Preprints.org

Online Reviews Classification and Summarization: A Comprehensive Analysis

Saima Abdullah , Humaira Ashraf * , NZ Jhanjhi *

Review

Online Reviews Classification and Summarization: A Comprehensive Analysis

Saima Abdullah [1,2], Humaira Ashraf [1,*] and NZ Jhanjhi [3,*]

[1] Department of Computer Science, Faculty of Computing, International Islamic University, Islamabad, Pakistan; Saima.abdullah30@gmail.com

[2] Department of Computer Science, Faculty of Computer Sciences, IBADAT International University, Islamabad, Pakistan

[3] School of Computer Science, SCS, Taylors University, Subang Jaya, Malaysia

* Email: humaira.ashraf@iiu.edu.pk, noorzaman.jhanjhi@taylors.edu.my

Abstract: The classification and summarization of online reviews using machine learning and deep learning-based approaches are the subject of this systematic literature review. Consumers, companies, and researchers can now find vital information from online reviews. Automatically categorizing and summarizing these reviews can offer useful information and facilitate decision-making. The purpose of this study is to offer a thorough overview of the current methodology, approaches, and developments in this area. The review concentrates on machine learning-based methods for categorization and summarizing tasks, highlighting their advantages, drawbacks, and new developments. This systematic review provides insights into the most recent methodologies and indicates prospective directions for further investigation by synthesizing the data from a wide range of pertinent research publications.

Keywords: online reviews; classification; summarization; machine learning; systematic literature review

1. Introduction

A systematic literature review (SLR) is a comprehensive method of identifying and evaluating existing research on a particular topic. It involves a structured and systematic approach for literature search, selection, and analysis to answer a specific research question. Conducting a systematic literature review for the classification and summarization of online reviews using machine learning approaches can provide valuable details about the current research in this field and guide about future research directions.

1.1. Research Title:

Classification and summarization of online reviews – A Machine Learning Approach

1.2. String Development:

Developing multiple search strings for a systematic literature review is a common practice, and can help to ensure comprehensive coverage of the literature.

- For this SLR, keywords were identified for the research title.
- Three synonyms for each keyword were identified.
- Strings were developed according to three synonym for each keyword.
- In total 15 strings were formed.

Table 1 shows the keyword with their synonyms identified:

2

Table 1. Keywords with synonyms.

Keyword	Synonym1	Synonym2	Synonym3
Classification	Categorizing	Grouping	Organization
Summarization	Abstract Generation	Encapsulation	Outline
Online Review	Online Evaluation	Online Audit	Online Survey
Approach	Framework	Perspective	Perception

Table 2 shows all the strings formed.

Table 2. Strings formed of keywords.

| No.	Strings
1.	Classification and summarization of online reviews- A machine learning approach
2.	Categorizing and summarization of online reviews – A machine learning approach
3.	Grouping and summarization of online reviews- A machine learning approach
4.	and summarization of online reviews- A machine learning approach
5.	Organization and summarization of online reviews- A machine learning approach
6.	Classification and Abstract Generation of online reviews- A machine learning approach
7.	Classification and encapsulation of online reviews- A machine learning approach
8.	Classification and summing up of online reviews- A machine learning approach
9.	Classification and summarization of online Evaluations- A machine learning approach
10.	Classification and summarization of online audits- A machine learning approach
11.	Classification and summarization of online Surveys- A machine learning approach
12.	Classification and summarization of online Surveys- A machine learning Framework
13.	Classification and summarization of online Surveys- A machine learning Perspective
14.	Classification and summarization of online Surveys- A machine learning Perception

A search strategy was made to categorize all the searches according to the search journals. Three journals are used for searching research articles. Moreover, the latest research articles were included for the last four years and filtered according to their title, abstract, and objectives.

1.3. Searching protocol:

The search protocol was based on the rule, according to which:
- All research articles published in the years
- 2019,
- 2020,
- 2021,
- 2022,
- 2023 were searched.
- Three research databases
  - IEEE,
  - ACM,
  - Elsevier   were used for searching.

Figure 1 shows the search strategy with number of research databases.
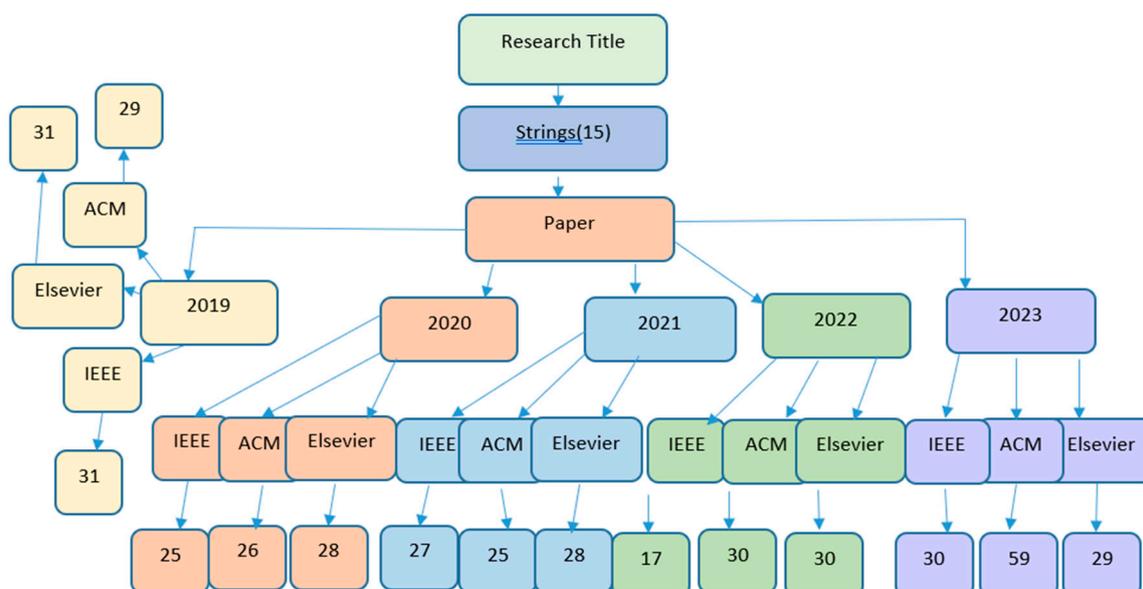
Figure 1. Searching Process using different research articles databases.

1.4. Inclusion Criteria:

The following criteria was followed for including research articles:

- All papers from years 2019, 2020, 2021, 2022, 2023 were included.
- Papers from three pages of research articles databases were included.
- Papers that were relevant to research strings were included.
- Papers that are not published yet were not included.

 1.5. Screening:

Research article screening is the process of reviewing a large number of research articles to identify the most relevant ones for a particular research project or study. For given research strings, in total 453 research articles were selected. Screening was done in phases:

2. Title Based screening:

Title-based screening is the first step in the systematic literature review process. It involves screening the titles of potentially relevant articles to determine whether they meet the inclusion criteria for the review. This initial screening is usually done by one or more reviewers and is often based on a pre-defined set of criteria.

For this SLR, all papers with titles not relevant to the research were excluded from selected papers of different databases. In total there were 453 papers were identified. After screening 80 papers were selected whose titles were related to research question.

3. Abstract Based Screening:

In this phase, the abstracts of articles identified in the initial search are reviewed to determine whether they meet the inclusion criteria for the review. In the second phase, abstracts of selected papers were studied. Papers with irrelevant abstracts were excluded. After screening, 37 papers were selected whose abstracts were most relevant with research for classification and summarization of online reviews.

4. Objective Based Screening:

In third phase, objectives of papers were considered. Objectives of every paper were identified and papers were categorized according to their objectives. As a result of objectives based screening, 13 papers were selected. Objectives that were identified are enlisted in Table 3.

**Table 3.** Objectives abbreviation Table.

Objective	Abbreviation
Precision	PR
Recall	RC
Accuracy	ACC
F-Score	F-1 SC
Summarization	SUM
Classification	CLS
Machine Leaning and Deep learning Techniques	ML &DL
Aspect Polarity Classification	APC
Emotion Detection	ED
Topic Modeling	TM

Table 4 presents the summary of Objectives based screening.

Table 4. Objectives based screening.

Reference.	PR.	RC	ACC	F-1 SC	SUM	CLS	ML&DL	APC	ED	TM
1	-	-	-	-	✓	-	-	-	-	-
2	-	-	-	-	✓	✓	✓	-	-	-
3	✓	✓	✓	✓	-	✓	✓	-	-	-
4	✓	✓	-	-	-	✓	-	-	-	-
5	-	-	-	-	-	-	✓	-	-	-
6	-	-	-	-	-	✓	-	✓	-	-
7	-	-	-	-	-	✓	-	-	✓	-
8	-	-	-	-	-	✓	✓	-	-	-
9	-	-	✓	-	-	✓	-	-	-	-
10	-	-	-	-	-	✓	✓	-	-	-
11	✓	✓	✓	✓	-	✓	✓	-	-	-
12	✓	✓	✓	✓	✓	✓	-	-	-	-
13	-	-	-	✓	-	✓	-	-	-	✓

Figure 2. shows results of screening at all phases.
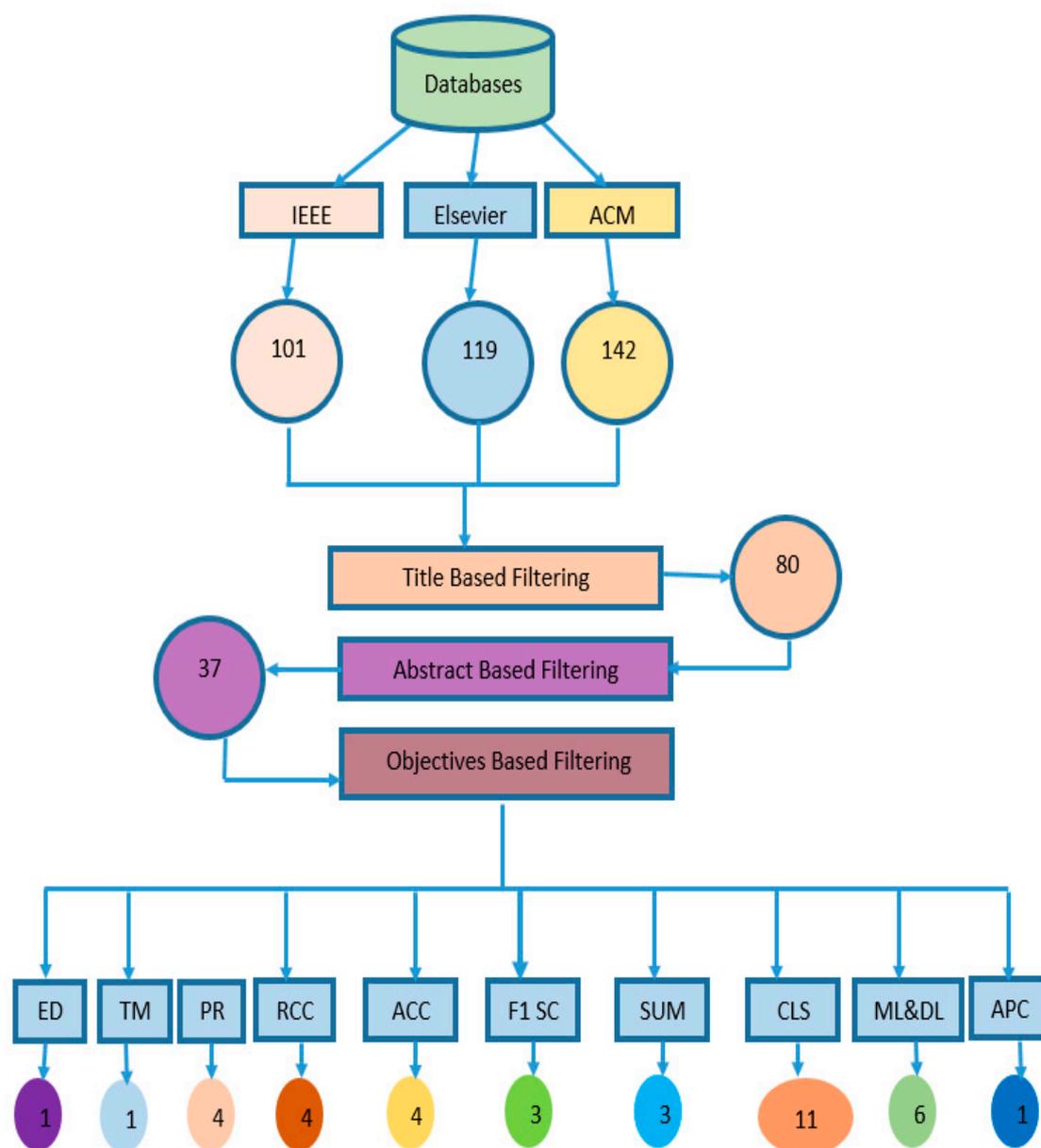
Figure 2. Screening for research articles selection.

Table 5. Objective notation table:.

Article	Objectives	Inclusion Criteria	Decision
1	To proposes a novel graph-based technique for generating abstractive summaries of duplicate sentences	Paper is relevant to research where graph based technique is used to generate summaries of sentences.	Included
2	To perform sentiment classification from the point of view of the consumer review summarization model	Paper is relevant to research where hybrid approach is used for sentiment analysis.	Included
3	To develop a hybrid model for sentiment analysis using convolutional neural network-long short-term memory	Relevant to research where ML &DL techniques are used.	Included
4	A study is presented to identify relevant features that could be used to classify customer satisfaction.	Relevant to research	Included

5	Combining machine learning (ML) techniques for consumer sentiment analysis (CSA) in the domain of hospitality and tourism	ML and DL techniques combined and analyzed for sentiment analysis	Included
6	To identify aspect terms in online reviews and performs aspect polarity classification.	The paper is the most recent and related to research.	Included
7	Analyze recent trends in sentiment analysis and text-based emotion detection.	The paper is relevant to the research.	Included
8	Comparing multiple deep learning models for sentiment classification using different datasets	Deep Learning techniques for sentiment analysis.	Included
9	Analyze the sentiments as positive and negative and classify the text based on feedback.	Different machine learning techniques are used for sentiment analysis to find the best one.	Included
10	To present an approach for automatic comment analysis and classification	The paper is relevant to the research where Machine Learning techniques are used.	Included
11	Comparison of multiple machine learning algorithms for sentiment analysis and classification	The Paper is relevant to the research where Machine Learning techniques are used.	Included
12	Classification of sentiments and summarization of reviews with optimal values for Precision, recall, Accuracy, F-1 Score.	Paper is relevant to research where Machine learning techniques are used and paper is latest one.	Included
13	Analyzing online reviews by integrating multiple techniques including topic modeling	The paper is relevant to the research.	Included

5. Detailed Literature:

The main aim of all the papers was to classify and summarize the online reviews using different Machine Learning and Deep Learning Techniques.

In [1], presents a comparative study of different techniques used for opinion summarization, including abstractive and extractive approaches that consider aspects to summarize sentences. This research work identifies gaps in previous studies and proposes a novel graph-based technique for generating abstractive summaries of duplicate sentences. This work contributes to the field of opinion summarization by proposing a new technique and providing a comprehensive comparison of different approaches.

In [2], research work focuses on developing a consumer review summarization model that uses Natural Language Processing (NLP) techniques and Long short-term memory (LSTM) to provide concise and meaningful data, enabling businesses to gain valuable insights into their customers' behavior and preferences. The authors have presented a hybrid approach for sentiment analysis. The first step is pre-processing of data then feature extraction and lastly sentiment classification is done. Using Natural Language Processing techniques, the pre-processing phase removes the undesirable data from input text reviews. The proposed approach in this study involves the use of a hybrid feature extraction method that combines review-related and aspect-related features to construct a distinct feature vector for each review. This approach aims to effectively extract features for sentiment analysis. The sentiment classification is carried out using a deep learning classifier called LSTM. The proposed model was experimentally evaluated on three different research datasets. The model achieved the average precision, average recall, and average F1-score of 94.46%, 91.63%, and 92.81%, respectively.

In [3], the objective is to develop a hybrid model for sentiment analysis using convolutional neural network-long short-term memory (CNN-LSTM). The model is designed with dropout, max

pooling, and batch normalization techniques to improve its performance. To transform text into numerical vectors, the Keras word embedding approach is used, which results in small vector distances between similar words. To evaluate the model's performance, experiments were conducted using the Airline quality and Twitter airline sentiment datasets and measured various parameters such as accuracy, precision, recall, and F1-measure. Our proposed model's performance was found to be superior to classical machine learning models in sentiment analysis. Our analysis of the results revealed that our model achieved an accuracy of 91.3% in sentiment analysis, outperforming other existing models.

In [4], a study is presented to identify relevant features that could be used to classify customer satisfaction based on 47,172 reviews of 33 Las Vegas hotels registered with Yelp, a social networking site. The Naive algorithm is used in the study which demonstrated high precision and recall in classifying hotel reviews with a low computational cost. The findings of this study are more reliable and accurate than prior statistical results based on limited sample data, and they provide insights into how hotels can enhance their services by focusing on staff experience, professionalism, tangible and experiential factors, and gambling-based attractions.

In [5], authors have presented a study aimed at determining the usefulness, scope, and applicability of combining machine learning (ML) techniques for consumer sentiment analysis (CSA) in the domain of hospitality and tourism. A systematic literature review is conducted to compare, analyze, explore, and identify research gaps in order to provide insights into the potential future developments of this approach. The main objective of this research is to investigate the use of ML techniques for CSA in online reviews related to hospitality and tourism, and its implications for service providers in terms of developing managerial strategies to meet the needs of consumers. Additionally, the study provides valuable information for researchers regarding potential future research directions in this field.

In [6], the authors have discussed the challenges of aspect polarity classification (APC) in natural language processing (NLP) and the importance of utilizing dependency syntax information with a graph neural network (GNN) to improve performance. The authors presented a multitask learning model that combines APC and aspect term extraction (ATE) tasks to simultaneously extract aspect terms and classify aspect polarity. They also use multi-head attention (MHA) to associate dependency sequences with aspect extraction, allowing the model to focus on words closely related to aspects. The experimental results on three benchmark datasets demonstrate that the proposed model enhances aspect polarity classification performance significantly.

In [7], author has focused on sentiment analysis and emotion detection in text. This review provides a comprehensive analysis of recent trends in text-based emotion detection, highlighting the shift from traditional sentiment analysis to emotion detection, and the challenges involved. The article summarizes recent works from the past five years and looks at the methods used, as well as the models of emotion classes that are generally referenced. The trend in text-based emotion detection has moved away from early keyword-based comparisons to machine learning and deep learning algorithms that offer more flexibility and improved performance.

In [8], the presented study aims to compare various deep learning models for sentiment classification using 13 different review datasets. The focus is on two types of input structures, word level and character level, and eight different deep learning models, three of which are based on convolutional neural networks and five based on recurrent neural networks. The study analyzes and discusses the classification performances of these models from various perspectives to provide meaningful implications for building sentiment classification models.

In [9], authors have presented an approach for examining customer reviews. The study has two main objectives: sentiment analysis to classify comments as positive or negative and text categorization to classify comments based on feedback about food taste, ambiance, service, and value for money. A manually annotated dataset of approximately 4,000 records was utilized to train and test several classification algorithms, including Naive Bayes Classifier, Logistic Regression, Support Vector Machine (SVM), and Random Forest. The performance of these algorithms was compared, and the Random Forest algorithm achieved the best accuracy of 95%. The paper's findings contribute

to understanding customer feedback about restaurants in Karachi and provide insights into the effectiveness of different classification algorithms for sentiment and text categorization.

In [10], the authors have proposed a system that can effectively determine the polarity of customer comments collected from Amazon and Flipkart data domains. Five supervised learning classifiers, namely Naïve Bayes, Linear Regression, SentiWordNet, Random Forest, and KNN. The paper also discusses the experimental results and challenges encountered during the study. The paper presents a comprehensive approach to automatic comment analysis and classification. It provides insights into various techniques and algorithms used in comment analysis and classification and their effectiveness in handling large sets of reviews. The paper also highlights the challenges encountered during the study, which can be useful for researchers working in this area.

In [11], the authors have compared the performance of five different machine learning classifiers (Bernoulli Naïve Bayes, Decision Tree, Support Vector Machine, Maximum Entropy, and Multinomial Naïve Bayes) on a dataset of movie reviews for analyzing sentiments and classifying those into positive, negative and neutral. The results of the analysis suggest that the Multinomial Naïve Bayes classifier achieves the best overall performance in terms of accuracy, precision, and F-score, while the Support Vector Machine classifier has the highest recall. The analysis also found that the Bernoulli Naïve Bayes classifier performed better than in a previous experiment.

It is important to note that these results may not necessarily generalize to other datasets or applications, and that the performance of different classifiers can vary depending on the specific characteristics of the data and the task at hand. Therefore, it is important to carefully evaluate the performance of different classifiers on a specific task and select the one that performs best.

In [12], a method is proposed for classifying and summarizing movie reviews automatically. The aim is to handle the vast amount of movie reviews available online and provide users with a quick way to identify the positive and negative aspects of a movie. The breakdown of the proposed method is consisting of Movie Review Classification and Movie Review Summarization. For classification, Bag-of-words feature extraction technique is used to extract unigrams (single words), bigrams (pairs of consecutive words), and trigrams (triplets of consecutive words) from the movie review documents. Then the review documents are represented as vectors using the extracted features. Naïve Bayes algorithm is used to classify the movie reviews as either negative or positive based on the feature vectors. For reviews summarization, the Word2Vec model is used to extract features from the classified movie review sentences. A semantic clustering technique is applied to group together review sentences that are semantically related, creating clusters of similar sentences. Various text features are used to compute the salience score of each review sentence within the clusters and the review sentences with the highest salience scores are chosen to form a summary of the movie reviews. The presented machine learning approach is evaluated and compared against other benchmark summarization approaches. As per the achieved results, the suggested method performs better than these benchmark approaches in terms of summarizing movie reviews.

In [13], the authors have presented the approach to analyze online student reviews using text analytics to identify the current strengths, weaknesses, opportunities, and threats (SWOT) of a university. The proposed approach integrates four techniques: topic modeling, sentiment analysis, root cause analysis, and SWOT analysis. A case study is used to demonstrate the feasibility and application of the presented approach. The results indicate that the method provides an efficient and cost-effective performance summary of the university and its competitors. It can be utilized by university leaders to enhance their recruitment and retention efforts. In our comprehensive analysis of online reviews classification and summarization, our research draws upon foundational insights presented in [24–35].

3. Critical Analysis:

Table 5 summarizes all schemes and algorithms for sentiment analysis, classification and summarization of online reviews used in reviewed literature. Limitations of the approach are also given.

The research work in [1] is the comparative study of extractive and abstractive techniques for opinion summarization. For abstractive summaries, it uses novel graph based approach and for extractive approach, it uses principal component analysis (PCA) for reducing the number of dimensions in data. The comparison is made between both approaches to discuss which method provides the most related and complete summary. A novel algorithm is presented that uses a combination of PCA and Singular value Decomposition (SVD). But PCA can only be used for smaller datasets. Also SVD can only be used on linear datasets so it is not appropriate for all datasets.

In [2], a customer's review summarization model is developed using natural language processing techniques and deep learning model that is LSTM. The process consists of pre-processing, feature extraction, and sentiment classification. Experimental evaluation was done on three different datasets. For sentiment classification Long Short-Term Memory (LSTM) algorithm is used but this requires a longer processing time.

In [3], novel hybrid classification model is proposed that is based on coupling of classification methods. The Classifier Collection was constructed using Naïve Bayes (NB), Support Vector Machine (SVM) and Genetic Algorithm (GA).

In [4], the naïve Bayes algorithm is used to classify reviews of hotels customers and results in high precision and recall values. This technique has worked well for limited data sample. But for larger dataset, the technique may not work well. Naive Bayes assumes that all predictors (or features) are independent, rarely happening in real life. This limits the applicability of this algorithm in real-world use cases. [15]

In [5], the primary objective of this study is to conduct a comprehensive analysis of the use of ML techniques for consumer sentiment analysis in the context of online reviews within the hospitality and tourism domain. The results shows that machine learning based techniques are very helpful in improving the performance of the model. But in this research, factors that are affecting the performance of the model are not drawn as current research articles are not included in this research[5].

In [6], the authors have proposed a multitask learning model that combined Aspect Term Extraction (ATE) and Aspect Polarity Classification (APC) that can extract aspect terms and also provide classification of aspect polarity simultaneously. Also multihead attention(MHA) is exploited to associate dependency sequences with aspect extraction. But the process of training of an MTL network can be more complicated as multiple tasks can compete with each other to get a better learning representation and one or more tasks can dominate the training process[16]. Also the loss function of MTL can be complex as there are multiple summed losses, therefore making the optimization more difficult [17].

In [7], a review is presented that examines the recent shift in focus from text sentiment analysis to emotion detection in natural language processing (NLP) tasks. In this paper work of authors have been reviewed. Some of the recent techniques used for emotion detection through text are including Na¨ıve Bayes, maximum entropy, SVM, Logistic regression, Na¨ıve Bayes, CNN, ensemble of CNN and many others. It can be seen that most researchers have preferred to use the Ekman model of emotions to define the emotion classes of their work. Neural networks, are often considered black boxes, making it difficult to interpret how and why they make specific emotion predictions. Deep learning models heavily rely on large amounts of labeled training data. However, emotion detection datasets are often limited in size and quality, making it challenging to train accurate and generalizable models. Insufficient or biased training data can lead to poor performance and limited applicability of the models.

In [8], authors have conducted a comparative analysis of various deep-learning-based model structures for sentiment classification. Specifically, the study compared eight deep-learning models, including three based on convolutional neural networks (CNNs) and five based on recurrent neural networks (RNNs). The comparison is performed on 13 different review datasets, and the classification performances are thoroughly examined from various perspectives. There can be an overall problem with all deep learning techniques that are studied in this research. In sentiment classification based on deep learning models, the best model structure depends on the characteristics of the dataset on

which this model is trained. Also, model is manually selected based on the domain knowledge of an expert or selected from a grid search of possible candidates [8].

In [9], presented research is two-fold. Firstly, it involves sentiment analysis where comments are analyzed and classified as positive or negative based on their sentiment. Secondly, text categorization techniques are employed to automatically classify comments based on feedback related to food taste, ambiance, service, and value for money. Naive Bayes Classifier, Logistic Regression, Support Vector Machine (SVM), and Random Forest algorithms were compared for performance and Random Forest algorithm had performed best with highest accuracy. But Random Forest algorithm has limitations in terms of generalization. Unlike some other algorithms, it cannot generalize beyond the range of previously observed labels. Instead, it provides predictions that are an average or combination of the labels observed during training. This behavior can pose challenges when the training and prediction inputs differ significantly in their range or distribution [19].

In [10], researchers have addressed, reviewed and compared different algorithms for automatic identification of the sentiments expressed in the English text for Amazon and Flipkart products. Used techniques are Random Forest and K-Nearest Neighbor techniques. Problems identified with both techniques are A Random Forest can't generalize. It can only make a prediction based on previously observed labels [19] and with KNN, classification is slow when dataset is larger. Also it cannot deal with missing values [20].

In [11], authors have compared the performance of different machine learning algorithms for movie review dataset that was collected. The evaluated several classifiers were Bernoulli Naïve Bayes (BNB), Decision Tree (DE), Support Vector Machine (SVM), Maximum Entropy (ME), and Multinomial Naïve Bayes (MNB), for sentiment analysis. The results revealed that MNB outperformed the other classifiers in terms of accuracy, precision, and F-score. Additionally, SVM demonstrated higher recall compared to the other classifiers. Also results showed that the BNB classifier achieved improved accuracy compared to previous experiments involving this classifier. However, there can be early convergence and the cold start issues, encountered in the multinomial models[21].

In [12], an approach is proposed for classification of movie reviews and summarization of those reviews. Bag-of-words feature extraction technique is used for extracting features for movie review classification. Then the Na¨ıve Bayes algorithm is used to identify the movie reviews as positive and negative. For movie review summarization, Word2vec feature extraction technique is used to extract features from classified movie review sentences. Semantic clustering technique is used to cluster semantically related review sentences. But used techniques do not always perform best. For larger data sets, bag of words technique cannot work best. As feature dimension is dependent on unique tokens. Also Bag of words does not preserve the relationships between tokens[22]. Word2Vec cannot understand the words that are not available in training data. Also it ignores the formation of words with same meaning[22].

In [13], proposed approach is an ensemble approach called Latent Dirichlet Allocation (E-LDA) topic models for the automatic identification of key features (topics) predominantly discussed by students. The comprehensive approach used in the research enabled to gain valuable insights into the university's performance and make informed strategic decisions based on the students' opinions. LDA technique has some limitations including limited number of generated topics, unsupervised and sentence structure is not modeled[23].

Table 5. Summary of critical Analysis of reviews classification and summarization techniques.

Detection Algorithm	Effort Year	Technique	Performance Metrics	Shortcoming
[1]	2021	Abstractive and Extractive summarization of reviews with the use of PCA and SVD.	Precision, Recall, F-Measure	PCA cannot be used for larger datasets and SVD is appropriate only for linear datasets

[2]	2023	Term frequency-inverse document frequency (TF-IDF), n-gram features & emoticon polarities , Long Short Term Memory (LSTM)	Precision , accuracy, Recall and F-1 Score, AUC	LSTM requires a longer time to process. [14]
[3]	2019	Hybrid Technique, KNN	Accuracy	KNN is not efficient when training data increases. It has Poor performance on imbalanced data If Optimal value of K is chosen incorrectly, the model will be under or over fitted to the data
[4]	2019	Reviews classification with Naïve Bayes	High Precision, recall with low computational cost.	• Naive Bayes assumes that all predictors (or features) are independent, rarely happening in real life. This limits the applicability of this algorithm in real-world use cases. [15]
[5]	2021	Machine Learning Techniques for reviews classification	Provide novel ML-CSA framework and provides guidelines that have never been provided in earlier SLRs. [5]	The proposed framework is based on research till year 2020, some of the research is not tested. [5]
[6]	2023	multitask learning model with main focus on aspect polarity classification	Accuracy, Macroaverage F1 Score	In the process of training of an MTL network, multiple tasks can compete with each other to get a better learning representation and one or more tasks can dominate the training process. [16] Also the loss function of MTL can be complex because of multiple summed losses, therefore making the optimization more difficult. [17]
[7]	2023	Review of Machine Learning and Deep Learning Techniques	Avg. Recall, Avg. Accuracy, Avg. Precision	Classical machine learning methods cannot work better for larger and higher dimensions data[18]. Emotion detection datasets are often

				limited in size and quality, making it challenging to train accurate and generalizable models .
[8]	2020	Deep Learning based techniques are compared	Area under the receiver operating characteristics curve (AUROC)	In sentiment classification based on deep learning models , the best model structure depends on the characteristics of the dataset on which this model is trained. Also, model is manually selected based on the domain knowledge of an expert or selected from a grid search of possible candidates. [8]
[9]	2020	Naive Bayes Classifier, Logistic Regression, Support Vector Machine (SVM), and Random Forest algorithms are compared. Random Forest performed best.	Accuracy, precision, Recall, F-1 Score	A Random Forest can't generalize. It can only make a prediction based on previously observed labels. [19]
[10]	2022	Random Forest , K-Nearest Neighbour (KNN)	Accuracy, precision, Recall, F-1 Score	A Random Forest can't generalize. It can only make a prediction based on previously observed labels.[19] With KNN, classification is slow when dataset is larger. Also it cannot deal with missing values. [20]
[11]	2019	Bernoulli Naïve Bayes (BNB), Decision Tree (DE), Support Vector Machine (SVM), Maximum Entropy (ME), as well as Multinomial Naïve Bayes (MNB)	Accuracy, Precision and F-1score, Recall.	Early convergence and the cold start issues, encountered in the multinomial models [21].
[12]	2020	Bag-of-Words for feature Extraction, Naïve Bayes for review classification, word2vec for feature extraction for summarization, Semantic Clustering for summarization generation	Accuracy, Precision and F-measure, Recall.	For larger data sets, bag of words technique cannot work best. As feature dimension is dependent on unique tokens. Also Bag of words does not preserve the relationships between tokens[22]. Word2Vec cannot understand the words

13

				that are not available in training data. Also it ignores the formation of words with same meaning[22].
[13]	2019	Latent Dirichlet Allocation (E-LDA)	F-1 Score	Limited number of generated topics, unsupervised and sentence structure is not modeled[23].

4. Research Challenges:

This section presents the issues and challenges of all techniques used for reviews classification and summarization in studied literature in Table 6. It briefly describes the limitations of all the schemes. These are the open research areas in which work can be conducted in the future to overcome the issues and challenges addressed by the schemes

Table 6. Identified challenges of reviews classification and summarization.

Table 6. Techniques with challenges.

Technique	Challenge
Machine Learning Based Algorithms [1,2,4,5,7,9,11,12]	With some algorithms, longer time is required to process the data. Naive Bayes has less in real-world use cases. This technique does not provide best results for larger and high dimensional dataset.
Latent Dirichlet Allocation (E-LDA) [13]	Old technique for topic modeling where limited number of topics can be generated. It is unsupervised technique.
Deep Learning Based Techniques [8,10]	The best model structure depends on the characteristics of the dataset on which this model is trained. Model is manually selected. With some models, classification is slow when dataset is larger. Also cannot deal with missing values. .
Multi Task Learning Models [6]	Multiple tasks can compete with each other while training of an MTL network and one or more tasks can dominate the training process.

6. Conclusions

Many Schemes are investigates that are used for reviews classification and summarization. The reviewed schemes encompass a range of approaches, including AI- and ML-based techniques, deep learning-based methods, multitask learning-based approaches, and E-LDA (Latent Dirichlet Allocation) based schemes. The SLR provides an extensive critical and comparative analysis of these schemes, focusing on evaluating their performance based on metrics such as accuracy, precision, F-1 score, and recall. By analyzing and comparing these schemes, the paper identifies gaps in the existing literature, indicating future research directions for analyzing online reviews and generating review summaries. Notably, recent studies have demonstrated the successful application of artificial intelligence systems for sentiment analysis, with deep learning-based schemes showing better results.

References:

1. S. Bhatia, "A Comparative Study of Opinion Summarization Techniques," in *IEEE Transactions on Computational Social Systems*, vol. 8, no. 1, pp. 110-117, Feb. 2021, doi: 10.1109/TCSS.2020.3033810.
2. Kaur, G., Sharma, A. "A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis" J Big Data 10, 5 (2023)

3. N. Yadav, R. Kumar, B. Gour and A. U. Khan, "Extraction-Based Text Summarization and Sentiment Analysis of Online Reviews Using Hybrid Classification Method," *2019 Sixteenth International Conference on Wireless and Optical Communication Networks (WOCN)*, Bhopal, India, 2019, pp. 1-6, doi: 10.1109/WOCN45266.2019.8995164.

4. Manuel J. Sánchez-Franco, Antonio Navarro-García, Francisco Javier Rondán-Cataluña, "A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services" , Journal of Business Research, Volume 101, 2019, Pages 499-506.

5. Praphula Kumar Jain, Rajendra Pamula, Gautam Srivastava, A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews, Computer Science Review,Volume 41, 2021.

6. Zhao, Guoshuai & Luo, Yiling & Chen, Qiang & Qian, Xueming. Aspect-based sentiment analysis via multitask learning for online reviews. (2023) Knowledge-Based Systems. 264. 110326. 10.1016/j.knosys.2023.110326.

7. Hung, Lai & Alias, Suraya. Beyond Sentiment Analysis: A Review of Recent Trends in Text-Based Sentiment Analysis and Emotion Detection. (2023). Journal of Advanced Computational Intelligence and Intelligent Informatics. 27. 84-95. 10.20965/jaciii.2023.p0084.

8. S. Seo, C. Kim, H. Kim, K. Mo and P. Kang, "Comparative Study of Deep Learning-Based Sentiment Classification," in IEEE Access, vol. 8, pp. 6861-6875, 2020.

9. K. Zahoor, N. Z. Bawany and S. Hamid, "Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning," 2020 21st International Arab Conference on Information Technology (ACIT), Giza, Egypt, 2020, pp. 1-6

10. Dadhich, A., Thankachan, B. (2022). Sentiment Analysis of Amazon Product Reviews Using Hybrid Rule-Based Approach. In: Somani, A.K., Mundra, A., Doss, R., Bhattacharya, S. (eds) Smart Systems: Innovations in Computing. Smart Innovation, Systems and Technologies, vol 235. Springer, Singapore.

11. A. Rahman and M. S. Hossen, "Sentiment Analysis on Movie Review Data Using Machine Learning Approach," 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, Bangladesh, 2019, pp. 1-4

12. Atif Khan, Muhammad Adnan Gul, M. Irfan Uddin, Syed Atif Ali Shah, Shafiq zAhmad, Muhammad Dzulqarnain Al Firdausi, Mazen Zaindin, "Summarizing Online Movie Reviews: A Machine Learning Approach to Big Data Analytics", Scientific Programming, vol. 2020, Article ID 5812715, 13 pages, 2020.

13. Sharan Srinivas, Suchithra Rajendran, Topic-based knowledge mining of online student reviews for strategic planning in universities, Computers & Industrial Engineering, Volume 128, 2019, Pages 974-984,

14. N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: a comparative study," Electronics, vol. 9, no. 3, p. 483, 2020.

15. Naïve Bayes Algorithm's Advantages and Disadvantages | Data Science and Machine Learning | Kaggle

16. Alex Kendall, Yarin Gal, Roberto Cipolla , "Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics". In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2018.

17. How to Do Multi-Task Learning Intelligently (thegradient.pub)

18. Cai, Y.; Li, X.; Li, J. Emotion Recognition Using Different Sensors, Emotion Models, Methods and Datasets: A Comprehensive Review. *Sensors* 2023, *23*, 2455. https://doi.org/10.3390/s23052455

19. A limitation of Random Forest Regression | by Ben Thompson | Towards Data Science

20. Introduction to Classification Using K Nearest Neighbours | by Ashwin Raj | Towards Data Science

21. Multinomial Naïve Bayes' For Documents Classification and Natural Language Processing (NLP) | by Arthur V. Ratz | Towards Data Science

22. All You Need to Know About Bag of Words and Word2Vec — Text Feature Extraction | by Albers Uzila | Towards Data Science

23. Latent Dirichlet Allocation. Latent Dirichlet Allocation (LDA) is… | by Harsh Bansal | Analytics Vidhya | Medium

24. Shahid, H., Ashraf, H., Javed, H., Humayun, M., Jhanjhi, N. Z., & AlZain, M. A. (2021). Energy optimised security against wormhole attack in iot-based wireless sensor networks. Comput. Mater. Contin, 68(2), 1967-81.

25. Wassan, S., Chen, X., Shen, T., Waqar, M., & Jhanjhi, N. Z. (2021). Amazon product sentiment analysis using machine learning techniques. Revista Argentina de Clínica Psicológica, 30(1), 695.

26. Almusaylim, Z. A., Zaman, N., & Jung, L. T. (2018, August). Proposing a data privacy aware protocol for roadside accident video reporting service using 5G in Vehicular Cloud Networks Environment. In 2018 4th International conference on computer and information sciences (ICCOINS) (pp. 1-5). IEEE.

27. Ghosh, G., Verma, S., Jhanjhi, N. Z., & Talib, M. N. (2020, December). Secure surveillance system using chaotic image encryption technique. In IOP conference series: materials science and engineering (Vol. 993, No. 1, p. 012062). IOP Publishing.

28. Humayun, M., Alsaqer, M. S., & Jhanjhi, N. (2022). Energy optimization for smart cities using iot. Applied Artificial Intelligence, 36(1), 2037255.
29. Hussain, S. J., Ahmed, U., Liaquat, H., Mir, S., Jhanjhi, N. Z., & Humayun, M. (2019, April). IMIAD: intelligent malware identification for android platform. In 2019 International Conference on Computer and Information Sciences (ICCIS) (pp. 1-6). IEEE.
30. Diwaker, C., Tomar, P., Solanki, A., Nayyar, A., Jhanjhi, N. Z., Abdullah, A., & Supramaniam, M. (2019). A new model for predicting component-based software reliability using soft computing. IEEE Access, 7, 147191-147203.
31. Gaur, L., Afaq, A., Solanki, A., Singh, G., Sharma, S., Jhanjhi, N. Z., ... & Le, D. N. (2021). Capitalizing on big data and revolutionary 5G technology: Extracting and visualizing ratings and reviews of global chain hotels. Computers and Electrical Engineering, 95, 107374.
32. Nanglia, S., Ahmad, M., Khan, F. A., & Jhanjhi, N. Z. (2022). An enhanced Predictive heterogeneous ensemble model for breast cancer prediction. Biomedical Signal Processing and Control, 72, 103279.
33. Kumar, T., Pandey, B., Mussavi, S. H. A., & Zaman, N. (2015). CTHS based energy efficient thermal aware image ALU design on FPGA. Wireless Personal Communications, 85, 671-696.
34. Gaur, L., Singh, G., Solanki, A., Jhanjhi, N. Z., Bhatia, U., Sharma, S., ... & Kim, W. (2021). Disposition of youth in predicting sustainable development goals using the neuro-fuzzy and random forest algorithms. Human-Centric Computing and Information Sciences, 11, NA.
35. Lim, M., Abdullah, A., & Jhanjhi, N. Z. (2021). Performance optimization of criminal network hidden link prediction model with deep reinforcement learning. Journal of King Saud University-Computer and Information Sciences, 33(10), 1202-1210.