniques: A review," in *2017 international conference on advances in computing, communications and informatics (ICACCI)*, 2017, pp. 79-85: IEEE.

[52] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of artificial intelligence research,* vol. 16, pp. 321-357, 2002.

[53] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, 2008, pp. 1322-1328: IEEE.

[54] I. Tomek, "Two modifications of CNN," 1976.